Modelling cultural shift: application to processes of language displacement

Neus Isern and Joaquim Fort

ICREA/Complex Systems Lab and Physics Department, University of Girona
C/ Mª Aurèlia Capmany 61, 17071 Girona, Catalonia, Spain.
e-mail: neus.isern@udg.edu

Abstract Cultural shift is present in many aspects of human history, from the adoption of consecutive exploitation techniques, to the assimilation of a new language in a region. Here we focus on language shift, describing the main processes that have led human societies to start speaking a new language, and discussing several models devised to reproduce such processes. In particular we present a model developed to study the specific case of language shift in which the indigenous language in a region is being replaced by the language of an adjacent region, which is perceived by the population as being socially and economically more advantageous (Isern and Fort, J. R. Soc. Interface 2014). The model can predict the evolution of the fraction of speakers of each language over time, as well as the speed of at which the linguistic border advances into the region. The model is tested with modern data on the retreat of the Welsh language during the twentieth century providing results consistent with the observed historical data.

1 Cultural shift and language shift

Cultural shift, understood as the change in one or several cultural or social traits undergone by a human population, is present throughout human history. Indeed, we can find examples of cultural shift from the development and spread of tool construction techniques in prehistoric times, to the present time adoption of smartphones in our everyday lives.

The span of this definition of cultural shift comprises from technical changes (such as the examples above), to changes in religious beliefs or the adoption of a new language by a society. It comprises global changes with crucial impact on the evolution of human history—such as the Neolithic transition, which besides the technical changes directly related to the adoption of agriculture, entailed as well changes housing organization, social structures and belief systems that may be the initial seed of the present sociocultural organization (Smith 1995). But cultural shift can happen at any level, always playing a role in shaping today's and the future cultural characteristics and diversity of our societies.

Here we will focus on a specific kind of cultural shift: language shift. And in particular, below we present a model devised to be applied to processes of language shift where the dominant language in an area is replaced by a foreigner language (usually from a neighbouring area), that for some reason is seen by the locals as being more advantageous to the population (Isern and Fort 2014).

As with the broader concept of cultural shift, there have been ongoing processes of language shift since the emergence of the first spoken languages. This is clearly evidenced by the amount of dead languages from which we have written testimony (Ancient Greek, Goth, Hittite, Tocharian, etc.), and

the probably much larger number of extinct non-literate languages from which we have little to none information; as well as by the historic testament of how dominant regional languages have changed during the last few millennia. However, processes of language change are not assimilated as easily or as often as, for example, the adoption of a new pottery style. As opposed to most technological innovations, language is frequently an important component of ethnic identity in a group (Barth 1998, Crystal 2000). Therefore, even though a new language can sometimes be related to the adoption of a new culture complex, this is not always the case and, in general, linguistic substitution (and probably other changes related to ethnicity) is not as readily integrated.

However, linguistic change does happen, and in the following section we will detail the most important processes that yield to the replacement of the dominant language in an area, before discussing several models and presenting a language shift model that can be applied to predict the temporal and spatial evolution in processes of language displacement (Isern and Fort 2014).

2 Processes of language shift

The predominant language spoken in a certain region in the world will most probably have changed several times since the settlement of the first humans using verbal communication. But what are the underlying processes that lead to language change? In broad outline we can identify two basic processes: the local birth of a new language and the displacement of the local language by an extraneous language.

The birth of a new language is a slow process that usually includes several successive minor processes that spread throughout the population over the course of millennia (Renfrew 1987), until eventually the language has diverged enough from the original language as for them to be mutually unintelligible. These are often considered random processes, analogous to genetic drift, which may include the invention of new words—e.g., for innovations—, acquisition of loanwords from other languages in contact, phonetic changes—e.g., the use of occlusive sounds (*p*, *t*, *k*) in Latin, Greek and Sanskrit, as opposed to the fricative sounds (*f*, *th*, *h*) used in the Germanic languages, all of which share the same Indo-European origin (Lightfood 1999)—and ultimately the apparition of new grammatical forms (Renfrew 1987). A millennium would seem to be the minimum time span for this linguistic divergence to yield the birth of a new language (Cavalli-Sforza et al. 1994). Historical well known examples of language birth are the Romance languages from Latin, or the Indo-European languages from a probable common origin.

