RAQUEL CASESNOVES FERRER and DAVID SANKOFF

TRANSMISSION, EDUCATION AND INTEGRATION IN PROJECTIONS OF LANGUAGE SHIFT IN VALENCIA

(Received 22 February 2004; accepted in revised form 29 March 2004)

ABSTRACT. We assess the relative importance of demographic and linguistic processes for the evolution of competence in understanding, speaking, reading and writing Catalan in the Autonomous Community of Valencia, Spain. Our projections involve cohorts defined by age, sex, geographic division, and language competence level, and incorporate sociolinguistic models of transmission, acquisition and immigrant integration. Based on projections for a large random sample of scenarios, multiple regression shows that school acquisition, immigration and linguistic integration of immigrants outweigh other processes in their long-term effects on competence levels.

KEY WORDS: bilingual acquisition, Catalan, demolinguistic projection and simulation software, geolinguistics, language revitalization, language shift, linguistic normalization, mathematical models, oral and written competence, scenario sampling

ABBREVIATIONS: CCE – Conselleria de Cultura i Educació de la Generalitat Valenciana; IVE – Institut Valencià d'Estadística; SIES – Servicio de Investigación y Estudios Sociolingüísticos de la Dirección General de Ordenación e Innovación Educativa y Política Lingüística de CCE

Introduction

The outcome of language shift and language revitalization processes is a burning question in bilingual communities around the world, and nowhere more so than in the Autonomous Community of Valencia in Spain. Efforts to halt or reverse the effects of Franco's forcible imposition of Castilian are complicated by such factors as a dramatic increase in overseas immigration, an extremely low birth rate, and foot-dragging or outright opposition to Catalan by authorities at various levels (Pradilla, 2002). Many factors may contribute to advancing or reversing the process of language shift in a community: language education, government practices, incentives and regulations, measures to integrate immigrants, the relative prestige of the competing languages and their social, political and ethnic associations within and outside the community. It is difficult to evaluate the

Language Policy **3:** 107–131, 2004.

relative importance of these factors, however, since they function simultaneously and because they operate against a backdrop of ongoing demographic processes, which could have a greater effect than all the attitudinal, social and political effects combined. This is complicated in the case of Valencia by the absence of systematic data on language usage except for some outdated small-sample surveys based on self-reports (CCE 1989a, 1992, 1995). At the time of this research (2002–2003), even the available census data on language competence dated to 1991, although results from 2001 are now becoming available.

The aim of this article is to assess the relative importance of the key demographic and linguistic processes for the evolution over the next 50 years of Community-wide competence in understanding, speaking, reading and writing Catalan. This is based on demographic projections, where the cohorts are defined by age, sex, *comarca* (county or canton), and language competence levels, into which we incorporate models of language transmission, acquisition and immigrant integration.

We generate a large random sample of scenarios based on six demographic and linguistic parameters, and apply multiple regression to the set of projections made under all of these scenarios to evaluate the relative importance of the parameter effects. Given that the input competence level data are severely out of date, we cannot place too much credence in the projections of the number of individuals at each level. But we will show that the estimates of the parameter effects depend very little on the linguistic input data, so that within the limitations of the models of language transmission, acquisition and integration, we can have confidence in our assessments of the relative importance of the various processes in play.

In the next section we present a sketch of the recent demography of the Community, as well as the available data on levels of competence in Catalan. In the following section, we discuss the processes of language transmission, acquisition and immigrant integration, present pertinent data from various sources and develop the models to be incorporated into our projections. We then describe the projections themselves. In the final few sections we first investigate the effects of the parameters one at a time, and then carry out the sampling and multiple regression analyses.

Because the true measure of the vitality of a language in a situation of language shift is the extent to which a language is used: how often, by whom and in which contexts, our projections of competence

can only be an indirect indication of how the situation will evolve. We take up this problem in the concluding discussion.

Trends in Valencian Demolinguistics

In Valencia, as in the rest of Spain and southern Europe, high birth rates in the sixties and seventies decreased in the ensuing period, reaching a nadir in the mid nineties. A birthrate of 82 per thousand in 1975 fell to 34 per thousand by 1996, and has more recently begun rising somewhat, to 38 per thousand by 2000. The total number of births is depicted in Figure 1, which also reflects a widespread postponement of childbearing among younger women. At the same time, life expectancy has increased, from 70 (men) and 76 (women) in 1975 to 74 and 81, respectively, in 1995 (IVE [http://www.ive.es]).

Recent years have seen an enormous influx of immigrants settling in the Community (Serra Yoldi, 2000), presently increasing the population by more than one percent per annum. Figure 2 shows that there are far more immigrants coming to Valencia than to either Catalonia or Andalusia, which have much larger populations. Of interest from the linguistic point of view is that in the last two years the greatest source of immigrants has been Latin America, virtually all from Castilian-speaking countries. Turning to the status of Catalan, the only census-based information available is based on responses to query items about understanding, speaking competence, and ability to read and write. This dates from the 1991 census, but the complete tabulations were not released until 1998. Complete

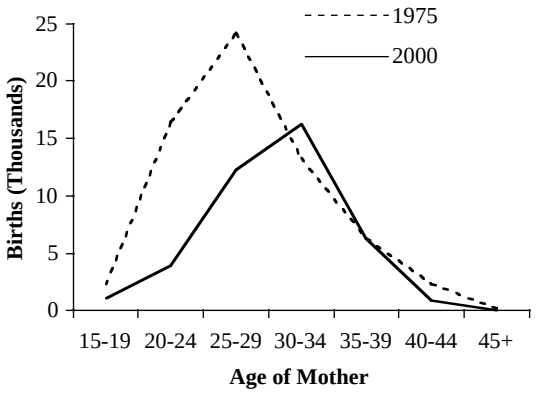

Figure 1. Number of births to women residing in the Valencia, by age group. Data from Instituto Nacional de Estadística (INE [http://www.ine.es]).

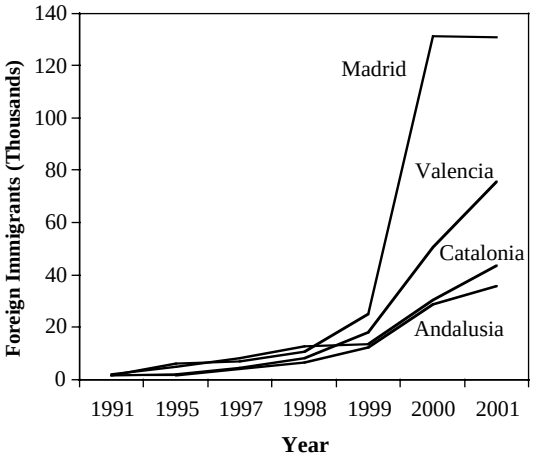

Figure 2. Foreign immigration to the four major receiving communities.

tabulations from the 2001 census are not yet available, though partial results are being released in late 2003.

