Miguel, Amblard, Barceló & Madella (eds.) Advances in Computational Social Science and Social Simulation
Barcelona: Autònoma University of Barcelona, 2014, DDD repository <http://ddd.uab.cat/record/125597>

Modelling cultural shift: Application to language decline and extinction

Neus Isern
Laboratori d'Arqueologia Quantitativa (LAQU), Departament de Prehistòria, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain
Email: neus.isern@uab.cat

Joaquim Fort
Complex Systems Lab and Physics Department, University of Girona, C/. Mª Aurèlia Capmany 61, 17071 Girona, Catalonia, Spain
Email: joaquim.fort@udg.edu

Abstract—**Cultural shift is present in many aspects of human history. Here we present a model developed to study the particular case of language shift when a minority language is in competition with another language, which is perceived by the population as being socially and economically more advantageous (Isern and Fort, J. R. Soc. Interface 2014). We show that this model can describe satisfactorily the decline on the fraction of Welsh speakers over the last century. We also apply our language shift model as an interaction term into a reaction-diffusion equation and use it to predict the spread of retreat of the area of prevalence of the Welsh language. We find that the predictions are consistent with observational data.**

I. Introduction

THROUGHOUT human history we find numerous cases of cultural shift, where a society undergoes a change in some of its social or cultural traits, by adopting a more advantageous trait. Examples of cultural shift might include technical changes (such as the adoption of agriculture, or the industrialization process), changes in religious believes, or the adoption of a new language. Particularly, here we will focus on the study of local language shift during the last century, and the geographical retreat of the area of influence of minority languages being replaced by a new dominant language seen as more advantageous [1].

In general, cultural shift can be due to a local innovation (such as the evolution of Romance languages from Latin [2]) or due to the transmission of a cultural trait from a neighboring region [3]. In linguistics, the evolution of a new language is a process that takes thousands of years [4], whereas the acquisition of a neighboring language can happen at much shorter timescales [5]. In addition, nowadays improved communications and globalization processes have accelerated the processes of language replacement [6], [7], having a substantially negative impact in language diversity. Indeed, currently about 96% of the population speaks only about 4% of the languages in the world [8], and about a 90% of the present languages might become extinct, or in process of extinction, by the end of the century [9].

Because of the importance of the decline in language diversity, as well as the interest of the language shift dynamics by itself, in recent years, several studies have proposed different mathematical and computations models to describe the ongoing processes of language shift [10]. In 2003, Abrams and Strogatz [11] developed a simple two-population model to describe the competition for speakers between two languages, A and B, and which has been the basis for several other studies on linguistic shift [12], [13]. The dynamics of the competition is expressed in terms of the temporal change in the fraction of the population speaking each language as follows [11]

$$\begin{cases} \dfrac{dp_A}{dt} = \gamma\left(s p_A^{\alpha} p_B - (1-s)\, p_A p_B^{\alpha}\right), \\ \dfrac{dp_B}{dt} = -\gamma\left(s p_A^{\alpha} p_B - (1-s)\, p_A p_B^{\alpha}\right), \end{cases} \quad (1)$$

where p_A and p_B are the fractions of speakers of each language, with $p_A + p_B = 1$, γ is a parameter that scales time, $s \in (0,1)$ reflects the status of language A relative to B, and α determines the relative importance of the population fractions in attracting speakers to language A.

This model was applied to describe the decline of minority languages in competition with more advantageous languages. In particular, they applied the model to cases where a minority language is spoken regionally in a country where the official language a different language, which is seen as socially and economically more advantageous [11]. As a consequence, the minority language loses speakers on behalf of the other language, since the shift is seen as an advantageous strategy. In this context, p_A corresponds to monolingual speakers of the high-status (official) language, and p_B to fraction of the population that can speak the regional language (either as monolinguals or bilinguals). Therefore language A has the higher status, $s > 0.5$.

