JASSS

Francesc S. Beltran, Salvador Herrando, Doris Ferreres, Marc-Antoni Adell, Violant Estreder and Marcos Ruiz-Soler (2009)

Forecasting a Language Shift Based on Cellular Automata

Journal of Artificial Societies and Social Simulation **12** (3) 5
<http://jasss.soc.surrey.ac.uk/12/3/5.html>

Received: 04-Nov-2008 Accepted: 08-May-2009 Published: 30-Jun-2009

Abstract

Language extinction as a consequence of language shifts is a widespread social phenomenon that affects several million people all over the world today. An important task for social sciences research should therefore be to gain an understanding of language shifts, especially as a way of forecasting the extinction or survival of threatened languages, i.e., determining whether or not the subordinate language will survive in communities with a dominant and a subordinate language. In general, modeling is usually a very difficult task in the social sciences, particularly when it comes to forecasting the values of variables. However, the cellular automata theory can help us overcome this traditional difficulty. The purpose of this article is to investigate language shifts in the speech behavior of individuals using the methodology of the cellular automata theory. The findings on the dynamics of social impacts in the field of social psychology and the empirical data from language surveys on the use of Catalan in Valencia allowed us to define a cellular automaton and carry out a set of simulations using that automaton. The simulation results highlighted the key factors in the progression or reversal of a language shift and the use of these factors allowed us to forecast the future of a threatened language in a bilingual community.

Keywords:Cellular Automata, Computational Simulations, Language, Social Dynamics

Introduction

1.1 When people talk about the languages of the past, such as Latin and Classical Greek, they often refer to them as *dead languages*. People refer to the languages that are currently spoken as *living languages*. This is probably because people often think of languages as living organisms: a language *is born, grows up* and finally *dies*. According to this biological metaphor, all languages will die sooner or later. However, these suppositions are not entirely true, because the fate of languages is not always death, but change. For example, in some sense, the speakers of Romance languages currently speak a form of Latin; not the Latin spoken in the Roman Republic in the first century BC, but a Latin that has changed continuously over the last twenty centuries.

1.2 However, it is also possible to find languages spoken in the past that did not produce a language spoken today, in the way that Latin produced today's Romance languages. These languages disappeared, i.e., they actually died out. Language death is not only an ancient event – as witnessed by the deaths of Etruscan, Egyptian or Hittite – but also a more recent one. The last speaker of Vegliot Dalmatian died in 1898, and languages still continue to die today. It is estimated that 90% of the approximately 6,000 languages currently spoken will have disappeared by the end of the 21st century (UNESCO 2003).

1.3 But if the *natural* process for a language is to change, why have some languages died out? There are two main reasons: (a) because the speakers disappeared due to an action such as direct genocide, or genocide through the destruction of their habitat or economic resources; or (b) because the speakers *decided* to speak a different language, i.e., due to a language shift (Mühlhäusler 1996).

1.4 Focusing on the second reason, a given language dies out as a result of a language shift when the community of speakers stop using their traditional language and speak a new one in all communication settings. The key factor for declaring that a given language becomes extinct is usage, not the linguistic competence of the speakers. If the speakers know a language but don't use it anymore, that language has become extinct regardless of their knowledge. Note that, for a language shift to occur, people must be fluent in at least two languages. A language shift is a phenomenon of bilingual and multilingual societies, not monolingual ones.

1.5 If there are two or more languages in a community, a hierarchical structure is always adopted, with one becoming the dominant language and the other the subordinate one. Both languages can coexist within such a hierarchy for long periods of time, but changes such as migrations,

political and economic events, etc., can break the equilibrium. In such cases, the pressure on the speakers of the subordinate language produces a change in their speech behavior. The speakers of the subordinate language may notice that their language has lost value relative to the dominant language. They may decide that it is no longer useful and stop speaking it in all domains of use.

1.6 The above shows that there are three phenomena involved in language death (Sasse 1992): (a) the *external setting*, i.e., the variety of cultural, historical, sociological and/or economic factors, which create pressure to abandon the language in the speakers' community; (b) the *speech behavior*, which includes the domains of use of the languages, attitudes towards the languages, etc.; and finally (c) the *structural consequences*, i.e., the linguistic changes observed in the morphology, phonology, syntaxes, etc. of the subordinate language. The three phenomena are interrelated: the pressure on the community created by the external setting compels the speakers to modify their speech behavior, which produces an impoverishment of the structure of the subordinate language. In this article we will focus on the speech behavior of the individuals.

1.7 Exhaustive empirical studies of language death are usually difficult to carry out because they take a long time, involve surveys and interviews with the speakers, etc. Although the process of language extinction is not usually exhaustively studied, the most complete studies (which examined two minority languages in Europe, a variety of Scottish Gaelic and an Albanian dialect spoken in Greece) made it possible to build up a complete model of a language shift (Sasse 1992). In reference to individual speech behavior, the Gaelic-Arvanitika model stated that a main factor in maintaining a language across generations is transmission within the family. If transmission of the language within the family fails, i.e., the parents speaks to their children in a language other than their own, the non-transmitted language will die in two generations. Note that the death of a language is not a slow process lasting several centuries, but a fast process of a few decades (see Figure 1).

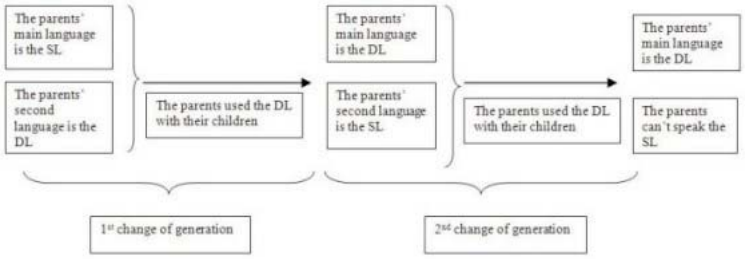

Figure 1. The interruption of language transmission according to the Gaelic-Arvanitika model of language shifts. Given a dominant language (DL) and a subordinate language (SL) in a community, the non-transmitted language (SL) become extinct after two generations.

1.8 A key issue in the study of language shift is to predict the future of threatened languages, i.e., to determine whether or not the subordinate language will become extinct in a given community with a dominant language and a subordinate language. If it is discovered that the subordinate language is threatened, language policies could be designed to reverse the language shift. Given that the external setting which triggers the language-shift process is usually very difficult to modify, the policies to reverse the language shift should mainly focus on individuals' speaking behavior. Thus, research on language shifts should concentrate on the factors involved in individuals' speaking behavior, and these factors should be used to forecast the extinction or survival of a subordinate language. Forecasting the values of variables is usually very difficult in the social sciences; however, the use of some mathematical tools such as the cellular automata theory (Hegselmann 1996;Hegselmann & Flache 1998;Nowak & Lewenstein 1996) and some methodological tools such as computer simulation (Gilbert 1996 , 2007;Goldspink 2002) helped us overcome this traditional difficulty. We used the cellular automata theory to study the speech behavior of individuals and forecast the language shift.