The other process of language shift, the displacement of the local language by a foreign one that becomes the new prominent language in the region, once started is usually a much faster process (McMahon 1994), and it can take place in as short a time as a single generation (Krauss1992). However, as mentioned above, language is often strongly related with ethnicity and not so readily assimilated as other changes, so what are the mechanisms that may trigger a language displacement process? Renfrew (1987) described three main mechanisms that would yield to language displacement: *demography/subsistence*, *élite dominance* and *system collapse*.

The first mechanism, denoted *demography/subsistence,* would entail the arrival of new population into the territory, bearing some new exploitative technology that would allow them to subsist at significantly higher densities, thus outnumbering the local population and prompting their language

to be the new dominant one. The spread of farming in Europe is known to have mostly taken place in such a way, and it may have well introduced their language into Europe alongside the farming technologies. Renfrew (1987) theorized that this language might be the Proto-Indo-European language from which most European and several Asian languages aroused. This theory has recently been supported by the results from a study that applied computational methods derived from evolutionary biology to infer an Indo-European language tree, as well as a chronology of the divergence times for every linguistic branch (Gray and Atkinson 2003). Their analysis estimated the initial Indo-European divergence to have taken place about 7800-9800 years BP, consistent with the initial spread of agriculture from Anatolia around 8000-9500 years BP (Gray and Atkinson 2003). Therefore, the spread of the Proto-Indo-European language into Europe would be an example of the *demography/subsistence* mechanism, and so would be some modern processes such as the spread of English (and its speakers) into Australia or North America—this last one including the use of the force of arms (Renfrew 1987).

The second mechanism, *élite dominance*, implies as well the arrival of foreign population but in this case, rather than introducing a new technology, the newcomers would be a reduced group with military superiority who would undertake the ruling of the region. This would lead to a period of bilingualism and the language of the élite may eventually become the dominant language. This is the case, for example, of the spread of the Latin language during the Roman empire; the Latin language was never imposed by the new Roman rulers, however, the language of the new rulers ended up being the dominant language at most of their empire (Rochette 2011). Nonetheless, *élite dominance* does not necessarily imply the displacement of the indigenous language, but it may be the new élite who end up being assimilated and their language forgotten (Renfrew 1987). This would be the case, for example, of the Norman invasion of Britain during the 11th century, where after a period of French dominance, English became again the dominant language at all social levels (especially reinforced by the posterior animosity between England and France), although with a clear French influence (Crystal 2003, Clairborne 1990).

Finally, the third main mechanism yielding to language displacement described by Renfrew (1987) was as a consequence of *system collapse*. In this case the language displacement would be the consequence of the collapse of a rapidly growing, highly specialized society whose central authority would not be able to maintain control in case of environmental adversities. The collapse could entail the exodus of local people as well as the loss of the control over the frontiers, thus losing terrain (politically and linguistically) to neighbouring better structured societies. And precisely for this reason, Cavalli-Sforza et al. (1994) englobed this mechanism and the *élite dominance* under a single denomination: *conquest by a minority*. The collapse of a previous system may have been the mechanism that yielded the Nahuatl language to become dominant with the Aztecs in Mexico (Renfrew 1987).

Besides these mechanisms, Cavalli-Sforza et al. (1994) additionally noted that language displacement processes may also be observed without the need of population movements introducing a new language. In this case, the new language would be acquired as a result of continuous contact with a neighbouring language, becoming the new main language near the border—we may call it *neighbouring acquisition*. As a consequence of this process the linguistic border would retreat, although without the need of political or military intervention, nor of population replacement. Over time, the new languages may completely displace the indigenous one and become the prevalent

language in the region. Cavalli-Sforza et al. (1994) observed this mechanism of language acquisition from a neighbour population with African pygmy people, and similar processes are observed nowadays in Europe with the shrinkage of the area of prevalence of several minority languages (Isern and Fort 2014).

Below we present a language competition model devised to predict the evolution of the number of speakers when an external language is displacing the native one. In the model, we are interested especially in language displacement processes which do not imply large movements of people. Therefore, this would mostly correspond to cases of language displacement due to the mechanisms of *élite dominance* or *neighbouring acquisition*, or a combination of both. In the next section we consider non-spatial models of such processes. In Sect. 4 we generalize them into spatial models (i.e., models of moving linguistic borders) to describe the case of neighbouring acquisition of language. They can be also useful to describe elite dominance, but in this case we note that (i) there is usually a military conquest first, after which the language substitution process takes place; (ii) the new language spreads possibly from one or several geographic centres of political power, rather than from an outside adjacent area.