Here we want to model the same kind of situations, but with special attention to the geographical aspect of the language shift process. That is, we want to estimate the

This work was partially funded by the MINECO projects SimulPast (Consolider-CSD2010-00034, all authors), FIS-2009-13050 and FIS-2012-31307 (JF) and the Generalitat de Catalunya project 2014-SGR-36 (JF).

speed of advance of the border between linguistic areas, since not only the fraction of speakers of the high-status language increases over time, but also its area of influence expands. However, we will not develop our models here applying (1) because, even though Abrams and Strogatz obtained good fits for the periods they studied, (1) presents several problems when extrapolating beyond these periods. As an example, for the case of the Quechua language modelled in [11], (1) predicts that only when the fraction of speakers of the high-status language (Spanish) is higher than 25% will the language shift start; otherwise, it would be the Spanish speakers who would learn Quechua; a result that is in opposition to historical events [7], [9]. In contrast, for other languages, such as the Welsh language, (1) predicts that, for very low proportions of minority speakers, the languages shift process would be reversed, which is historically unrealistic (see a detailed mathematical discussion on these issues in [1]).

Therefore, here we present an alternative model to (1) that can also describe satisfactorily the historical data on the decline of endangered minority languages, but without the problems affecting (1) [1]. We use this new model as interaction term in a reaction-diffusion equation in order to explore the geographical aspect of language shift and to estimate the speed of retreat of the border between linguistic regions. We will apply our model to the case of the Welsh language, and see that we obtain estimates consistent with historical data.

II. MODEL

We want to define a model that can estimate the speed at which the border between two linguistic regions advances, when the speakers of one of the languages cease to speak it in favor of the neighbor language, which they perceive as having a higher status. We will first present an equation to describe the dynamics of the transfer of speakers from one language to the other, and then apply this equation as interaction term into a reaction-diffusion equation.

A. *Language shift model*

As in (1), in order to describe the dynamics of linguistic replacement, we divide the population into two groups —those who can speak the low-status languages (B), and those who cannot, and who can only speak the high-status language (A). Then, we propose to describe the temporal change of the fraction of speakers of each language, p_A and p_B, as follows [1]

$$\begin{cases} \dfrac{dp_A}{dt} = \gamma p_A^{\alpha} p_B^{\beta}, \\ \dfrac{dp_B}{dt} = -\gamma p_A^{\alpha} p_B^{\beta}, \end{cases} \tag{2}$$

where γ is a parameter that scales time, and the parameters $\alpha, \beta \geq 1$ are related to the attraction or perceived value of each language. Since $p_A, p_B \leq 1$, α and β may be regarded as a measure of the difficulty of language A to attract speakers (α), and the resistance of language B to loose speakers (β).

Note that, as opposed to (1), the model represented by (2) only allows for the speakers of the low-status language to change language, but not the reverse shift. This simplifies the equations, solves the problems and limitations from (1), and is a reasonable assumption in cases where no linguistic policies are applied (such as with the Welsh language up to the 1970s).

In general, in order to apply (2) as an interaction term into a reaction-diffusion equation, we need to rewrite it in terms of the population density (n_i), rather than the population fraction ($p_i = n_i/(n_A + n_B)$). Assuming that the total population ($n_A + n_B$) does not vary significantly over time (a realistic assumption for the Welsh during the period considered here [14]), the temporal variation of the population density speaking each language may be expressed as

$$\begin{cases} \dfrac{\partial n_A}{\partial t} = \dfrac{\gamma}{(n_A + n_B)^{\alpha+\beta-1}} n_A^{\alpha} n_B^{\beta}, \\ \dfrac{\partial n_B}{\partial t} = -\dfrac{\gamma}{(n_A + n_B)^{\alpha+\beta-1}} n_A^{\alpha} n_B^{\beta}. \end{cases} \tag{3}$$