1.9 According to findings in the field of social psychology, individuals change or maintain their social behavior depending on their ability to maintain their behavior, physical immediacy and the number of neighbors influencing that behavior (Latané 1981;Nowak, Szamrez & Latané 1990). Similarly, we propose that given individuals will change their speech behavior in regard to the subordinate language if they are weakly engaged with it and/or a considerable number of their neighbors maintain a different speech behavior. Thus, a language shift is:

a. A local behavior in time and space: the decision to shift languages affects an individual or home unit at a given time.
b. An independent behavior: the external setting puts pressure on each individual or home unit to make the decision to shift languages, but this shift occurs without an explicit consensus with the members of the living community.
c. A mass behavior: a great number of individuals and home units make the decision to stop using their usual language and use the dominant language.
d. A parallel behavior: the individuals and home units make the decision to stop using

their usual language and use the dominant language at approximately the same time.

All these properties produce a self-organized emergent social phenomenon because there is no centralized unit[1] guiding the process and the overall result, i.e., the extinction of a language, is not explicit in individual behavior.

1.10 The behavior of a cellular automaton exhibits properties of localism, parallelism, emergence, etc., as occurs empirically during a language shift. The transition rules of a given cellular automaton are frequently simple, but their global behavior is analytically unpredictable, i.e., it is only possible to know the state of the cells in a given future time $t+k$ by running the automaton from $t = 0$ to $t = k$. Similarly, it is possible to assume that the language shift is regulated by a set of simple rules at the local level (the individual level) which produces global behavior at the social level. Thus, if it is possible to define the transition rules that describe the main features of a language shift, running the automaton will make it possible to predict the future of a subordinate language given different scenarios in the present. This article introduces the properties of a cellular automaton built to simulate the speech behavior of a given bilingual community, introduces an example of how the cellular automaton works using empirical data on the language shift and shows the preliminary results of the simulations using the automaton.

Model

2.1 Let's imagine a community that uses two languages, a dominant language (DL) and a subordinate language (SL). In our model, depending on the attitude towards the SL (i.e., the strength or weakness of individuals' engagement with the SL), the social pressure favoring the use of the DL and the number of neighbors engaged with the DL, the speech behavior of each person can be categorized as in one of three main states:

a. State 0 or monolingual state: The person only speaks the DL.
b. State 1 or bilingual state with preference for the DL: The person usually speaks the DL, but also speaks the SL, depending on the communication setting. The person transmits the DL to his or her children.
c. State 2 or bilingual state with preference for the SL: The person usually speaks the SL, but also speaks the DL, depending on the communication setting. The person transmits the SL to his or her children.

Given the fact that if there are two languages in a community a hierarchical structure is adopted, usually all people know the DL, but only a percentage of people know the SL. So, in reference to the knowledge of the languages, it is possible to find a percentage of monolinguals of the DL, but not monolinguals of the SL. (That is the case of the empirical example of language shift used to test the model in our simulation. See the next section *An empirical example of a language shift: Catalan in Valencia*)

2.2 Each state number indicates the level of engagement with the SL, from zero (0) to maximum strength (2). Furthermore, as stated above, the number of neighbors in each linguistic state also determines a given individual's use of the DL or the SL. Thus, a factor in determining the use of a given language is the number of interactions where it is possible to use that language. This includes *the submission rule*, a typical behavior of state-2 speakers, who tend to use the DL automatically when they address a DL speaker, even if the DL speaker is competent in the SL. (For a complete explanation of the submission rule, mathematical modeling and effects, seeMelià 2004).

2.3 To include the information about the speech behavior of individuals provided by the Gaelic-Arvanitika model, the definitions of states 1 and 2 include transmission of the DL or the SL to the next generation. Obviously, the speakers in state 0 transmit the DL to their children. The bilinguals transmit their preferred language to the next generation (the state 1 bilinguals transmit the DL and the state 2 bilinguals transmit the SL). Our cellular automaton does not include the *birth or death* of cells, but each cell inherits the transmitted language when the generation is renewed.

2.4 The community of our model is a discrete two-dimensional torus-shaped world. The world contains 105x64 cells, with each cell containing an individual. The amount data provided by the 6,720 cells facilitates to do both statistical descriptions and visual analysis on the computer screen. At each unit of time, a cell can only be classified in one of the three possible language states (0, 1 or 2), indicating the individual's strength in the use of the SL. Each cell has eight adjacent neighbors on the side and the vertex (a Moore neighborhood with a radius of 1) and the sum of neighbor values indicates the social pressure on the individual to use the DL or the SL. A low sum value means an individual has few opportunities to interact with his/her neighbors using the SL, but if the sum value increases, the individual's opportunities to interact using the SL also increase.

2.5 The transition rule determines the future state in $t+1$ of a given cell, which has a given state in t. The new state is determined by the sum of the neighborhood values, including the cell target, i.e., whether or not the sum surpasses a previously defined threshold. The sum can be a value between 0 (all cells in the neighborhood are classified in state 0) and 18 (all cells are classified in state 2). There are three thresholds:

a. S_a: a sum value below the threshold produces a sharp transition, i.e., state 2 changes sharply to state 0.
b. S_b: a sum value below the threshold produces a transition from a higher-value state to a lower-value state, but a sum value above the threshold produces a transition from a lower-value state to a higher-value state.
c. S_c: a sum value above the threshold produces a transition from a lower-value state to a higher-value state.

2.6 The threshold values should be $S_a < S_b < S_c$. The threshold values indicate the individual's level of engagement with the SL. When there is a greater level of engagement, the individual needs a lower threshold value to move up to a higher-value state. So the individual increases his/her usage and eventually the transmission of the SL with only a minimal number of current neighbors using the SL. Conversely, when there is a lower level of engagement, the individual needs a higher threshold value to move up to a higher-value state. So the individual decreases his/her usage and eventually the transmission of the SL if there are not a large number of current neighbors using the SL. The transition rule is described in detail in Table 1.

Table 1: The transition rule of the cellular automaton that simulates the language shift. The state of the cell at t changes at $t+1$ if the sum value of its neighbors surpasses the value of a threshold (S_a, S_b or S_c). Note that the transition from state 0 to state 2 is difficult to observe empirically, because it involves a non-subordinate-language speaker becoming a strong subordinate-language speaker.