Figure 3 summarizes the competence data from 1986 (CCE, 1989b) and 1991 (SIES, 1998). It should be kept in mind that with the exception of a small proportion of foreign residents, virtually everyone in Valencia understands and speaks Castilian. That understanding and speaking competences in Catalan are so low in the 3–9 age group reflects the fact that most children are not acquiring it at home, but rather in school. The 1986 curve for speaking and, to

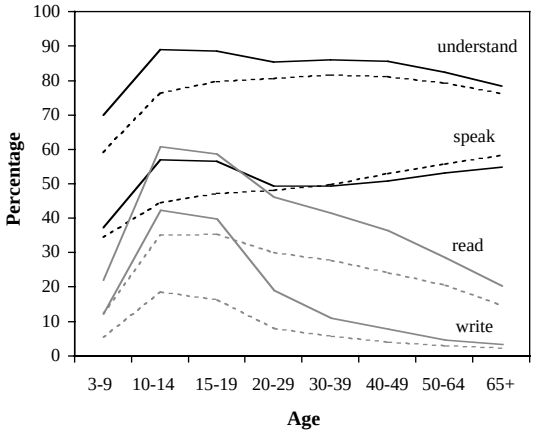

Figure 3. Competence in Valencian, by age. Dashed lines – 1986, solid lines – 1991.

some extent, that for understanding, reflect the steady attrition in Catalan during the dictatorship. The 1986 curves for reading and writing as well as all four curves for 1991 show distinct downturns after the 14–19 age group, which would have been the first age group to have benefited, at least for some period of time, from the Catalan courses instituted as a part of the linguistic normalization program. The continuing fall-off for reading and writing in the oldest age groups may be attributable to decreased levels of literacy. It is important to note that the diversity in competence levels is not only a function of age group, but also varies according to the origin (place of birth), social and political affiliations, and *comarca*. In our projections, we take into account the first and last of these.

Figure 4 shows that immigrants from Catalonia have competence levels higher than those of native Valencians, while those from the Balearic Islands have less speaking competence but claim more writing ability. Immigrants from non-Catalan-speaking origins of course tend to have lower scores on all four measures of competence, but it is of interest that while the Spanish immigrants can understand and speak more, foreign immigrants indicate that they can read and write more. Figure 10a in the Results section portrays the geographical distribution of oral competence in Catalan. The western and southern-most *comarcas* with lowest levels (least darkly shaded) are parts of the Community that have traditionally been Castilian-speaking. In addition, the southern *comarcas*, as well as those around the cities of Valencia and Alicante, are the major recipients of immigration.

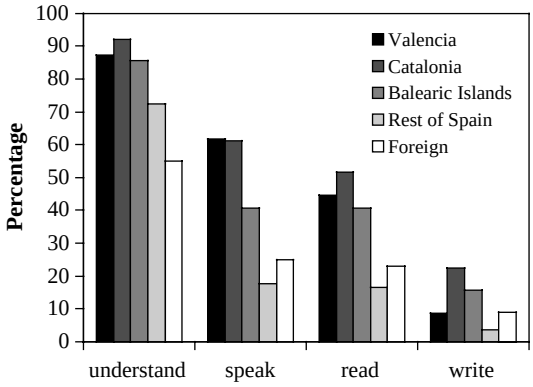

Figure 4. Distribution of competences in Valencian, by place of birth.

Models of Linguistic Processes

Our projections involve cohorts defined only by age (18 five-year groups), sex, *comarca* (34) and competence level (five: can write, can read but not write, can speak but not read, can understand but not speak, cannot understand), numbering $18 \times 2 \times 34 \times 5 = 6120$ in all. Demographic projections usually add or remove individuals from cohorts through birth, death or migration only. In adding the linguistic dimension, we must also allow for: (i) Assigning newborns to one of five linguistic categories, representing their predicted competence. In the next subsection, we model the linguistic transmission process in terms of the linguistic composition of the *comarca*, involving the prediction of the proportion of linguistically endogamous and exogamous marriages, and the effect on transmission rate of mother's and father's languages, and the relative density of the two languages in the *comarca*. (ii) Changing individuals from one linguistic category to another, representing the acquisition of language skills in school. This is not a monolingual context where acquisition may be modeled by a rapid increase from 0% to 100% within a few years for speaking and understanding and, in literate societies, from zero to the literacy level within a few more years for reading and writing. The situation here is far more complex, as is clear from Figure 3. First, the competence level in the initial age group varies greatly from *comarca* to *comarca*, depending on the rate of transmission. Second, childhood acquisition of understanding and speaking is not as rapid, as it is largely school-acquired and not home-acquired. Third, the peak competence level, achieved towards the end of high school, is generally far from 100%. Fourth, there is a drop in competence level for those born before the mid-1970s, who received at least part of their schooling during the period of linguistic repression. Fifth, reading and writing competence tails off for the older age groups, indicating a degree of illiteracy in earlier generations. The effects of schooling, which is of key importance in our projections, are modeled later in this section. (iii) Incorporating immigrants of various provenances into linguistic categories, representing whether they will be highly integrated linguistically or not. In the present version of our projections, we do not maintain separate cohorts according to origin. We do have statistics about how many immigrants of various origins acquire speaking and understanding of Catalan in each *comarca*. This will enable us below to model the integration process in terms of the density of Catalan speakers in the *comarca*. We assign immigrants to

linguistic cohorts according to their *comarca*, age and sex, in proportion to a pre-calculated integration rate. This does introduce a very small bias in the competence rates, since it condenses the time required for integration, whether one year, two years or five years, to effectively zero, but it is difficult to avoid this since we have no pertinent data on the dynamics of the integration process. Moreover it allows us to keep the number of cohorts in our projection to a manageable number, and to assign all newborns in each *comarca* in a uniform way, independent of the origins of their parents, only whether the parents are integrated or not.

Intergenerational Transmission

Whether a child speaks Catalan as a first language has been shown to depend on parents' languages and on the relative prevalence of Catalan and Castilian in the environment; Querol (2000: Chapter 6) reports the figures in Table 1 for a sample of high-school students representative of the Valencian Community. Based on these binomial data (e.g. 5 out of 213 children in the traditionally Catalan zone with two Castilian-speaking parents themselves speak Catalan as a first language), we carried out a maximum likelihood logistic regression of the rates of transmission

$$\log\left(\frac{p}{1-p}\right) = \mu + y_{\mathrm{m}} + y_{\mathrm{f}} + y_{\mathrm{z}} \tag{1}$$

where y_{m}, y_{f} and y_{z} represent the effects of mother's first language, father's first language and language zone, respectively. The values obtained are presented in Table 2.

TABLE 1
Child's first language by parents' first language and zone.