B. *Reaction-diffusion model with language shift*

We now apply (3) as an interaction term in a reaction-diffusion equation that describes the temporal and spatial evolution of the population number density of speakers of each language, n_A and n_B, as follows [1]

$$\begin{cases} \dfrac{\partial n_A}{\partial t} = D\dfrac{\partial^2 n_A}{\partial x^2} + a n_A\left(1 - \dfrac{n_A + n_B}{K}\right) + \dfrac{\gamma n_A^{\alpha} n_B^{\beta}}{(n_A + n_B)^{\alpha+\beta-1}}, \\ \dfrac{\partial n_B}{\partial t} = D\dfrac{\partial^2 n_A}{\partial x^2} + a n_B\left(1 - \dfrac{n_A + n_B}{K}\right) - \dfrac{\gamma n_A^{\alpha} n_B^{\beta}}{(n_A + n_B)^{\alpha+\beta-1}}, \end{cases} \tag{4}$$

where D is the diffusion coefficient, a is the intrinsic growth rate, and K is the carrying capacity. In (4), we have chosen the local *x*-axes such that the linguistic front moves in the *x*-direction.

We can simplify (4) by assuming that, for modern cases of language replacement, the total population is fairly constant in time and nearing the carrying capacity. Then, if we apply that $n_A + n_B \approx K$ into (4), we find the following expression [1]

$$\begin{cases} \dfrac{\partial p_A}{\partial t} = D\dfrac{\partial^2 p_A}{\partial x^2} + \gamma p_A^{\alpha} p_B^{\beta}, \\ \dfrac{\partial p_B}{\partial t} = D\dfrac{\partial^2 p_B}{\partial x^2} - \gamma p_A^{\alpha} p_B^{\beta}, \end{cases} \tag{5}$$

that can be expressed again in terms of population fractions, which may now be defined as $p_i = n_i / K$ for $i = A, B$.

It is possible to find the speed of the linguistic front from (5) (or (4)) by applying numerical integration. We set the initial conditions so that, initially, the grid is divided in two sections —a range of x is occupied only by speakers of language A, and the rest by speakers of language B. Running the numerical integration, we obtain a moving front, and from its position at each time step we calculate the front speed.

It is not possible to find an analytic expression for the speed of the linguistic front from (5). However, we can derive an analytic range containing the real front speed by applying variational analysis. Following the method described by Benguria and Depassier [15], we can find the following lower bound for the front speed [1]

$$c_L = \sqrt{\gamma D} \max_{\delta\in(0,1)} 2\delta\sqrt{1-\delta}\,\frac{\Gamma\left(1+\frac{\beta}{2}\right)\Gamma\left(\frac{\alpha}{2}+\delta-\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}+\frac{\alpha}{2}+\frac{\beta}{2}+\delta\right)}, \quad (6)$$

where the gamma function is defined by the following integral $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$, for $x>0$ [16].

It is also possible to find an analytic upper bound following the variational analysis in [17], which is expressed as follows [1]

$$c_U = 2\sqrt{\gamma D}\sqrt{\sup_{p_A\in(0,1)}[\alpha p_A^{\alpha-1}(1-p_A)^{\beta} - \beta p_A^{\alpha}(1-p_A)^{\beta-1}]}. \quad (7)$$

The bounds are obtained from (6) and (7) by searching for the maximum result of the right-hand side expression for values of $\delta\in(0,1)$ for the lower bound, and for values of $p_A\in(0,1)$ for the upper bound. Besides the limitations explained above of the Abrams-Strogatz model (1), another problem with this model is that the variational analysis above does not work (because the interaction has two terms, and their difference can be either positive or negative, depending on the values of the population fractions).

III. RESULTS

Here we apply our spatial model with language shift, (4), to predict the speed of advance of the English linguistic front replacing Welsh in the UK. We will compare the results with the actual front speed, estimated from linguistic maps to be within the range $0.3-0.6$ km/yr [13].

We first find the values of the parameters α, β and γ from (2) that better fit the historical data on the decline of the Welsh language. We use data on the fraction of the population able (p_B) or unable (p_A) to speak Welsh in Monmouthshire for the period 1900–1980 (figure 1c in [11]). We use these data, rather than the data on all of Wales, also presented in [11], because Monmouthshire is a rather rural, area representative for most of the extension of Wales, and thus of the region where the front speed was estimated in [13]. The data from all of Wales, by contrast, contains data from the large agglomerations near Cardiff (about 50% of the population lives in 10% of the area of Wales), where the language shift dynamics may well differ from that on the rest of Wales.