	To state:			
		0	1	2
From state:				
	0	$\Sigma \leq S_b$	$\Sigma > S_b$	----
	1	$\Sigma < S_b$	$S_b \leq \Sigma \leq S_c$	$\Sigma > S_c$
	2	$\Sigma \leq S_a$	$S_a < \Sigma < S_b$	$\Sigma \geq S_b$

An empirical example of a language shift: Catalan in Valencia

2.7 To evaluate our model, we used empirical data on a current potential language shift. The availability of empirical data from recent language surveys in Valencia on the knowledge and use of Catalan prompted us to choose Catalan in Valencia as an empirical example with which to evaluate our model. Catalan is a Romance language currently spoken by approximately ten million people. Its area of influence extends along almost the entire Mediterranean Arc, which runs along the Mediterranean coast from Southern Spain to the South of France, and includes the Balearic Islands and the village of L'Alguer in Sardinia. This area is divided politically into four countries, Andorra, France, Italy and Spain, each of which grants a different official status to Catalan. Valencia, currently an autonomous region of Spain, belongs for the most part to the Catalan language area and Catalan is a joint official language. In Valencia, the subordination of Catalan to Spanish has proceeded at least for three centuries. Valencia is currently divided into two areas: a large Catalan-speaking area and a smaller Spanish-speaking area (see Figure 2). This and the coexistence of two languages in the Catalan area of Valencia have made Valencia a privileged arena in which to study the language shift.

2.8 In recent years several language surveys have been carried out in Valencia. The data obtained from these surveys do not show a clear trend of progression or reversal of the language shift, and different forecasts have been made without rigorous analysis, as influenced by the observer's point of view and interests. Therefore, furnishing the cellular automaton with empirical data from the language surveys in Valencia allowed us to highlight the ability of our cellular automaton to forecast the language shift in a more rigorous manner.

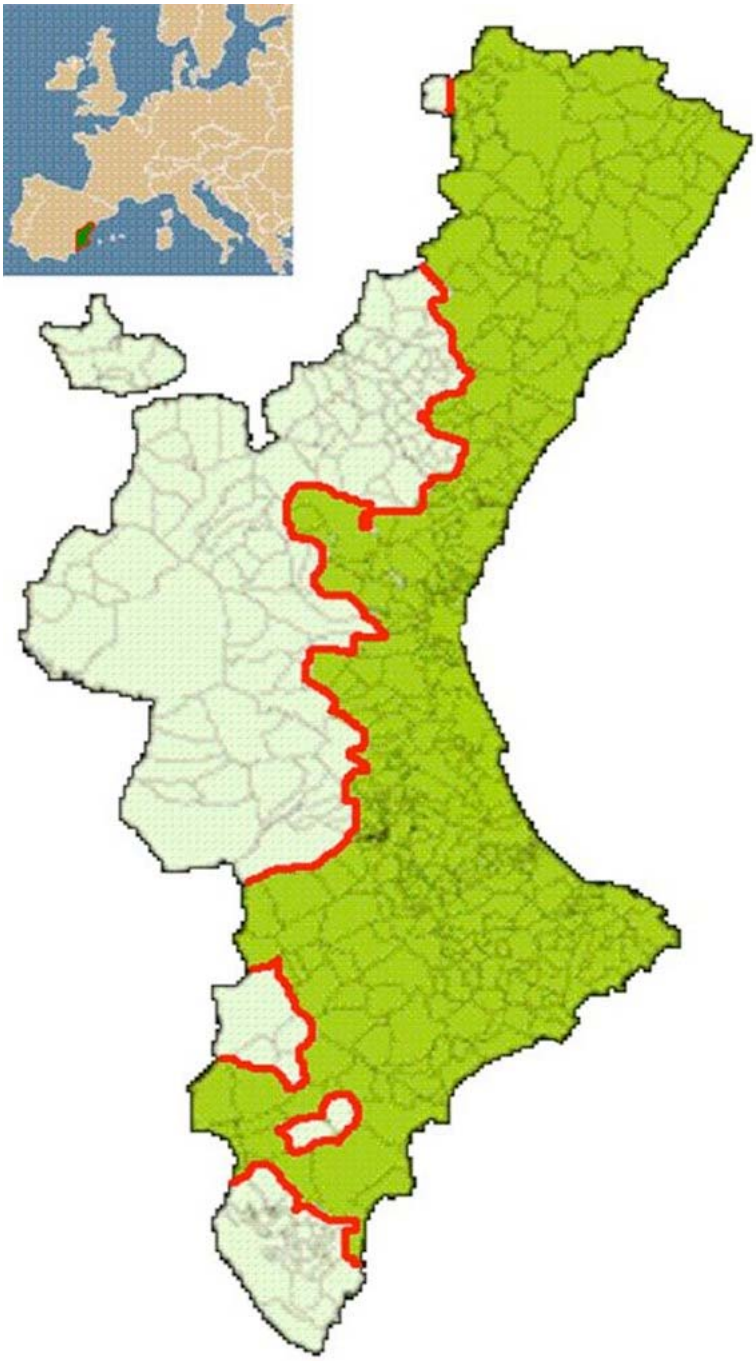

Figure 2. Language map of Valencia introducing the two language areas: the Catalan area (dark green) and the Spanish area (light green). The map in the top-left corner shows the location of Valencia in Europe.

2.9 We used empirical data from a recent language survey carried out by the Valencia government's *Servici d'Investigació i Estudis Sociolingüístics* (Bureau of Sociolinguistic Research and Studies) (Ninyoles 2005). The data from the survey were collected in 2005 from a sample of 6,600 people aged 15 and over. The survey was carried out in the two language areas of Valencia, Catalan and Spanish, and sought data on the knowledge and use of Catalan with regard to different variables such as age, gender, educational level, place of residence, etc. We used data from these language surveys as the initial values of the cellular automaton and tested different simulated scenarios to forecast the progression or reversal of the language shift in the future. The percentages for the use and knowledge of Catalan obtained in the survey gave us the number of cells containing each state at the beginning of the simulation ($t = 0$), and the variation of the threshold values gave us different scenarios of language engagement for the individuals.

Method

3.1 The cellular automaton was implemented using a Microsoft® Excel 2007 spreadsheet. We defined three sheets in an Excel book. The first sheet made it possible to visualize the cells and their states at each time unit. The state of the cell was shown by the color: white for state 0, orange for state 1 and green for state 2. The first sheet also displayed the frequency of the states at each time unit. Cell color and frequency values were updated at each time unit while the automaton was running. A second sheet made it possible for the user to define the

number of cells classified in each state at $t = 0$, the threshold values (S_a, S_b and S_c) and the number of simulations, given an initial number of states and threshold values. The number of cells classified in each state was determined by indicating the probability of each cell falling into one of the three states at $t = 0$. Finally, a third sheet summarized the final frequency of states for each simulation when the automaton became stabilized and also displayed the number of iterations required before stabilization.

3.2 When we defined the number of simulations, the automaton ran automatically and the data were displayed on the third sheet. But we could can also run the automaton step by step and observe the temporal evolution of the states of the automaton on the first sheet (see Figure 3). The initial states could be randomly seeded across the cells or defined by the user. In the latter case, it was possible to define cell sets of a given state. The Excel macros used to define the automaton and the main instructions to run it can be downloaded from http://www.ub.es/comporta/gcai.htm(go to *download* in the main menu).