Mother–Father	Number of children			
	Traditionally Catalan zone		Traditionally Castilian zone	
	Catalan	Castilian	Catalan	Castilian
Catalan–Catalan	69	8	2	1
Catalan–Castilian	14	24	0	4
Castilian–Catalan	9	24	0	1
Castilian–Castilian	5	208	0	54

TABLE 2
Effects of parents and zone on transmission of Catalan.

Parameter	Catalan effect	Castilian effect
Mean μ	−1.04	
Mother y_m	0.99	−0.40
Father y_f	0.88	−0.33
Zone y_z	0.14	−0.79

Because we have no details of the sampling of the individual *comarcas* within each zone, we assume a linear dependence of transmission parameters on the level of knowledge of Catalan within the *comarca*. Then, instead of y_z for the entire zone, we have separate values y_i, satisfying

$$y_i = k_t \lambda_i + \omega \tag{2}$$

where λ_i is the proportion of the population in *comarca i* competent to speak Catalan, $k_t = 2.915$ and $\omega = -1.049$ are constants calculated by fitting the two data points (λ_C, y_C) and (λ_V, y_V) to Equation (2), where $y_C = -0.79$ and $y_V = 0.14$ are obtained from Table 2, and $\lambda_C = 0.089$ and $\lambda_V = 0.312$ are the overall proportions of Catalan speakers in the Castilian zone (11 *comarcas*) and the Catalan zone (23 *comarcas*), respectively (as calculated from the number of parents in each category in Table 1). In our projections we allow the *comarca* effect y_i to change over time as a function of λ_i, the proportion of Catalan speakers, while the overall effect μ and the parent effects y_m and y_f remain constant.

In order to incorporate transmission rates into our model, we need to input the proportion of marriages of each type (endogamous: Catalan–Catalan, Castilian–Castilian, and exogamous: Catalan–Castilian and Castilian–Catalan) in each *comarca*. Under a random marriage model we could predict these quantities to be x^2, $(1-x)^2$, $x(1-x)$ and $x(1-x)$ from knowledge of the proportion x of Catalan speakers in each *comarca*. Querol's sample as presented in Table 1, however, suggests a disproportionately high proportion of linguistically endogamous marriages compared to a random mating model.

Lacking any census reports on the linguistic composition of marriages, we can roughly estimate the effects of linguistic endogamy by setting up and analyzing a three-compartment model, which can then be applied to each *comarca*. The model consists of two exclusively endogamous subpopulations and one random marriage sub-

population of proportions θx, $\theta(1-x)$ and $(1-\theta)x + (1-\theta)(1-x)$, respectively, where x is the proportion of Catalan speakers in the *comarca* and θ is constant across all *comarcas*. This compartmentalized model is, of course, only a first-order correction to a pure random marriage model, but it is worthwhile to attempt some degree of correction since marriage type is so influential on transmission rate, and the simple adjustment suggested here is all that is warranted by the very limited data in Table 1. Solving this model for the proportions of the four marriages and for θ, and using the data for the entire Catalan zone in Table 1, we have found (Casesnoves Ferrer, 2003) that $\theta = 0.54$. Surprisingly, the very different data in the Castilian zone yields a similar result, $\theta = 0.5$.

Then in the projection exercises we undertake below, after the number of births is calculated in a *comarca*, the proportion of Catalan speakers in that *comarca* is used to predict the prevalence of each type of marriage. For each type, we use the logistic-linear equation (1) to predict the proportion of children who will turn out to be Catalan-speaking. These are weighted by the predicted prevalence of the marriage type in producing an overall transmission rate p_i for the *comarca*.

Acquisition through the School System

In a given *comarca*, let p_e be the proportion of speakers of age e with speaking competence in Catalan. In our school-based acquisition model, we assume that, until the age of 19,

$$w_e = \log\left(\frac{p_e}{1-p_e}\right) = w_i + (e-3)\alpha, \tag{3}$$

i.e., the logit of p_e is linear in age, where $w_3 = w_i$, the logit of the *comarca*-specific transmission probability p_i calculated above, and α is a universal acquisition rate.

Let u and v be the proportion of the 3–9-year-olds and 10–14-year-olds, respectively, with speaking competence in Catalan. Then, assuming the transmission rate w_i is relatively stable in these two cohorts, the acquisition rate can be estimated by

$$\hat{\alpha} = \frac{1}{6}\log\left(\frac{v(1-u)}{u(1-v)}\right). \tag{4}$$

As can be seen in Figure 5, this rate is fairly stable across the Community (mean 0.18, standard deviation of 0.055). We may also assume that it will not vary much over time. The transmission rate w_i,

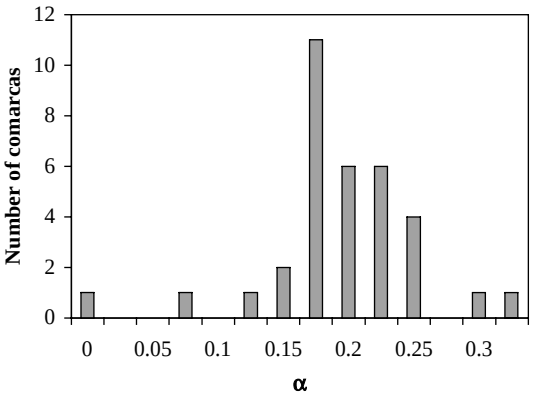

Figure 5. Distribution of acquisition rate across *comarcas*.

which varies from *comarca* to *comarca* could also be estimated from u and v according to this model, but this cannot be used as a constant in our projections since we can expect it to change as the overall rate of speaking competence evolves in each *comarca*. Thus in our projections, the transmission rates are updated on an annual basis according to our transmission model.

Integration of Immigrants

The SIES analysis of the 1991 census data distinguishes between the competence of native-born Valencians, those born in Catalonia or the Balearic Islands, immigrants from the rest of Spain, and those born overseas.