3 Modelling language shift

The study of language evolution and language shift has become a field of interest for many disciplines in the recent times, including the application of computational and mathematical methods to model these processes.

A fruitful area for the application of computational methods to linguistics has been the study of the emergence of internal linguistic changes leading to language birth—lexical emergence and diffusion, phonetic change, appearance of grammar structures—through the application of approaches such as game theory or probabilistic inference (Wang et al. 2004, Baronchelli et al. 2008, Nowak et al. 2002, Bouchard-Côté et al. 2013). In general, many of these applications can be of interest both for the study of current language evolution as well as to the historical events of linguistic divergence and language birth, although several studies are particularly focused on historical linguistics. The work by Bouchard-Côté et al. (2013) is precisely devoted to the reconstruction of protolanguages by means of probabilistic inference of sound change over time, producing results very close to those obtained through manual reconstruction by linguists. Another important line of computational research, related to historical linguistics and process of language birth, has been the inference of language trees through the application phylogenetic methods. Such methods have been applied to infer linguistic relationships and to estimate chronologies of language divergences for the Celtic languages (Forster and Toth 2003), the Indo-European languages (Gray and Atkinson 2003), and even to the Eurasiatic macrofamily (Pagel et al. 2013).

Language displacement has also been an important topic of research both for computational and mathematical modelling, although the substantial difference in the mechanisms leading to the shift in language described above affect significantly the applied models. In the *demography/subsistence* method, the means to the language shift is mostly related to population displacement. Therefore, an appropriate modelling approach here would be the application of wave-of-advance models (Ammerman and Cavalli-Sforza 1973). Indeed, such models have been widely applied to the spread of the Neolithic in Europe (Ammerman and Cavalli-Sforza 1973, Fort and Méndez 1999, Fort 2012)—

which seems to have been related to the expansion of the Proto-Indo-European language (Renfrew 1987, Gray and Atkinson 2003). However, the Neolithic expansion in Europe was not a purely demic process in the whole continent, but in some regions agriculture was also transmitted by an acculturation process (Fort 2012). In those regions language may have or may have not been transmitted alongside agriculture—Refrew (1987) suggests that the assumed pre-Indo-European languages such as Basque or the now extinct Etruscan, may have survived in those processes of cultural transmission. But if in some areas language was indeed transmitted as well, it must have been part of the whole "Neolithic package," rather than a simple linguistic shift, and thus the process may be well described by demic-cultural models devised to describe the Neolithic transition as a whole (Fort 2012). The colonization of North America—with the subsequent spread of the English language—has also been successfully described with a wave-of-advance model modified to include the colonizing intent (Fort and Pujol 2007).

Language displacement may also take place in situations where there is little or no population movement. These would be the cases for the mechanisms of *élite dominance* and *neighbouring acquisition*. Although the reasons behind the language shift differ, after the new élite is established in the first mechanisms, both cases can be assumed mostly equivalent for modelling purposes. Both mechanisms lead to a competition for dominance between two languages, one of which having a higher status—either because it is the language of the élite, or because the neighbouring language is seen as more advantageous for some reason—within a population that is mostly unchanged. Of course, a linguistic imposition in the case of *élite dominance* may accelerate the displacement process, or on the contrary, it may cause a resistance effect giving a higher status to the indigenous language; such effects can be easily included into a language competition model by means of adapting the parameter values.