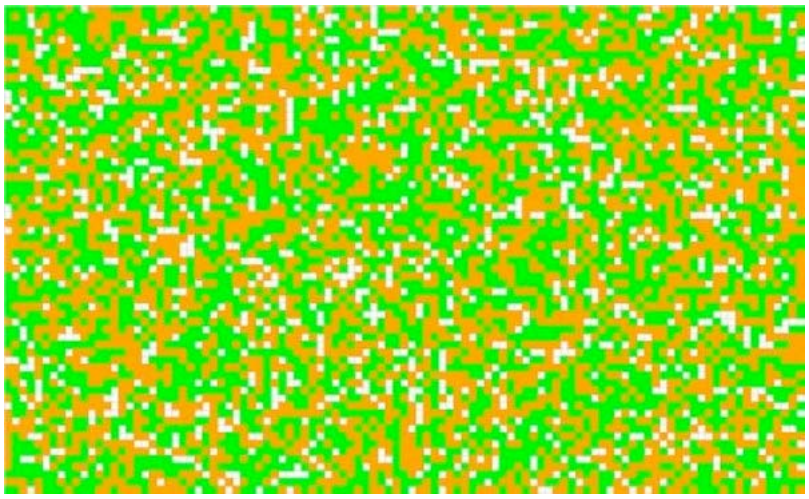

Figure 3a ($t = 0$)

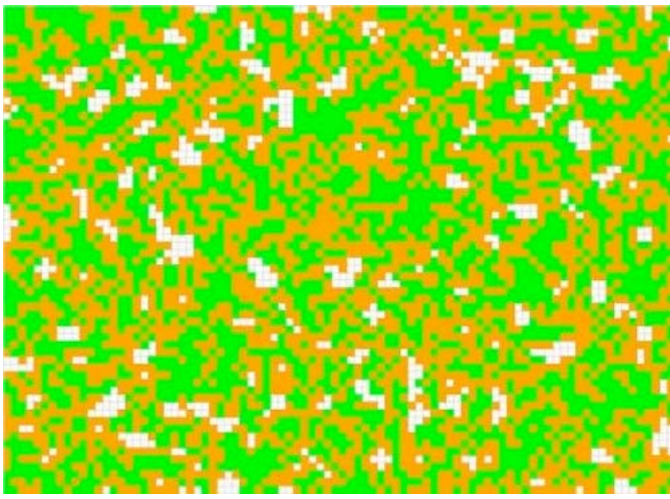

Figure 3b ($t = 1$)

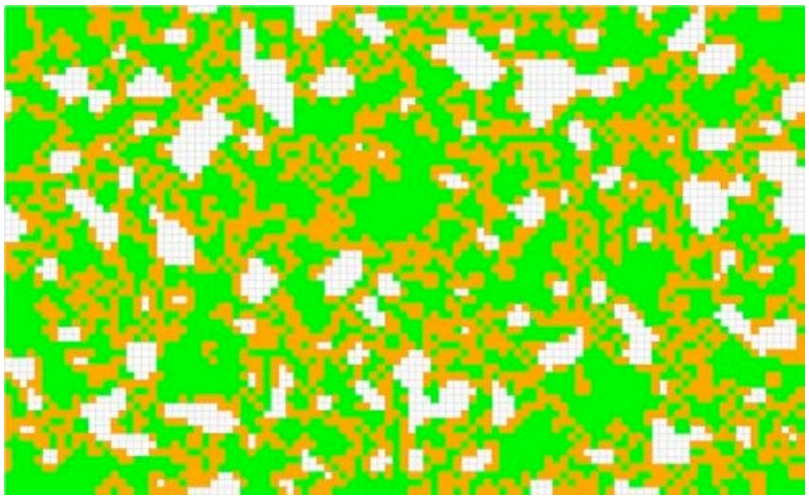

Figure 3c ($t = 4$)

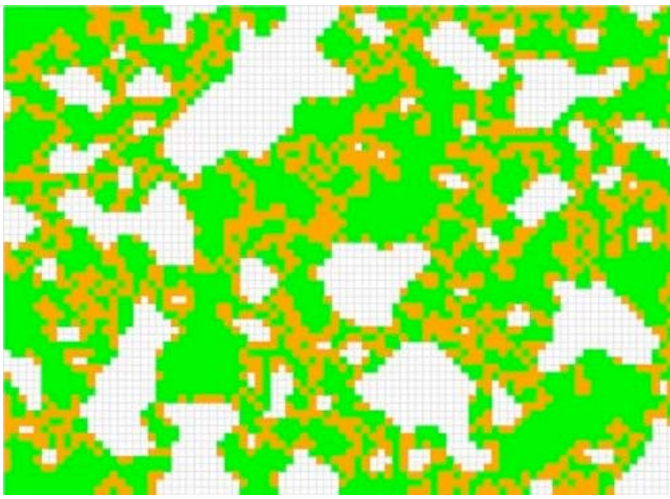

Figure 3d ($t = 56$, the automaton stabilizes)

Figure 3. An example of the dynamics of the cellular automaton (a to d) that simulate a language shift. The graphics were obtained from the first sheet of the sp which the cellular automaton was implemented. The white cells indicate state 0, the orange cells indicate state 1 and the green cells indicate state 2. At $t = 0$ the accounted for 15%, the state-1 cells represented 40%, and the state-2 cells accounted for 45%. The threshold values were set at $S_a = 3$, $S_b = 9$ and $S_c =$

Simulations 1 and 2. Exploring the effects of the initial size of the states and threshold values

3.3 In preliminary simulations, the values of the thresholds were varied and the initial size of the states was kept constant. The results showed the automaton's extreme sensitivity to variations in threshold Sb, compared with variations in thresholds S_a and S_c. We therefore focused on the effects of threshold S_b. We decided to explore the effects of threshold S_b exhaustively before entering the empirical data from the survey in Valencia. Thus, we made the first simulation with an initial percentage of states by setting 15% for state 0, 40% for state 1 and 45% for state 2. The size of the states was set arbitrarily, but we introduced a realistic linguistic context whereby states 1 and 2 were approximately the same size and state 0 was small but still of significant size. Thresholds S_a and S_c were also set to 3 and 15, respectively, and threshold S_b was varied across nine conditions (values 5 to 14). As the threshold values should be $S_a < S_b < S_c$, the values 3 and 15 allow to vary S_b across a wide range of values.

3.4 We made 150 simulations per condition and in each simulation we recorded the frequency of each state when the automaton stabilized. The criterion of stabilization was three steps

without change in the frequency of the three states. The states were randomly seeded in the cells of the automaton at $t = 0$. Although our automaton allows to define cell sets of a given state, we prefer to seed randomly the states because two main reasons: (a) we obtain information about the behavior of the automaton under some kind of base-line condition and (b) the Catalan and Spanish speakers in the Catalan-speaking area of Valencia were in fact very mixed, so a random seeding indicates the spatial distribution of the states in our empirical example.

3.5 The results showed that for S_b values 10 to 14, states 1 and 2 disappeared, i.e., the SL became extinct. But condition $S_b = 9$ resulted in the three states with the same percentage. The remaining S_b values reversed the situation and state 0 disappeared (states 1 and 2 had approximately the same percentage), i.e., the SL survived.