Fig. 1 shows the data on the evolution of the fraction of Welsh speakers in Monmouthshire over time (squares), as well as the best fit obtained with (2) (line) . The parameter values yielding this best fit are $\alpha=2.23$, $\beta=1.76$ and $\gamma=0.237$. We can see in Figure 1 that we can obtain a very good fit of the data, which even improves the best fit that can be obtained with the Abrams-Strogatz model (1), since here we obtain a lower value of the sum of squared errors ($\chi^2_{Eq(2)}=1.41\cdot10^{-4}$ while $\chi^2_{Eq(1)}=1.46\cdot10^{-4}$). (Reference [1] shows how (2) can be satisfactorily applied to model also the decline of other languages.)

We now apply the parameters found above into our spatial model, (5), in order to find an estimation of the front speed both numerically and analytically. To do so we will consider two realistic values of the diffusion coefficient, $D=5.08$ km^2/yr and $D=6.72$ km^2/yr. Both are estimated from $D=\langle\Delta^2\rangle/4T$ [18] and using values of the generation time ($T=25$ yr [13]) corresponding to modern human populations. Then, the first value of the diffusion coefficient is estimated from mobility data on modern populations in the Parma Valley, Italy, during the twentieth century ($\langle\Delta^2\rangle=508$ km^2 [19], [20]), and thus coetaneous with the data in Fig. 1. The second value is estimated from mobility data in Catalonia, Spain, during the eighteenth and nineteenth centuries ($\langle\Delta^2\rangle=672$ km^2 [21]).

We present the predicted front speeds in Table I for the two values of the diffusion coefficient. The second column corresponds to the results of the numerical simulation, and thus, the exact front speed of the linguistic front for a system whose dynamics may be described by (4). Comparing these values with the speed range estimated from data, $0.3-0.6$ km/yr [13], we see then that we obtain good agreement between model and observations.

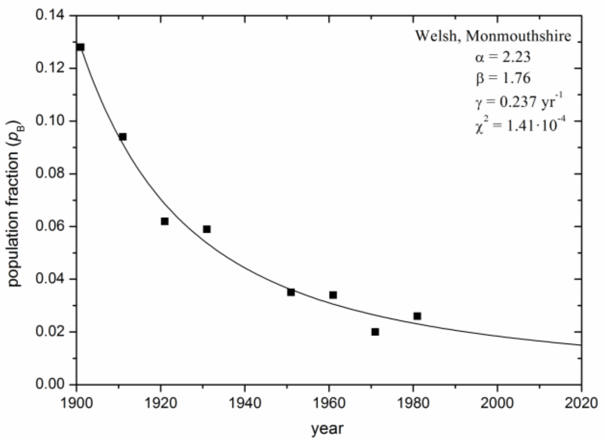

Fig. 1 Decline of the proportion of Welsh speakers over time (squares) and best fit (line) obtained with (2). Adapted from [1].

TABLE I.
NUMERICAL (c) AND ANALYTIC (c_L, c_U) PREDICTIONS OF THE ENGLISH LINGUISTIC FRONT REPLACING THE WELSH LANGUAGE

D **(km^2/yr)**	***c*** **(km/yr)**	**c_L (km/yr)**	**c_U (km/yr)**
5.08	0.557	0.356	0.934
6.72	0.641	0.409	1.750

In addition, the last two columns in Table I contain the values of the lower and upper analytic bounds calculated using (6) and (7), respectively. We see that, as expected, the exact solution lies within those bounds. But, what is more important, we see that the ranges obtained are also fairly consistent with the observed data values, and thus, we can use (6) and (7) to find a first approximation of the expected front of linguistic replacement without the need to apply numerical integration.

IV. CONCLUSIONS

We have presented a model that has been developed to describe the dynamics of language shift in a region where the speakers of a native language are under the influence of a neighbour language regarded as being socially and economically more advantageous [1]. Applying the model to the case of the Welsh language, we find that it is able to reproduce satisfactorily the decline in the fraction of Welsh speakers over time.