Therefore, language competition models are a good approach to model language displacement when there are no important changes in the population density due to immigration or to the increase in the population density that is sustainable. In the recent years, several authors have developed mathematical and computational language competition models (for a review, see Kandler 2009). In 2003, Abrams and Strogatz (2003) developed a simple two-population model to describe the competition for speakers between two languages, A and B, coexisting in the same region, and which has been the basis for several other studies on linguistic shift (Patriarca and Heinsalu 2009, Fort and Pérez-Losada 2012). This model describes the evolution in time of the fraction of speakers of each language (p_A and p_B), with the fraction of speakers defined as the ratio between the number of speaker of a given language over the total population (e.g., $p_A = N_A / N = N_A / (N_A + N_B)$, and therefore $p_A + p_B = 1$). The evolution over time, represented by the time derivative, is expressed mathematically according to the following equations (Abrams and Strogatz 2003),

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

In broad lines, this equation shows that the evolution of the fraction of speakers of each language follows the same dynamics as the other one, thought with an opposed sign; this means that the speakers lost by one language become speakers of the other one. In addition, the minus sign before $(1-s)$ indicates that a language may lose or gain speakers depending on the fraction of the population speaking each language, as well as the values of the of the parameters. In the model by Abrams and Strogratz (2003), γ is a parameter that scales time, so it accelerates or decelerates the process; s, with a value between 0 and 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.

Although this model has been applied rather successfully to describe language evolution (Abrams and Strogatz 2003), it yields some problems when trying to extrapolate the model beyond the data over which they applied the model. From a mathematical point of view, these problems arise because of the existence of stable and unstable equilibrium points, depending on the parameter values (see a detailed mathematical discussion in Isern and Fort 2014). To put it in more general terms, we shall describe one of the possible problematic outcomes. Depending on the parameter values chosen, the language with a higher status displaces the other one until it is nearly extinct in the region, and then the process stops. This means that the model predicts that, without adding any extra particularity (such as part of the population living in a very secluded area), the language will remain alive as the main language for a reduced part of the population forever.

Such behaviour is historically unrealistic, and for this reason we opt for an approach conceptually simpler and which does not present the same extrapolating problems (Fort and Isern 2014). This alternative model also describes the dynamics of the transfer of speakers between two languages A and B in competition, one of which is seen by the population as being socially or economically more advantageous. This model is described as follows (Fort and Isern 2014)

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

As with the model by Abrams and Strogatz (2003), the temporal evolution of the population fraction (described mathematically as a derivative), depends on the population fraction speaking each language, p_A and p_B, and the values of three parameters. γ is a parameter that scales time, so it accelerates or decelerates the process. 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, again, the speakers lost by language B become speakers of language A (both equations have the same form but with a minus sign in the second equation). However, with this model, only one of the languages can gain speakers and the other loses them; in particular, A is the language seen as more advantageous and thus gaining speakers and displacing language B. In general, this is a reasonable simplification for processes where a foreign language is displacing the indigenous language in a given region. It is true that it cannot directly describe all historical situations, such as

the case of the Norman invasion of Britain, where the French initially gained speakers, but the English language eventually recuperated its prevalence (Clairborne 1990); though neither can it be directly described by the model in Eq. 1. A reasonable alternative would be to divide the whole period into two subperiods, each with a different language defined as the high status one; after all, the status of the language is defined by the subjective perception of it by the population, rather than its political position.

Therefore, the new model in Eq. 2 is conceptually a reasonable approach to model processes of language displacement when a language is perceived by the population as being more advantageous socially or economically. Such cases could be related to processes of language shift due to *élite dominance* or *neighbouring acquisition* or combinations of both (for example, when the new élite come from an adjacent region, which may describe the situation of minority languages in current times when they are co-official in their territory, or even not officially recognized).

The process of language displacement after a situation of *system collapse* mentioned in the previous section may take place in many different ways, with or without population movement, with a change in the dominant élite, etc. and for this reason we may not propose specific models of application in such event.

4 Moving linguistic borders

From a geographical point of view, linguistic displacement can also take place in different ways. For example, when an incoming new élite takes the ruling power of the region, a possible pattern may be the apparition of several language shift sources—e.g., near government, education or religious emplacements. On the other hand, when the language shift mechanism implies population displacement (due to immigration and a subsequent rapid growth), there will be a moving linguistic border between the two languages that will understandably progress with the incoming population; usually such situation can be seen as an advancing front driven by the population growth and thus can be easily described through wave-of-advance models. However we can also have a moving linguistic frontier without the need of an immigrant population front. This is the case when the new language is introduced from a neighbouring region.

We can mathematically model the progress of a linguistic frontier over time and space, when the displacement mechanism is due to language acquisition rather than population substitution, with a reaction-diffusion model similar the wave-of-advance models; that is, a model where the population dynamics is simplified to short-range migration (e.g., due to marriage), and increase or decrease in the population number is due to factors such as population growth or conversion into another population group. However, as opposed to the classic wave-of-advance model (Ammerman and Cavalli-Sforza 1973), now the main driver will be the language shift (conversion into another linguistic group) rather than the population growth.