3.6 We performed a second simulation to determine whether or not the size of the states at $t = 0$ modified the results obtained in simulation 1. Thresholds S_a and S_c were set at 3 and 15, respectively, and threshold S_b was varied across the same conditions as simulation 1 (5 to 14). The initial percentage of the states was the same for all three states. We also carried out 150 simulations per condition and recorded the frequency of each state when the automaton stabilized. The criterion of stabilization was also three steps without change in the frequency of the three states, and the states were randomly seeded at $t = 0$.

3.7 The results showed the same trend observed in the first simulation: after a given value of S_b, the extinction of the SL was reversed. The change in the initial percentage of states in reference to simulation 1 produced a displacement of the reversal point from threshold $S_b = 9$ to threshold $S_b = 8$. Moreover, although state 0 disappeared after the reversal, state 1 was twice the size of state 2.

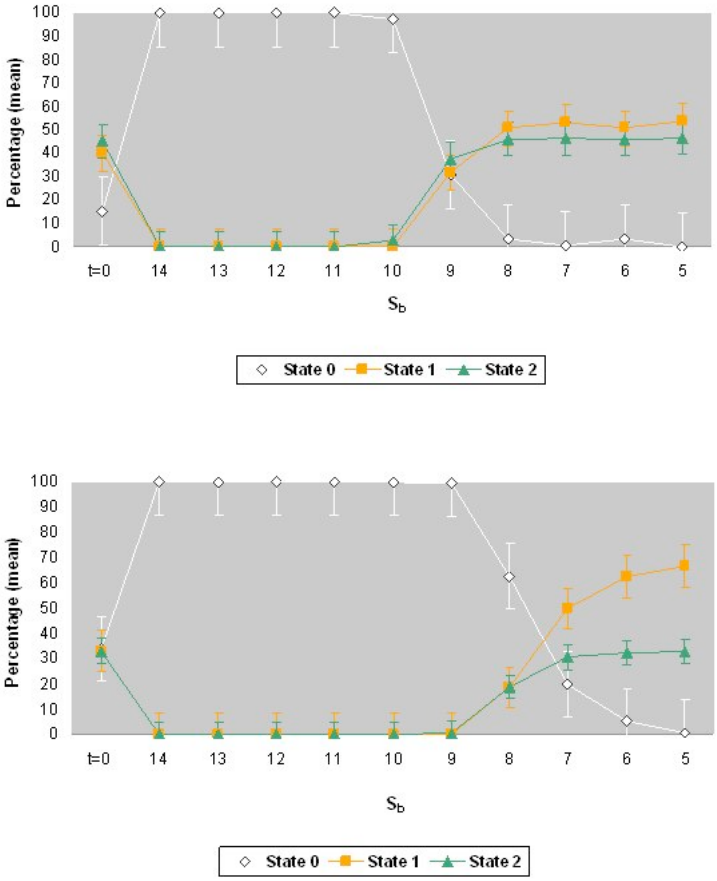

Figure 4. The mean percentages and standard deviations of states 0, 1 and 2 for the values of threshold S_b when the automaton stabilized (values of $S_b = 5$ to 14). The top graph shows the results for different percentages of the states (simulation 1) and the bottom graph shows the results for the same percentage of the states (simulation 2). The percentage of initial values ($t = 0$) is also shown.

3.8 Simulations 1 and 2 showed that below a given S_b threshold, state 0 disappeared and the SL survived, but above a certain S_b threshold, states 1 and 2 disappeared and the SL consequently became extinct. Although the effect remained the same, the initial percentage of the states produced differences in the percentages of states 1 and 2 obtained when the automaton was stabilized (see Figure 4). Simulations 1 and 2 showed the automaton's sensitivity to the size of the states at $t = 0$ and the value of threshold S_b. The next simulations used the values empirically collected from the language survey in Valencia as the initial values of the automaton, i.e., the size of the states at $t = 0$, and varied the values of

threshold S_b.

Simulations 3 and 4. Differences between Valencia's two language areas

3.9 As explained above, Valencia is divided into two language areas. Part of the language survey carried out by Ninyoles (2005) gathered information on the oral comprehension of Catalan in the two language areas. This information is summarized in Table 2. Although, as mentioned above, the critical factor in declaring a language extinct is usage, not the linguistic competence of the speakers, the latter could be useful in determining the potential of the SL, because linguistic competence is a necessary premise for the use of a given language. Moreover, the results of the simulation using oral comprehension data for the initial values of the cellular automaton (simulations 3 and 4) can be compared with the results of the simulations about the data on usage (simulations 5 and 6).

Table 2: Percentage of oral comprehension of Catalan in the two language areas of Valencia. The data were obtained in 2005 from a sample of 6,600 people aged 15 and over (Ninyoles 2005). The table also includes the states of the cellular automaton assigned to each oral comprehension category.

	Catalan area	Spanish area	State of the automaton
Do Not Understand Catalan	4.5	18.7	0
Understand Some Catalan	17.3	41.8	1
Understand Catalan Well	21.8	21.0	2
Understand Catalan Very Well	56.3	18.3	2
Did Not Answer			

3.10 Using the data from the survey, we performed a third simulation with initial states set at 4.5% for state 0, 17.3% for state 1 and 78.1% for state 2, which represented the percentage of oral comprehension of Catalan in the Catalan area. The initial percentage of each state was obtained by converting the survey responses to the automaton states. Thus, "Do Not Understand Catalan" was assigned to state 0; "Understand Some Catalan" was assigned to state 1; and "Understand Catalan Well" and "Understand Catalan Very Well" were assigned to state 2. Thresholds S_a and S_c were set at 3 and 15, respectively, and threshold S_b was varied across six conditions, values 5 to 10. These conditions were chosen because simulations 1 and 2 showed that the variation of these values produced a variation in the forecast of a progression or reversal in the language shift. We carried out 150 simulations per condition and recorded the frequency of each state when the automaton stabilized. The criterion of stabilization was three steps without change in the frequency of the three states, and the states were randomly seeded at $t = 0$.

3.11 We also performed a fourth simulation by setting 18.7% for state 0, 41.8% for state 1 and 39.3% for state 2, which represented the approximate percentage of oral comprehension of Catalan in the Spanish area according to the data from the language survey. The system of converting the survey responses to the automaton states was the same one used to convert the data on the Catalan area in the third simulation. The values of thresholds S_a, S_b and S_c were also the same as in the third simulation. 150 simulations per condition were made, the criterion of stabilization was three steps without change in the frequency of the three states and the states were randomly seeded at $t = 0$.

3.12 The results showed different forecasts depending on the language area. The Catalan area showed no change in the states across the values of threshold S_b, but in the Spanish area the results showed a language shift reversal at the $S_b = 8$ threshold (see Figure 5). In accordance with the results of simulations 1 and 2, the results of simulations 3 and 4 confirmed that, in addition to the S_b threshold, the initial percentage of states (i.e., the initial percentage of different speaker categories – states 0, 1 and 2) was also critical in forecasting the extinction or survival of the SL.