We have also applied our language shift model as an interaction term in a reaction-diffusion model in order to estimate the speed at speed at which the more advantageous languages spreads geographically, increasing its range of prevalence and, in consequence, diminishing the area of influence of the minority language. We have tested our model with historical data on the retreat of the area of influence of the Welsh language, obtaining a good agreement between model and observations.

In the context of present-day linguistics, our model can be used as a tool to assess how endangered a minority language is, and thus be able to design actions to control and/or reverse the destruction of language diversity. However, on a wider context, the model presented here could be applied to the study of other cases of cultural shift, present or past, where an advantageous cultural trait is spread, overcoming the prevalence of a local trait.

REFERENCES

[1] N. Isern, J. Fort, "Language extinction and linguistic fronts," *J. Roy. Soc. Interface,* vol. 11, paper 20140028, 2014.

[2] M. Metzeltin, *Las Lenguas románicas estándar: historia de su formación y de su uso*. Oviedo: Academia de la Llingua Asturiana, 2004.

[3] L. L. Cavalli-Sforza, P. Menozzi, and A. Piazza, *The History and Geography of Human Genes*. Princeton: Princeton University Press, 1994, pp. 102-103.

[4] C. Renfrew, *Archaeology and Language: The Puzzle of Indo-European origins*. London: Johnatan Cape, 1987.

[5] A. M. S. McMahon, *Understanding language change*. Cambridge: Cambridge University Press, 1994.

[6] J. Aitchinson, and H. Carter, *Language, economy and society: The changing fortunes of the Welsh language in the twentieth century.* Cardiff: University of Wales Press, 2000.

[7] A. Ortiz Rescaniere, *El quechua y el aymara*. Madrid: Editorial MAPFRE, 1992.

[8] D. Crystal, *Language death*. Cambridge: Cambridge University Press, 2000.

[9] M. Krauss, "The world's languages in crisis," *Language,* vol. 68, pp. 4-10, 1992.

[10] A. Kandler, "Demography and language competition," *Hum. Biol.*, vol. 81, pp. 181-210, 2009.

[11] D. M. Abrams, and S. H: Strogatz, "Modelling the dynamics of language death," *Nature*, vol. 424, p. 900, 2003.

[12] M. Patriarca, E. Heinsalu, "Influence of geography on language competition," *Physica A*, vol. 388, pp. 174-186, 2009.

[13] J. Fort, and J. Pérez-Losada, "Front speed of language replacement," *Hum. Biol.*, vol. 84, pp. 755-772, 2012.

[14] R. Chappell, *Focus on People and Migration: 2005*. Hampshire: Palgrave Macmillan, 2005. Available at: http://www.ons.gov.uk/ons/rel/fertility-analysis/focus-on-people-and-migration/december-2005/index.html Accessed 07 January 2014.

[15] R. D. Benguria, and M. C. Depassier, "Validity of the linear speed selection mechanism for fronts of the nonlinear diffusion equation," *Phys. Rev. Lett.*, vol. 73, pp. 2272-2274, 1994.

[16] R. S. Murray, and J. Liu, *Mathematical handbook of formulas and tables*, 2nd edn. New York: McGraw-Hill, 1999.

[17] R. D. Benguria, and M. C. Depassier, "Speed of fronts of generalized reaction-diffusion equations," *Phys. Rev. E*, vol. 57, pp. 6493-6496, 1998.

[18] J. Fort, and V. Méndez, "Time-delayed theory of the Neolithic transition in Europe," *Phys. Rev. Lett.*, vol. 82, pp. 867-870, 1999.

[19] N. Isern, J. Fort, and J. Pérez-Losada, "Realistic dispersion kernels applied to cohabitation reaction-dispersion equations," *J. Stat. Mech Theor. & Exp.*, paper P10012, 2008.

[20] L. L. Cavalli-Sforza, and W. F. Bodmer, *The genetics of human populations*. New York: Dover, 1999.

[21] M. Heras de Puig, *Biografía, ó, explicació del arbre geneològic de la descendencia de casa Heras de Adri, 1350-1850*. Girona: CCG Edicions, 2000.