A general reaction-diffusion model to describe the dynamics of two linguistic groups, where a language A is gaining speaker in detriment of language B, may be

$$\begin{cases} \dfrac{\partial n_A}{\partial t} = D\dfrac{\partial^2 n_A}{\partial r^2} + a n_A\left(1-\dfrac{n_A+n_B}{K}\right) + C(n_A, n_B), \\ \dfrac{\partial n_B}{\partial t} = D\dfrac{\partial^2 n_B}{\partial r^2} + a n_B\left(1-\dfrac{n_A+n_B}{K}\right) - C(n_A, n_B). \end{cases} \quad (3)$$

Note that the equations now do not deal with fractions of speakers (p_A and p_B) but with population densities (n_A and n_B). These equations estimate the evolution over time of the density of speakers of each language at every position and time instant in terms of three processes: diffusion, population growth and conversion. The first term on the right-hand side of the equations is the diffusion term which is related to short-range migrations without colonizing intent. This diffusion is characterized by a diffusion coefficient D. The second term is related to the population growth. Population growth is often described by a logistic function where populations with low densities grow exponentially, with a growth rate a, but the process is self-limiting when the population density nears a saturating density defined as the carrying capacity K. In the equation above, Eq. 3, the growth is limited by the densities of both populations, since they all share the same land and resources (Isern and Fort 2010). In addition, since we assume that they all have, in principle, similar ways of live, the parameters D, a and K are the same for both linguistic groups. Finally, the last term corresponds to the conversion of speakers from language B to language A, with the shift rate depending on the densities of speakers of each language at every location. As in the previous Eqs. 1 – 2, the opposed sign in this last term of Eq. 3 means that the speakers lost by language B become speakers of language A.

However, since the introduction of the new language does not yield, in this case, to the assimilation of new technologies leading to a rapid population growth, we may assume that the total population number will vary slowly over time, especially in comparison with the language shift rate. Therefore, as a first approximation, the growth term in Eq. 3 (second term on the right-hand side) can be dropped. Such approximation simplifies the described dynamics, since now we only have to deal with population diffusion and language shift, but it also allows us to rewrite Eq. 3 in terms of the population fraction, thus enabling us to replace the generic conversion term by the language displacement model introduced in the previous section, Eq. 2. The model is then expressed as follows (Isern and Fort 2014),

$$\begin{cases} \dfrac{\partial p_A}{\partial t} = D\dfrac{\partial^2 p_A}{\partial r^2} + \gamma p_A^{\alpha} p_B^{\beta}, \\ \dfrac{\partial p_B}{\partial t} = D\dfrac{\partial^2 p_B}{\partial r^2} - \gamma p_A^{\alpha} p_B^{\beta}. \end{cases} \quad (4)$$

This simplified system describes, for every point in the region, the evolution over time of the fraction of speakers of each language within a population that is not experimenting substantial changes in the total population number. This evolution depends on short-range migrations of the population and the language shift process, according to which the indigenous population acquires a new language that they see as more advantageous socially or economically.

Since we are assuming that the new language is introduced from an adjacent region, the language shift will happen initially near the border, and then the new language will be progressively

introduced further into the territory. From the model above, described by Eq. 4, we can measure the speed of the linguistic frontier by resolving the equation numerically (that is, with a computational simulation). It is also possible to derive mathematical expressions from which it is possible to obtain a range within which lies the real speed, without having to resort to computational simulations. This is possible by assuming that the moving frontier is mostly planar (which is realistic if the language shift "source" is a political border) and through variational analysis of Eq. 4 (Benguri and Depassier 1994, 1998), which leads to the following expression for the upper bound (Isern and Fort 2014)

$$c_U = 2\sqrt{\gamma D}\sqrt{\sup_{p_A\in(0,1)}[\alpha p_A^{\alpha-1}(1-p_A)^{\beta} - \beta p_A^{\alpha}(1-p_A)^{\beta-1}]}, \tag{5}$$

and the following one for the lower bound (Isern and Fort 2014)

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

where the gamma function is defined by the following integral $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$, for $x>0$ (Murray and Liu 1999).