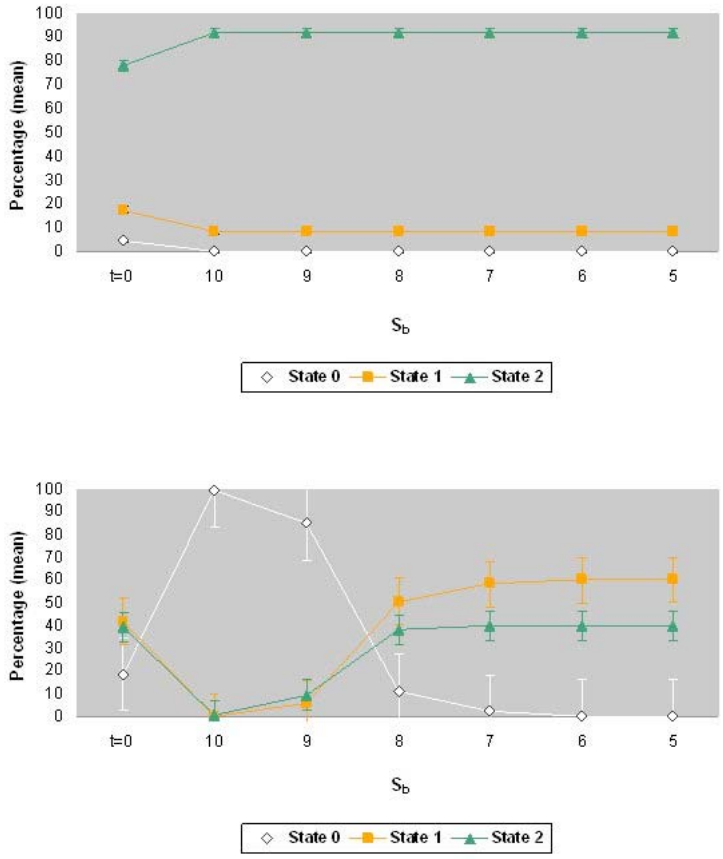

Figure 5. The mean percentages and standard deviation of states 0, 1 and 2 for the values of threshold Sb when the automaton stabilized (values of S_b= 5 to 10). The top graph shows the data on oral comprehension of Catalan in the Catalan area of Valencia (simulation 3) and the bottom graph shows the data on oral comprehension of Catalan in the Spanish area of Valencia (simulation 4). The percentage of initial values (t = 0) is also shown. If the standard deviations are very small, error bars are not displayed.

3.13 The size of the states at t = 0 in simulations 3 and 4 came from empirical data from the language survey, but the survey responses used as initial data in the automaton referred to people's knowledge of the language, not their speech behavior. However, as discussed above, the key indication for declaring a language extinct is usage, not linguistic competence. Hence, we made new simulations from the survey data using the responses on the use of Catalan.

Simulations 5 and 6. The use of Catalan in different social contexts

3.14 From the same survey used in simulations 3 and 4 (Ninyoles 2005), we obtained data on the use of Catalan in the Catalan area of Valencia in two contexts: the private context (at home) and the public context (with friends). These survey data are summarized in Table 3.

Table 3: Percentage of Catalan use in the Catalan area of Valencia in two social contexts. The data were obtained in 2005 from a sample of 6,600 people aged 15 years and over (Ninyoles 2005). The table also includes the states of the automaton assigned to each usage category.

Language spoken and frequency	Home	Friends	State of the automaton
Always Catalan	32.6	26.5	2
Frequently Catalan	2.5	4.0	2
More Catalan than Spanish	1.3	2.3	2
Equal Catalan and Spanish	6.2	13.8	2
More Spanish than Catalan	2.0	2.1	1
Frequently Spanish	4.4	4.9	1
Always Spanish	48.1	45.5	0
Other Language	1.4	---	---
Did Not Answer	1.1	0.9	---

3.15 We carried out a fifth simulation with the initial percentage of cells set at 48.1% for state 0, 6.4% for state 1 and 42.6% for state 2, which represented the percentages of the use of Catalan at home. The initial percentage of each state was obtained by converting the survey responses into the automaton states. Thus, "Always (speak) Spanish" was assigned to state 0;

"More Spanish than Catalan" and "Frequently Spanish" were assigned to state 1; and "Always (speak) Catalan", "Frequently Catalan", "More Catalan than Spanish" and "Equal Catalan and Spanish" were assigned to state 2. Thresholds S_a and S_c were set at 3 and 15, respectively, and threshold S_b was varied across five conditions (values 5 to 10). We made 150 simulations per condition and recorded the frequency of each state when the automaton stabilized. The criterion of stabilization was three steps without change in the frequency of the three states, and the states were randomly seeded at $t = 0$.

3.16 We also carried out a sixth simulation with initial state sizes set at 45.5% for state 0, 7% for state 1 and 46.6% for state 2, which represented the sizes of the use of Catalan with friends according to the language survey data. The system of converting the survey responses to the automaton states was the same one used to convert the data on the use of Catalan at home. The values of thresholds S_a, S_b and S_c were the same as in the simulation of Catalan at home. Once again, 150 simulations per condition were carried out, the criterion of stabilization was three steps without change in the frequency of the three states and the states were randomly seeded at $t = 0$.

3.17 The simulations showed the reversal of the language shift for a given value of threshold Sb in the two contexts: at home and with friends (see Figure 6). These results, compared with those obtained in simulation 3 in the Catalan area (Figure 5, top graph), showed a major difference in the forecast, depending on whether the person questioned understands Catalan or uses it effectively. In the same language area, the percentages for oral comprehension of Catalan and effective use revealed by the survey were very different. Given that a critical factor in obtaining a language shift in the automaton was the percentage of states at t =0, differences between the initial percentage of oral comprehension and usage also produced different forecasts in the automaton. In the case of oral comprehension, a language shift was not observed, but in the case of effective usage it was.