Figure 6 shows to what extent immigrants integrate to the Catalan-speaking community. Two observations may be made. First, immigrants from other regions of Spain are more resistant than immigrants of foreign origin. Second, integration may be modeled by a piecewise linear function of overall competence levels in the *comarcas*. Let x represent the proportion of all the individuals in the *comarca* who can speak Catalan and y the proportion of immigrants who can. Then in *comarcas* with less than 80% speaking competence,

$$y = 0.44x + 1.3 \tag{5}$$

for foreigners but only

$$y = 0.21x + 1.7 \tag{6}$$

for immigrant Spaniards. In *comarcas* with a very high concentration of Catalan speakers, however, the integration rate increases

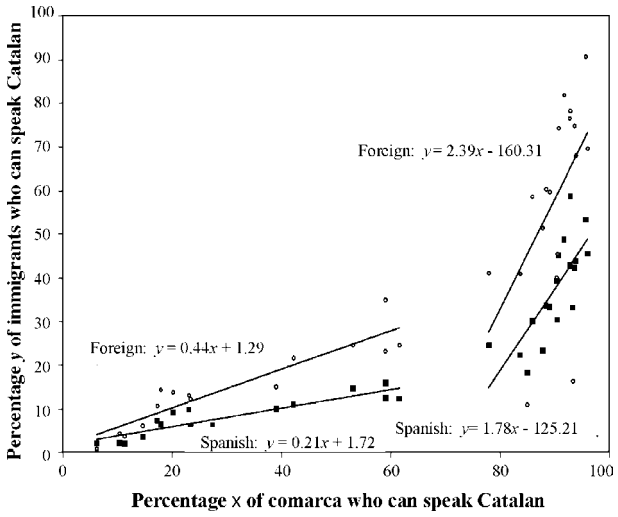

Figure 6. Integration of foreign and Spanish immigrants as a function of overall speaking competence level in the comarca.

very rapidly. Where there is more than 80% overall speaking competence,

$$y = 2.39x - 160.31$$
$$y = 1.78x - 125.21 \quad (7)$$

for foreigners and Spaniards, respectively.

As mentioned above, in our projections, we did not want to maintain separate cohorts for immigrants, because this would have multiplied the number of cohorts by a factor of four, incorporating the distinctions among the Valencian-born and immigrants classified by three types of origin, or even eight, since the children of immigrants may acquire Catalan in a manner different from their peers and their parents. Instead, we immediately assign all immigrants, in the year of their immigration, to Catalan-speaking or non-Catalan-speaking cohorts according to their *comarca*, age and sex, in proportion to the appropriate integration rate in Figure 6. This convenience comes at the cost of slightly inflating competence levels in *comarcas* with receiving very large numbers of immigrants.

The Program

We have implemented an experimental FORTRAN program in UNIX incorporating the models developed in the previous section into

demographic projections. At present this is specific to Valencian territorial divisions into *comarcas* and traditionally Catalan-speaking and Castilian-speaking regions, and to the structure of the data available on Valencian demography and linguistic competence levels. We intend to release a more general version for other operating systems, with an interface allowing flexibility in the input of territorial, demographic and linguistic information.

The Data

The population of each comarca in Valencia by sex and age group on January 1, 2000. Our source for recent population figures was: Padrón Municipal de Habitantes 2000, IVE (http://www.ive.infocentre.gva-es/padron/p2000).

Migratory movements in 2000. We included both internal (between *comarcas* in Valencia) and external migration (differentiating, only in the case of immigration, whether it comes from the other two Catalan-speaking communities, Catalonia and the Balearic Islands, other regions of Spain, or foreign sources) (source: Estadística de Variaciones Residenciales 2000, IVE [http://www.ive.infocentre.gva.es/vresi/vr2000]). These data are available broken down only by *comarca*, not by age or sex. To estimate the age and sex distribution, we could only obtain 1999 age and sex-specific immigration (broken down by origin) and emigration statistics pertinent to the Community as a whole, from the Instituto Nacional de Estadística (Estadística de Variaciones residenciales 1999, Inebase). We then broke down our year 2000 figures for each *comarca* in proportion to these Community-wide values.

Number of births by age of mother, and death by age and sex. We again used data for the Community as a whole (source: Movimiento Natural de la Población 1999, INE [http://www.ine.es.inebase]).

Competence in Catalan in 1991. This is the proportion of the population in each *comarca*, by age and geographic origin (Valencia, elsewhere in Spain and foreign), with the following abilities in Catalan: capacity to understand, speak, read and write (source: Servicio de Investigación y Estudios Sociolingüísticos de la Dirección General de Ordenación e Innovación Educativa y Política Lingüística, Consellería de Educación de la Generalitat Valenciana [http://www.gva.es]).

There are a number of limitations in our data. They come from a variety of sources, and are not all broken down by *comarca*, age and sex. In addition the competence data is only broken down into six age groups, so that a conversion routine is required to estimate corre-

sponding values for the 18 five-year groupings used for all the other data. Of course the major shortcoming is the outdated nature of the linguistic data.

The Projections

The population data were originally obtained as the number of individuals of each sex in each age group in each *comarca*. To break this number down into the linguistic cohorts necessary for initializing our projections, we used the 1991 percentages of each type of language competence level by age group and *comarca*. From these we derived the proportion of the age group who could write, read but not write, speak but not read, understand but not speak, and not understand Catalan. Each population cohort from the year 2000 data was then multiplied by these proportions to estimate the five demolinguistic cohorts. We used the same proportions for males as females, lacking data on different rates.

Birth and death rates were derived for the Community as a whole, based on 1999 numbers, divided by the January 2000 total population. Age and sex-specific migration numbers were estimated as discussed above. Raw numbers were retained for immigration coming from outside the community, while age and sex-specific rates were calculated for internal migration and emigration.

We used a time increment of one year. Births, deaths and migrations were calculated using the rates and numbers described above, and then one-fifth of each cohort was transferred to the (next older) cohort of the same sex, in the same *comarca*, with the same competence level.

The births were divided into six cohorts (no births entered the write or read cohorts) per *comarca*, using a male births proportion of 0.525 and a transmission rate for the *comarca* recalculated each year according to our transmission model. For the proportion of births not entering the 'can speak' cohorts, these were divided among 'can understand' and cannot understand' cohorts in equal proportions. We arrived at this division empirically, after studying curves like Figure 3 for each *comarca*.

For individuals between four and nineteen, some were transferred each year from the non-speaking cohorts to the speaking cohorts, using our acquisition model. Again the allocation of the transferred individuals among the cohorts 'speaks', 'reads' and 'writes' was determined empirically to be 20%, 10% and 70%, while these individuals were drawn from the corresponding non-speaking cohorts in proportion to the sizes of these cohorts.

Immigrants from outside Valencia were allocated to speaking or non-speaking cohorts according to our integration model, depending on the proportion of the *comarca* who can speak Catalan. For non-speaking immigrants younger than 20, this means a certain proportion will learn it later, possibly even to write or read, in school. The non-speaking immigrants are divided among 'can understand' and 'cannot understand' cohorts depending on the proportion of the *comarca* who can speak Catalan.

Results

Current Projections

Figure 7 depicts the evolution of linguistic competence in Valencia, under current demographic and linguistic processes, after 5, 10, 25 and 50 years, as indicated.