The values of the bounds are obtained from the previous Eqs. 5 – 6 by searching for the maximum result of the right-hand side expression for values of δ in the range $(0,1)$ for the lower bound, and for values of p_A in the range $(0,1)$ for the upper bound.

5 Application to a modern example of language shift

As evidenced in Sect. 2, there have been ongoing processes of language shift throughout history up to the present day. In fact, in recent times we have been experiencing a global process of linguistic convergence towards a few stronger languages—such as English, Spanish or Chinese—leading to the endangerment or extinction of many languages (Krauss 1992, Crystal 2000). Indeed, linguists estimate the about 96% of the population speaks only about 4% of the languages in the world (Crystal 2000), and that about 90% of the current linguistic diversity may become extinct by the end of the century (Krauss 1992).

It is actually this concern over the future of minority languages that have prompted the development of several existing language competition models (e.g., Abrams and Strogatz 2003, Kandler et al. 2010, Isern and Fort 2014); studies that in turn provide a source of data to exemplify here the model presented above. We will focus on the evolution of the Welsh language during the twentieth century. Up to the 1970, when linguistic policies started to be applied, the number of Welsh speakers had been decreasing year after year (Kandler et al. 2010), and the linguistic border had been progressively retreating—at an approximate speed of 0.3–0.6 km/y, according to the estimates obtained from linguistic maps by Fort and Pérez-Losada (2012).

It corresponds therefore to a case of language acquisition from a neighbouring region with a moving linguistic border, and thus a good subject to test the models described above. Of course, since Wales is part of the United Kingdom, and with English being the official language, there is a factor related to the *élite dominance* described above. However, Wales have been part of the UK for a long time

and the most relevant language shift process has taken place in the recent time (Aitchison, and H. Carter 2000); indeed, in the early 1900s, half of the population still spoke Welsh, while by 1980 less than 20% of the population could speak the language (Abrams and Strogatz 2003).

Applying the language shift model described by Eq. 2 to the evolution of the Welsh and English languages during the twentieth century in the region of Monmouthshire yields a very good fit between model and data.[1] This is evidenced by the results in Fig. 1, where the line in the figure represents the best approximation to the historical demographic data (squares) obtained for Eq. (2). The parameter values yielding this best fit are $\alpha = 2.23$, $\beta = 1.76$ and $\gamma = 0.237$.

We now can apply the parameters found above into our spatial model, i.e. the model represented by Eq. 4, where we take into account the geographical evolution in addition to the temporal change, and by the expressions in Eqs. 5 – 6 that provide an estimated range of speeds for the progress of the linguistic border. To do so we will consider two realistic values of the diffusion coefficient, $D = 5.08$ km^2/y and $D = 6.72$ km^2/y. Both are estimated from the expression $D = \langle \Delta^2 \rangle / 4T$ (Fort and Méndez 1999), where $\langle \Delta^2 \rangle$ is an estimation of the mobility of a population over a generation, and T the time span of a generation. In both cases we use a value of the generation time corresponding to modern human populations, $T = 25$ y (Fort and Pérez-Losada 2012). Then, the first value for the diffusion coefficient is estimated from mobility data on modern populations in the Parma Valley, Italy, during the twentieth century and thus coetaneous with the data for the Welsh language; in this case $\langle \Delta^2 \rangle = 508$ km^2 (Isern et al. 2008, Cavalli-Sforza and Bodmer 1999). The second value is estimated from mobility data in Catalonia, Spain, during the eighteenth and nineteenth centuries, with $\langle \Delta^2 \rangle = 672$ km^2 (Heras de Puig 2000).