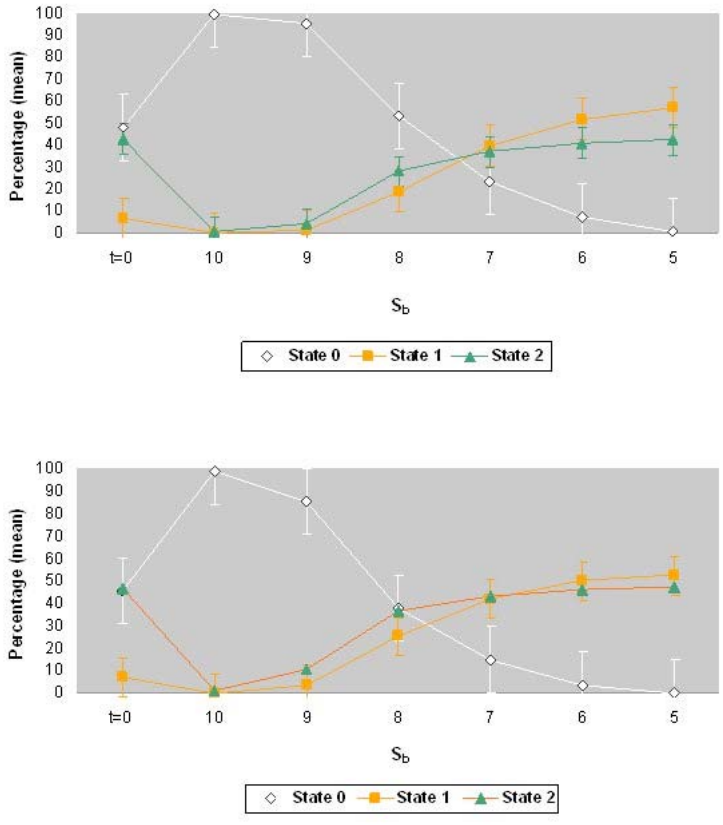

Figure 6. The mean percentages and standard deviations of states 0, 1 and 2 for the values of threshold S_b when the automaton stabilized (values of S_b= 5 to 10). The top graph shows the data on the use of Catalan at home in the Catalan area (simulation 5) and the bottom graph shows the data on the use of Catalan with friends in the same area (simulation 6). The percentage of initial values ($t = 0$) is also shown.

3.18 Also, given the fact that the initial size of the states was similar when the survey asked about the use of Catalan at home and with friends, the results were also very similar in both cases. Interaction with close friends was probably considered a private communicative setting instead of a public one by the people responding to the survey, so the speech behavior of the individuals was the same in both cases.

3.19 Simulations 3 and 5 showed differences between forecasts using data for oral comprehension and data for effective usage, as argued by the sociolinguistic studies. Furthermore, the Gaelic-Arvanitika model stated that the language shift happened over a short period of time,

namely two generations. The next simulations tested that statement by seeing if the cellular automaton could forecast the progress or reversal of the language shift within a small number of time units.

Simulations 7 and 8. The temporal dimension of the automaton

3.20 According to the Gaelic-Arvanitika model, when transmission of the SL at home stops, the language shift can occur in two generations. In all our previous simulations, the criterion for recording data was the stabilization of the automaton, which occurred at very disparate time units depending on the initial size of the states and the value of threshold S_b. We wondered if the automaton would show a clear trend after only a small number of time units, which would suggest a possible correlation between the time units of the automaton and chronological time at the empirical level. We replicated simulations 5 and 6, but ran the automaton step by step and recorded the frequency of the states after four time units, considering arbitrarily that each human generation corresponded to two of the automaton's time units.

3.21 Then we performed a seventh simulation and set the initial percentages of the states to 48.1% for state 0, 6.4% for state 1 and 42.6% for state 2, which represented the percentages for the use of Catalan at home according to the data of the language survey used as the initial values of the automaton in simulation 5. The system of converting the survey responses to the automaton states was the same one used for converting the data in simulation 5. Thresholds S_a, S_b and S_c were set at 3, 9 and 15, respectively. The value of 9 for threshold S_b was chosen because the previous simulations indicated that this value precedes the reversal of the language shift. A total of 10 simulations per condition were carried out and the states were randomly seeded at $t = 0$.

3.22 We also performed an eighth simulation and set the initial sizes of the states to 45.5% for state 0, 7% for state 1 and 46.6% for state 2, which represented the percentages for the use of Catalan with friends. The system of converting the survey responses to the automaton states was the same one used for converting the data on the use of Catalan at home. The values of thresholds S_a, S_b and S_c were the same as in the simulation of Catalan at home. Another 10 simulations per condition were carried out and the states were randomly seeded at $t = 0$.

3.23 The results indicated that after four time units, the percentage of the states showed a clear trend, making it unnecessary to wait for the stabilization of the automaton (see Figure 7). Therefore, the results suggested a possible correlation between the time units of the cellular automaton and chronological time.

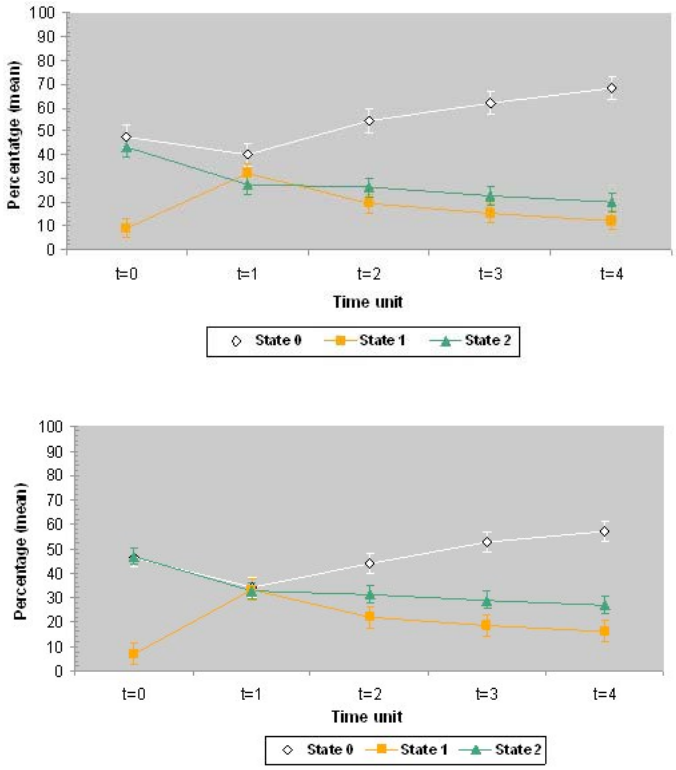

Figure 7. The mean percentages and standard deviations of states 0, 1 and 2 for the values of S_a=3, S_b=9 and S_c=15 after four time units. The top graph shows the data for the use of Catalan at home in the Catalan area of Valencia (simulation 7) and the bottom graph shows the data for the use of Catalan with friends in the same area (simulation 8). The percentage of initial values ($t = 0$) is also shown.

Discussion

4.1 The results of simulations 1 and 2 showed the importance of the initial values of the simulation and threshold S_b. When high values for S_b were set, states 1 and 2 completely disappeared, but when lower values were set, the trend of the results changed and state 0 disappeared, i.e., depending on the value of threshold S_b, the SL either died out (because states 1 and 2 had disappeared) or the SL survived.

4.2 Using the results of a language survey carried out in Valencia on the oral comprehension and effective use of Catalan to determine the initial size of the states, simulations 3 to 6 confirmed the results obtained in simulations 1 and 2, i.e., given an initial size of the states, the value of threshold S_b determined whether the SL died out or not. Furthermore, the difference in the results obtained in simulation 3 (data on oral comprehension) and simulation 5 (data on effective usage) confirmed that the data on the use of the SL were more significant than the data on the knowledge of the SL in terms of producing effective forecasts.