We note first that the irregular form of the year zero curves is largely due to the necessity of converting the data classified by the age groups in Figure 3 to the five-year groups used in our projections. All these graphs show, to a greater or lesser extent, the competing effects of schooling in increasing competence levels, and immigration in decreasing them. Since the former process is focused on children and young adults, whereas the latter affects all age groups and especially working-age adults, the curves evolve a characteristic shape where competence rises from a low level (determined by transmission rate) to a peak or plateau in early post-educational adulthood, followed by a steady drop reflecting the increasing proportion of immigrants with no training in Catalan in the older age groups. Because the normalized educational process has by now had its effect on an entire generation, the initial portion of the curves evolves rapidly into a steady state, but since the massive immigration from abroad is a very recent phenomenon, the shape of the curves at the older age groups continue to evolve over the fifty years of the projection, with the minimum competence attained at an increasingly advanced age. There is a contrast between competence in understanding and speaking on one hand, where competence levels in the oldest groups continue to decline over the projection period, and competence in reading and writing on the other, which actually increases steadily for all adult age groups.

The continuities evident between Figures 3 and 7 suggest that our projection procedures are adequately modeling ongoing processes of linguistic shift and normalization.

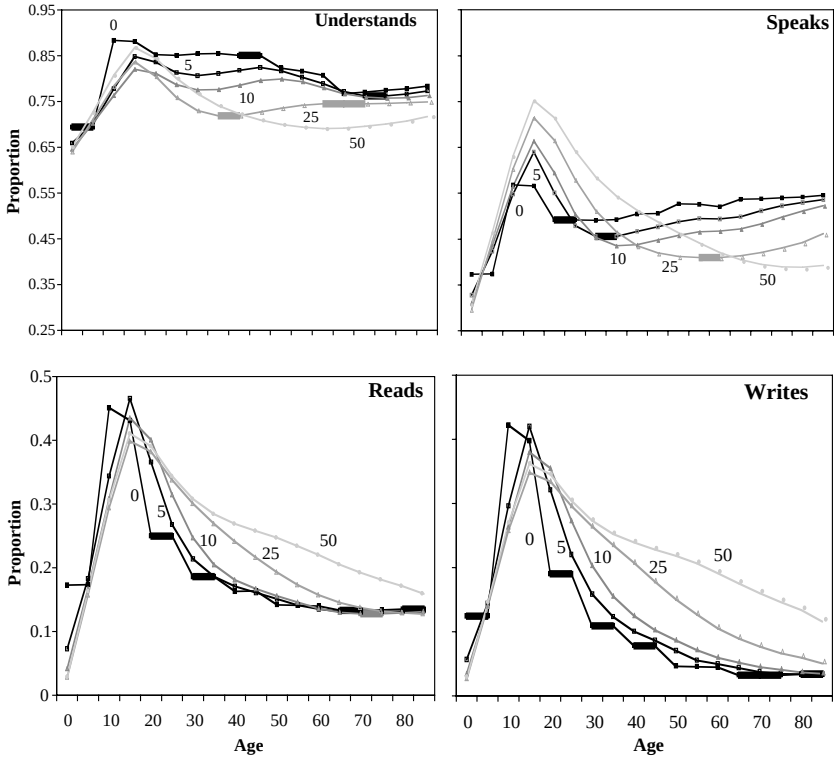

Figure 7. Projections of the proportion of the Community capable of understanding, speaking, reading and writing Valencian, after 5, 10, 25 and 50 years.

It is of interest that at present levels of fertility, density-specific intergenerational transmission, schooling, immigration and integration, the current proportion of speakers with speaking competence in Catalan (50.5%) will remain relatively stable over the 50 years projected, diminishing slightly to 50.1%. We refrain from advancing this as a prediction, because of the great instability in demographic factors we documented above, the great sensitivity of schooling and integration rates on unpredictable political and social trends, the out-of-date status of the linguistic competence data and the questionable pertinence of these data to actual use of Catalan. Thus while many of the speakers who are competent to speak Catalan actually do so in some non-school contexts, it is clear that the relationship between competence and use is becoming attenuated.

Projected Response to Perturbation in Individual Parameters

As a first step in evaluating the relative importance of the individual processes for competence levels, we undertake a series of sensitivity studies, perturbing one parameter at a time and observing the effect of this over time on competence. Some of the results are depicted in Figures 8a–d.

Figure 8a depicts changes in the projected proportion of the population with speaking competence over 50 years, as the schooling parameter α discussed above is modified. Figure 8b shows analogous results for the parameter μ used in our transmission model. Figure 8c treats the factor ϕ representing potential cutbacks in external immigration rates, and Figure 8d the consequences of changing integration rates. For $\gamma \leq 1$, the diminished integration rate is γy, where y is given by equations (5–7). For $\gamma \geq 1$, the increased integration rate is $(\gamma-1)x + [1-(\gamma-1)]y$, which may be thought of as representing a proportion of immigrants with rate determined by the equation $y = x$ rather than equations (5–7), while for the remaining proportion, the rate is still determined by the latter formulae. The predominance of the educational effect is clear in these studies. Not only does increasing the rate of acquisition have large long-term effects, but also

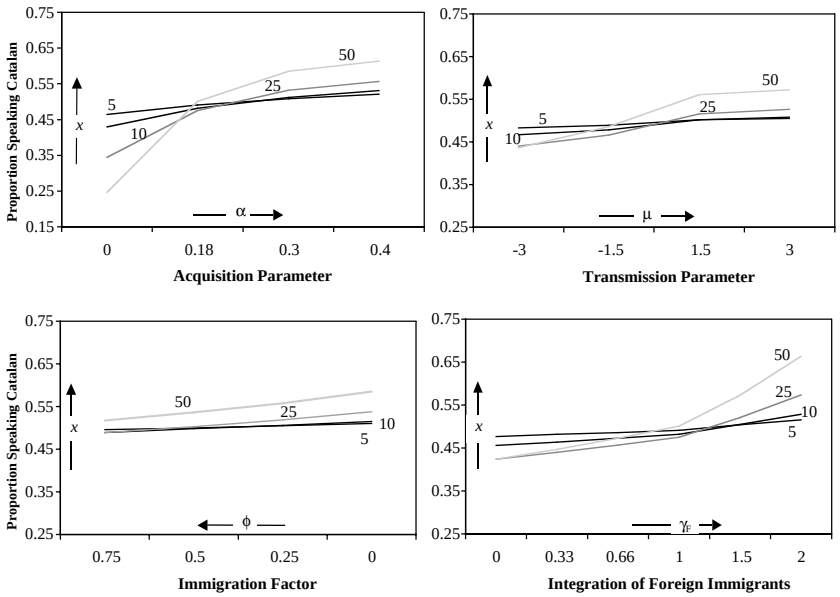

Figure 8. Effect of perturbations of individual parameter values on the projected proportion of Catalan speakers, after 5, 10, 25 and 50 years.

decreasing it results in a dramatic drop in the Catalan-speaking proportion of the population. This proportion is also sensitive to changes in both the transmission rate and the immigration rate, but to a much lesser extent, as long as acquisition in the schools continues to recruit both first-language Castilian speakers and children of immigrants of other origins into the Catalan-speaking cohorts. Note that increasing the integration rate has a very strong effect in bolstering the Catalan-speaking proportion, since it is equivalent to replacing immigration of both Spaniards and foreigners by a corresponding amount of population increase in each *comarca* with a level of competence approaching that of the Valencian-born residents.