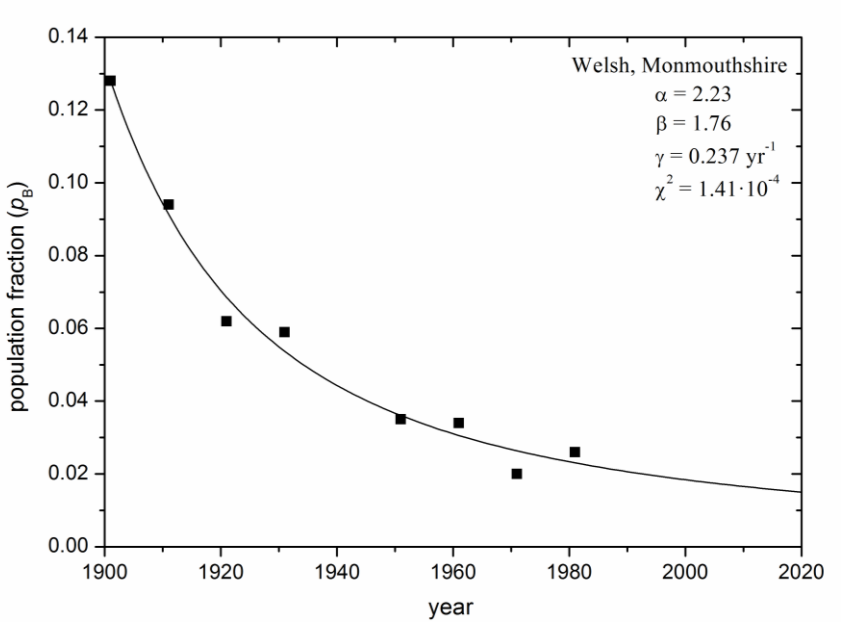

Fig. 1 Decline of the fraction of Welsh speakers over time (squares) and best fit (line) obtained with Eq. 2. (Adapted from Isern and Fort 2014)

[1] As evidenced by Isern and Fort 2014, this model can also be satisfactorily applied to describe the evolution of the fraction of speakers for other current instances of language competition.

To compute these theoretical values we have obtained the parameters from data corresponding only to the region of Monmouthshire rather than the data on all of Wales (also available in Abrams and Strogatz 2003). The reason behind this decision is that 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 by Fort and Pérez-Losada (2012). 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.

The predicted speeds of the linguistic frontier are shown in Table 1, for each of 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 Eq. 4; i.e. a population without substantial population growth and where the prevalent language from an adjacent regions is displacing the indigenous language. Comparing these theoretical values with the speed range estimated from data, $0.3-0.6$ km/yr (Fort and Pérez-Losada 2012), we see that we obtain good agreement between model and observations. Therefore, the model can indeed predict with a fairly good accuracy the actual speed of the linguistic frontier.

In addition, the last two columns in Table 1 contain the values of the upper and lower analytic bounds calculated using Eqs. 5 – 6, 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, these equations can be used as a first approximation of the expected front of linguistic replacement without the need to apply numerical integration.

Table 1 Numerical (c) and analytic (c_L, c_U) predictions of the English linguistic front replacing the Welsh language

D	c	c_L	c_U
km^2/y	km/y	km/y	km/y
5.08	0.557	0.356	0.934
6.72	0.641	0.409	1.750

6 Conclusions

Processes of cultural shift, and in particular of language shift, are present throughout history and have had a great importance into shaping the current day society and diversity. Language is an inherent part of what people use to define their identity, and therefore, the study of the processes that yield a population to abandoning their own language in favour of an alien language is important to understand human evolution. Linguistic studies have described several mechanisms leading to language shift, such as language displacement due to a demographic substitution, due to being militarily conquered by speakers of another language, or as a result of language acquisition from a neighbouring region.

We are particularly interested in the last process, since it produces an advancing linguistic frontier whose speed can be predicted by a mathematical model. 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 (Isern and Fort 2014). We have also applied our language shift model as an interaction term in a reaction-diffusion model in order to estimate the 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. Testing this model over modern data corresponding to the retreat of the Welsh-English border has yielded very accurate results thus indicating that the model provides a good description of the process. In addition, on a wider context, the model presented here could be applied as well to the study of other cases of cultural shift for traits also related to national or ethnic identity, such as religious affiliations.

Acknowledgement

This work was funded in part by FBBVA grant Neodigit-PIN2015, ICREA (JF) and the MINECO grants SimulPast-CSD2010-00034 and FIS-2012-31307 (all authors).

References

Abrams A and Strogatz SH, “Modelling the dynamics of language death,” Nature, vol. 424, p. 900, 2003.

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

Ammerman AJ and Cavalli-Sforza LL, “A population model for the diffusion of early farming in Europe.” In: Renfrew C, editor. The explanation of culture change. London: Duckworth, 1973, pp 343-357.