4.3 According to the results of simulations 1 to 6, given an initial size of the current speech behavior of the individuals (the automaton states), the value of threshold S_b became critical in explaining the dynamics observed in the simulation. A low value of S_b led the people in state 0 to become like the people in state 1, and the people in state 1 to become like the people in state 2. In other words, if, as stated above, the value of threshold S_b indicated the engagement of an individual with the SL, a higher level of engagement led the non-SL speakers to become bilingual, and the bilingual people that normally used the DL but also spoke the SL became bilingual people who usually spoke the SL. Also, transmission of the SL to the next generations increased because the number of bilingual people who transmitted the SL grew.

4.4 Simulations 7 and 8 showed that key information about the behavior of the automaton was given in the initial time units, i.e., it was not necessary to wait for the automaton to stabilize. These results agreed with the Gaelic-Arvanitika model, which anticipated a quick language shift where the future of the SL was decided in very few human generations, i.e., the four time units simulated in the automaton.

4.5 Finally, the results obtained by running the automaton provided some answers about the future of Catalan in Valencia, the empirical example of a language shift used to evaluate the model. It should first be pointed out that different forecasts were produced based on whether data for oral comprehension or effective usage of the language were used. In the first case, the automaton predicted that Catalan would survive, regardless of the value of threshold Sb (see the top graph in Figure 5). In other words, in this case it seems that nothing need be done to ensure Catalan survives in the Catalan language area. However, in the case of the data on the use of Catalan in the Catalan language area, the automaton predicted that the survival or extinction of Catalan depended on the threshold S_b value (see the bottom graph in Figure 5). Thus, because sociolinguistics tells us that the key to maintaining a language is its effective use, if we want to ensure that Catalan survives, it will be necessary to implement a language policy that favors speech behavior in accordance with a low value for threshold S_b. Although designing an appropriate policy to reverse a language shift is not an easy task (Fishman 1991), the main factors highlighted in the results of our simulations may help. Given a linguistic setting with a different level and number of speakers of the SL surrounding the individual, the individual's engagement with the SL becomes critical in determining his/her speech behavior with regard to the SL. Hence, a language policy should be developed to convince people to use Catalan even if there are few neighboring Catalan speakers. (Note that in the case of monolinguals who do not know Catalan, i.e., state 0 of the cellular automaton, the language policy should also provide opportunities to learn the language.)

4.6 In summary, modeling the speech behavior of individuals through the transition rules of a cellular automaton proved to be a useful tool to gain an understanding of a language shift, and provided a promising framework for future theoretical and empirical development of language shift studies. In the future, research on language shifts should focus on refining transition rules and applying them to different empirical settings to evaluate its predictive power. That is, the model should be calibrated by testing it with known language shift events (unfortunately, language shifts are currently a worldwide social phenomenon and it is easy to find a great many examples). Also, the research could incorporate a redefinition of the current neighborhood used in the simulations introducing connections between spatially distant cells. That would allow us to test the effects of the communicative encounters between local and more distant people (for example, between distant professional colleagues). Moreover, given the fact that in the current model the threshold values are the same for all individuals it would be of main interest to test the model under non homogeneous conditions of the threshold values. Finally, the time factor explored in this article in terms of the human generations needed to complete or reverse a language shift should also be investigated in order to demonstrate clearly a possible correlation between the time units of the automaton and the chronological change over generations.

4.7 Predicting the future of threatened languages can be a useful way to determine the use of language policies to reverse a language shift. As the example of Catalan in Valencia has illustrated, modeling the language shift and carrying out simulations based on cellular automata can highlight some relevant factors in the speech behavior of individuals with regard to the SL, which allows the future of the SL to be forecast in a given sociolinguistic

context.

Acknowledgements

This project was partially supported by a grant from the Directorate General for Research of the Government of Catalonia (2005SGR-0098). The authors would like to thank Sergi-David Diez for their assistance in running the simulations, and three anonymous reviewers for their comments on an earlier version of the manuscript.

Notes

[1] The external setting that triggers a shift from a subordinate language to a dominant one is usually a process guided by the dominant group or community, which puts pressure on the subordinate community, but the language shift itself is an autonomous individual *decision* made by the members of the subordinate community.

References

FISHMAN, J A (1991) *Reversing language shift. Theoretical and empirical foundations of assistance to threatened languages*. Clevedon, UK: Multilingual Matters.

GILBERT, G N (1996) Simulation as a research strategy. In Troitzsch K G, Mueller U, Gilbert G N and Doran J E (Eds.) *Social science microsimulation* (pp. 448-454), Berlin: Springer.

GILBERT, G N (2007) *Agent-based models*. Beverly Hills, CA: Sage.

GOLDSPINK, C (2002) Methodological implications of complex systems approaches to sociality: Simulation as a foundation for knowledge. *Journal of Artificial Societies and Social Simulation* 5 (1) 3 http://jasss.soc.surrey.ac.uk/5/1/3.html.

HEGSELMANN, R (1996) Understanding social dynamics: The cellular automata approach. In Troitzsch K G , Mueller U, Gilbert G N and Doran J E (Eds.) *Social science simulation* (pp. 282-306), Berlin: Springer.

HEGSELMANN, R and Flache, A (1998) Understanding complex social dynamics: A plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation* 1 (3) 1 http://jasss.soc.surrey.ac.uk/JASSS/1/3/1.html.

LATANÉ, B (1981) The psychology of social impact. *American Psychologist*, 36, pp. 343-365.

MELIÀ, J L (2004) Com es destrueix la llengua dels valencians: Un model binomial pels efectes de la regla de submissió lingüística [How the language of the Valencians is destroyed: A binomial model of the effects of the linguistic submission rule]. *Anuari de Psicologia de la Societat Valenciana de Psicologia*, 9 (1), pp. 55-68.

MÜHLHÄUSLER, P (1996) *Linguistic ecology. Language change and linguistic imperialism in the Pacific region*, London: Routledge.

NINYOLES, R L (2005) *Coneixement i ús social del valencià (síntesi de resultats)* [Knowledge and usage of Valencian (Summary of Results)], Servici d'Investigació i Estudis Sociolingüístics, Direcció General de Política Lingüística, Generalitat Valenciana, Valencia.

NOWAK, A and Lewenstein, M (1996) Modeling social change with cellular automata. In Hegselman R, Troizsch K and Muller U (Eds.) *Computer simulation from the philosophy of science point of view* (pp. 249-285), Dodrecht: Kluwer.

NOWAK, A, Szamrez, J and Latané, B (1990) From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97 (3), pp. 362-376.

SASSE, H-J (1992) Theory of language death. In Brezinger M (Ed.) *Language death. Factual and theoretical explorations with special reference to East Africa* (pp. 7-30), New York: Mouton de Gruyter.

UNESCO (2003) *Language vitality and endangerment*, Document by UNESCO Ad Hoc Expert Group on Endangered Languages, Paris.