Not shown in Figure 8 is the effect of increasing the birthrate. We examined this effect for increases up to 1965 levels, but this had virtually no effect on competence proportions.

The importance of the immigration and integration parameters are further illustrated in Figure 9, where the 50-year projection of the age-dependent speaking-competence profile under current parameter values is compared to the situations where no further foreign immigration is received and where, in addition, the Spanish immigrants are integrated to the same levels as Valencian-born residents. Up to now, we have discussed results at the level of the entire community. To what extent is the current heterogeneity of the Community maintained over 50 years according to our model. Briefly, the effects of universal education, internal migration and

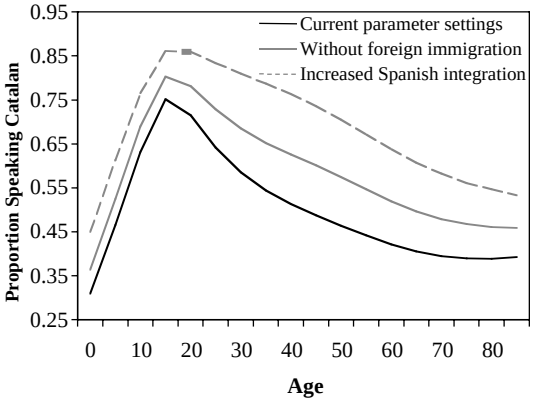

Figure 9. Projections under three hypotheses relative to immigration and integration: current trends, no external immigration, and no external immigration with maximal integration of Spaniards.

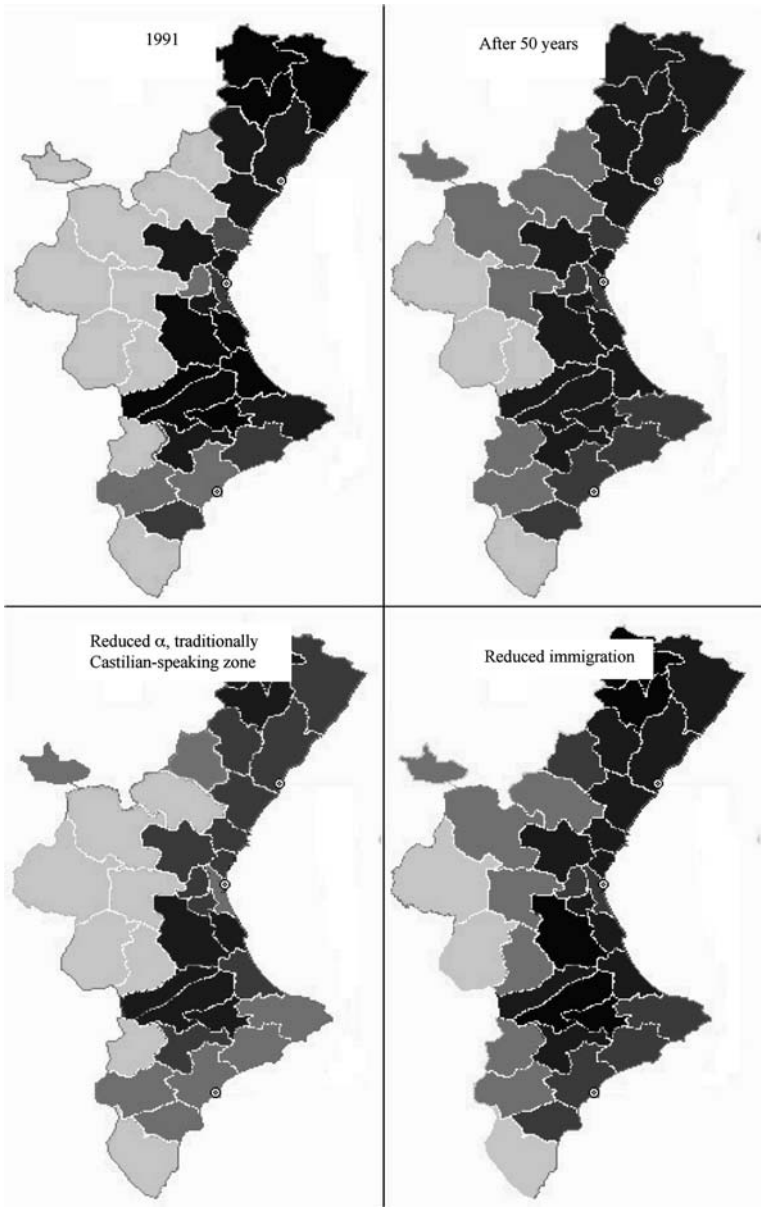

Figure 10. Geographical distribution of competence (a) in 1991, (b) after 50 years, current parameters (c) with reduced education parameter and (d) with reduced immigration.

immigration tend to homogenize the competence levels to some extent. In our estimates of α in the *comarcas*, we found no systematic difference between the traditionally Catalan-speaking *comar-*

cas and the traditionally Castilian-speaking ones, although the latter had consistently lower competence levels at all ages. Nevertheless, the educational requirements pertaining to the teaching of Catalan are absent or reduced in the Castilian speaking areas. This suggests that our model should be modified for these *comarcas*. Indeed, if we arbitrarily reduce α by a factor of 1/2 there, this maintains much of the linguistic distinctiveness of these *comarcas* over the 50-year projection. The maps in Figure 10 illustrate this, as well as the effect of reducing foreign immigration by a factor of 1/2. In all these maps, the darkest areas represent *comarcas* where speaking competence is 80% or more, and the four lighter shades of gray have thresholds of 60%, 40%, 20% and 0%, respectively.

Toward an Analytical Form for Demolinguistic Proportions

A Regression Experiment

We undertook a sampling experiment to assess the dependence of the competence proportions on all the independent parameters simultaneously. Over 1000 different scenarios were sampled according to a uniform distribution on the six-dimensional lattice defined by the parameter values in Table 3. For each scenario, we projected total population as well as proportion of Catalan speakers after 5, 10, 20, 30, 40 and 50 years. We ran multiple regression analyses of competence with forced entry of the six independent variables α, μ, ϕ, the integration factors γ_F, γ_S for foreign and

TABLE 3
Parameter values for the scenario sample.

Process	Parameter[a]	Values					
Education	α	0.00	0.09	0.18	0.27	0.36	
Transmission	μ	−3.12	−2.08	−1.04	0.00	1.04	
Immigration	ϕ	0.00	0.25	0.50	0.75	1.00	1.25
Integration	γ_F	0.00	0.50	1.00	1.25	1.50	1.75
Integration	γ_S	0.00	0.50	1.00	1.25	1.50	1.75
Fertility	v	0.00	0.25	0.50	0.75	1.00	

[a] Parameters μ and α as described in text for transmission and acquisition models. The factor ϕ multiplies external immigration rates (year 2000). Subscripts on γ indicate foreign (F) or Spanish (S) immigrants. For $0 \le \gamma \le 1$, the factor γ multiplies integration rate y as calculated in equations (5)–(7). For $1 \le \gamma$, the factor $\gamma-1$ indicates proportion of immigrants integrating at a maximal rate ($y = x$). For each female cohort, birthrate $= v \times 1965$ rate $+ (1-v) \times 2000$ rate.

Spanish immigrants, respectively, and the birthrate factor ν, testing for interactions among them and, given the curvilinear shape in Figure 8a, allowing both the linear and quadratic term α^2 for the effect of schooling. Schooling also showed long-term interaction with transmission, while an interaction between the immigration and integration parameters for foreigners completely absorbed the effect of integration, but not that of immigration, as could be expected. The final regressions were thus based on α, α^2, μ, $\alpha\mu$, ϕ, $\phi\gamma_F$, γ_S and ν. In the context of this study, change in birthrate has virtually no effect on the competence proportions, but we included ν anyway in our regression, both to illustrate its insignificance in comparison with the other parameters and for eventual comparison with other studies.

The regression coefficients for the standardized variables are given in Table 4. Figure 11 shows the evolution of these 'beta' coefficients in the equation over the term of the projection experiment. The most important results are the predominance of the schooling effect over all others, as well as the increasingly curvilinear dependence of competence levels on schooling, seen from the compensatory increase in linear and quadratic terms (cf. Figure 8). The immigration effects ϕ and $\phi\gamma_F$ are also important, affecting competence levels in opposite directions, particularly in the early years of the experiment. The transmission μ effect remains relatively stable. The increasing size of the schooling and transmission interaction parameter $\alpha\mu$ is a genuine effect, and not simply due to the lack of additivity of proportions in those scenarios where both of the corresponding linear parameters have high values, since repeating the entire analysis with the logistic of the competence level gives nearly identical values of all the beta.

The small magnitude of the birthrate effect is contrasted with its importance in an analysis of total population size in the same sampling experiment. Table 4 also gives the standardized coefficients for this regression. Since the relative sizes of the linguistically different age, sex and *comarca* cohorts, controlled by α, μ and the γ, do not affect total births, deaths or total migration out or into a *comarca*, these variables do not enter into the regression.

Note that our procedure for comparing effects of the different factors, while rigorous, admits to a degree of subjectivity at an early stage, namely with the construction of the six-dimensional space of parameter values. To some extent the upper and lower endpoints on each dimension are arbitrary, although chosen to represent the researchers' best guess at what may be socially, politically and his-

TABLE 4
Coefficients of (standardized) variables.

Variable	Parameter	Dependent variable (standardized)		
		Proportion of speakers	Same, changed initialization	Population size
		Multiple regression coefficients		
Education	α	1.43	1.28	
Transmission	μ	0.44	0.41	
Immigration	ϕ	−0.34	−0.23	0.88
Integration	γ_S	0.18	0.12	
Fertility	ν	0.03	0.05	0.46
Education × transmission	$\alpha\mu$	−0.21	−0.14	
Education × education	α^2	−0.68	−0.48	
Immigration × integration	$\phi\gamma_F$	0.26	0.18	

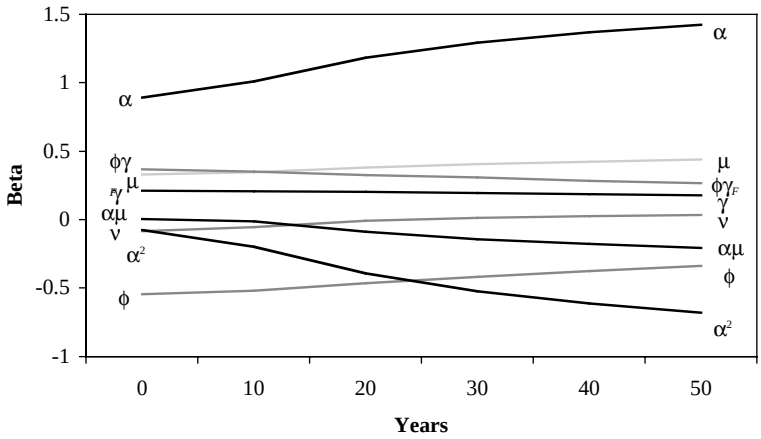

Figure 11. Evolution of effects over 50 years.

torically conceivable. Nevertheless, the arbitrariness of the range leads to some arbitrariness in the standard deviations used to standardize the variables in the regression analysis, and hence the magnitudes of the beta coefficients.

Stability of the Estimates

Projections of total populations are of course extremely dependent on the input structure of the population as well as of birth, death and

immigration rates. While we have used the latest compatible data available for these quantities, our major concern has not been to produce accurate projections of absolute population numbers, but rather to study proportions of the population at various competence levels. Since birth (and death) rates are assumed to be the same for all competence levels, inaccuracies and variation over time in these rates will tend to cancel out to a large degree when comparing different portions of the population at a given time. Moreover, given that the input competence data is severely outdated, it was not even our goal to accurately project the proportions at different competence levels. What we did set out to establish, and what we would like to validate, are the effects of the various demolinguistic parameters on the competence levels, and to what extent our estimates of these effects depend on the input data.

We undertook the following experiment. Starting with the original input data, we cut each cohort's understanding, speaking, reading or writing competence in half, and enlarged accordingly the corresponding no-competence cohort. We then selected a new sample of about 1000 scenarios, and repeated the projections, followed by the same regression analysis as above. Figure 12 shows that this large perturbation in initial values, while it predictably cuts the projected competence levels (e.g. the estimated constant in the unstandardized regression equation for the original experiment is 0.323 while in the replications with diminished input competences, it is 0.211), has almost no effect on the estimates of the parameter effects. This can also be seen in comparing the corresponding columns in Table 4. The only effect is a moderate shrinking of the absolute value of some of the estimates, but the approximate size and especially the relative importance of the effects remains unchanged. The compensatory changes in α and α^2 represent not a change in the importance of schooling but a reduced convexity of the effect.

Discussion and Conclusions

Our goal in this work was to develop a way to evaluate the relative importance of the demographic and linguistic factors that influence the progress of language shift and linguistic normalization processes. To accomplish this, we postulated a series of models for language transmission, acquisition and integration, whose parameters could be estimated from various data and which could then be incorporated into demographic projections. The design of the projections required

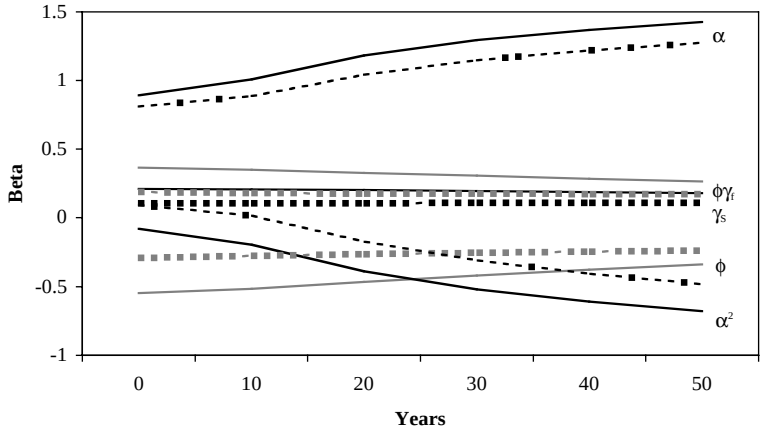

Figure 12. Replication where input competence levels are reduced by one half. Solid lines are the original effect curves; dotted lines are the results from the replication. Not shown: curves for μ, $\alpha\mu$, and v which are virtually identical in the original and the replication.

a trade-off between efficiency and the proliferation of cohorts subdivided by linguistic category.

In order to be able to summarize the interplay of the factors, we sampled a large number of scenarios from the six-dimensional space they define, and carried out a 50-year projection for each scenario. The results were then analyzed by multiple regression to derive an analytical expression for the proportion of Catalan speakers in the population as a function of the parameters.

The results underline the key role of the acquisition of Catalan in educational programs, and to a lesser extent, the flow of immigrants and the linguistic integration of immigrants. Birthrates have little impact, although this may be due to the uniform way this was modeled across the community. Transmission rate appears to have an ancillary role when compared to Catalan instruction in the schools, but this may be a misleading result due to our methodological focus on competence rather than usage.

Competence is a necessary condition for usage, but it is not sufficient. Indeed in the Valencian context, it is commonplace that the youth, who manifest the highest rates of competence to speak Catalan, actually use it the least. Thus, long-term projections about the maintenance or increase in competence may be masking an actual decrease in the vitality of the language. Furthermore, although competence is acquired in school, usage habits may be acquired in the

home, so that transmission rate would rank higher than school acquisition in projections of language use.

Scientifically then, it would be preferable if the dependent variable in our study were Catalan usage. Unfortunately, however, this is not feasible at present. First, in contrast to speaking competence, which may be coded 'yes' or 'no' for a given individual, language usage is context-dependent and variable, making it much harder to model and to predict. Moreover, there are very little data on usage, and even these consist of self-reports, which are notoriously less reliable than self-reports of competence, or of small-scale observational studies that are not statistically generalizable to entire populations. These problems are the major difficulties faced in our work and any attempts to predict language shift and normalization from language questionnaire data.

Even if we settle for competence projections for the time being, the fact that the data on this characteristic is over a dozen years old is a serious impediment to accurate prediction. We may only hope for the release of the 2001 Census data on this subject, and its pre-analysis by the SIES.

In the meantime there are a number of promising directions for the present work. Increased efficiency of the projection software should allow us to subdivide cohorts according to origin, leading to more detailed models of transmission, acquisition and integration. Given that the linguistic situations in the other Autonomous communities with 'co-official languages' have many elements in common with (and many different from) Valencia, and that much of the same kind of official data is available throughout Spain, a comparative study of the predicted output of all the programs of linguistic normalization becomes possible based on our software.

Acknowledgements

This work was made possible through the material support of the Centre de recherches mathématiques and the Institut Universitari de Lingüística Aplicada, and was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada. We thank Maria Teresa Turell for her advice and constant encouragement, and Charles Castonguay, Marc Termote, and Ernest Querol for helpful discussions. DS holds the Canada Research Chair in Mathematical Genomics and is a Fellow of the Canadian Institute for Advanced Research.

References

Casesnoves Ferrer, Raquel (2003). *Els efectes de la transmissió intergeneracional de la llengua, l'escolarització i la integració lingüística dels immigrants en les previsions de la competència en valencià al País Valencià*. [The effects of intergenerational language transmission, schooling and the linguistic integration of immigrants on projections of competence in Valencian in the Valencian Country.] Papers de l'IULA. Sèrie Monografies. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada

CCE (Conselleria de Cultura i Educació) (1989a). *Enquesta sobre coneixement i ús del valencià 1989*. [Survey of the knowledge and use of Valencian 1989.] Document. Direcció General de Política Lingüística, SIES. Valencia: CCE, Generalitat Valenciana.

CCE (Conselleria de Cultura i Educació) (1989b). *Coneixement del valencià. Anàlisi dels resultats del Padró Municipal d'Habitants de 1986*. [Knowledge of Valencian. Analysis of the results of the municipal census of residents of 1986.] Valencia: CCE, Generalitat Valenciana.

CCE (Conselleria de Cultura i Educació) (1992). *Enquesta sobre coneixement i ús del valencià 1992*. [Survey of the knowledge and use of Valencian 1992.] Document. Direcció General de Política Lingüística, SIES. Valencia: CCE, Generalitat Valenciana.

CCE (Conselleria de Cultura i Educació) (1995). *Enquesta sobre coneixement i ús del valencià 1995*. [Survey of the knowledge and use of Valencian 1995.] Document. Direcció General de Política Lingüística, SIES. Valencia: CCE, Generalitat Valenciana.

Pradilla, Miquel Àngel (2002). La política lingüística contemporània al País Valencià. [Contemporary linguistic politics in the Valencian Country.] *Treballs de Sociolingüística Catalana*, 16, 101–119.

Querol, Ernest (2000). *Els valencians i el valencià*. [The Valencians and Valencia.] Valencia: Denes 10.

Serra Yoldi, Inmaculada (2000). La inmigració estrangera: balanç i actituds. [Foreign immigration: overview and attitudes.] In Rafael Lluís Ninyoles (Ed). *La societat Valenciana: estructura social i instituciona* (pp. 35–63), Alzira: Bromera.

RAQUEL CASESNOVES FERRER
Centre de Recherches Mathématiques
Université de Montréal
Case Postale 6128, Succursale Centre-Ville
Montréal, Québec, Canada H3C 3J7
&
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra
La Rambla, 30–32, 08002 Barcelona, Spain

DAVID SANKOFF
Department of Mathematics and Statistics
University of Ottawa
585 King Edward Avenue, Ottawa, Canada, K1N 6N5
E-mail: sankoff@uottawa.ca