Barth F, Ethnic groups and boundaries: the social organization of culture difference. Prospect Heights: Waveland Press, 1998.

Baronchelli A, Loreto V and Steels L, “In-depth analysis of the naming game dynamics: the homogeneous mixing case,” Int. J. Mod. Phys. C, vol. 19, pp. 785-801, 2008.

Benguria RD and Depassier MC, “Validity of the linear speed selection mechanism for fronts of the nonlinear diffusion equation,” Phys. Rev. Lett., vol. 73, pp. 2272-2274, 1994.

Benguria RD and Depassier MC, “Speed of fronts of generalized reaction-diffusion equations,” Phys. Rev. E, vol. 57, pp. 6493-6496, 1998.

Bouchard-Côté A, Hall D, Griffiths TL and Klein D, “Automated reconstruction of ancient languages using probabilistic models of sound change," Proc. Natl. Acad. Sci., vol. 110, pp. 4224-4229, 2013.

Cavalli-Sforza LL and Bodmer WF, The genetics of human populations. New York: Dover, 1999.

Cavalli-Sforza LL, Menozzi P and Piazza A, The History and Geography of Human Genes. Princeton: Princeton University Press, 1994.

Clairborne R, The life and times of the English language: the history of our marvellous native tongue. London: Bloomsbury, 1990.

Crystal D, Language death. Cambridge: Cambridge University Press, 2000.

Crystal D, The Cambridge encyclopedia of the English language. Cambridge: Cambridge University Press, 2003.

Forster P, Toth A, “Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European,” Proc. Natl. Acad. Sci., vol. 100, pp. 9079-9084, 2003.

Fort J, “Synthesis between demic and cultural diffusion in the Neolithic transition in Europe,” Proc. Natl. Adac. Sci., vol. 109, pp. 18669-18673, 2012.

Fort J and Méndez V, “Time-delayed theory of the Neolithic transition in Europe,” Phys. Rev. Let., vol. 82, pp. 867-870, 1999.

Fort J and Pérez-Losada J, “Front speed of language replacement,” Hum. Biol., vol. 84, pp. 755-772, 2012.

Fort J and Pujol T, “Time-delayed fronts from biased random walks,” New J. Phys., vol. 9, paper 234, 2007.

Gray RD and Atkinson QD, “Language-tree divergence times support the Anatolian theory of Indo-European origin,” Nature, vol. 426, pp. 435-439, 2003.

Heras de Puig, Biografía, ó, explicació del arbre geneològich de la descendencia de casa Heras de Adri, 1350-1850. Girona: CCG Edicions, 2000.

Isern N, Fort J and Pérez-Losada J, “Realistic dispersion kernels applied to cohabitation reaction-dispersion equations,” J. Stat. Mech Theor. & Exp., paper P10012, 2008.

Isern N and Fort J, “Anisotropic dispersion, space competition and the slowdown of the Neolithic,” vol. 12, paper 123002, 2010.

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

Kandler A, “Demography and language competition,” Hum. Biol., vol. 81, pp. 181-210, 2009.

Kandler A, Unger R, Steele J. “Language shift, bilingualism and the future of Britain’s Celtic languages,” Phil. Trans. R. Soc. B, vol. 365, pp. 3855–3864, 2010.

Krauss M, “The world’s languages in crisis,” Language, vol. 68, pp. 4-10, 1992.

Lightfoot, D, The development of language: acquisition, change and evolution. Malden: Blackwell, 1999.

McMahon AMS, Understanding language change. Cambridge: Cambridge University Press, 1994.

Murray RS and Liu J, Mathematical handbook of formulas and tables, 2nd edn. New York: McGraw-Hill, 1999.

Nowak MA, Komarova NL and Niyogi P, “Computational end evolutionary aspects of language,” Nature, vol. 417, pp. 611-617, 2002.

Pagel M, Atkinson QD, Calude AS and Meade A, “Ultraconserved words point to deep language ancestry across Eurasia,” Proc. Natl. Adac. Sci., vol. 110, pp. 8471-8476, 2013.

Patriarca M, Heinsalu E, “Influence of geography on language competition,” Physica A, vol. 388, pp. 174-186, 2009.

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

Rochetter, B, “Language Policies in the Roman Republic and Empire.” In: Clackson, J, editor. Blackwell companion of the History of the Latin Language. New York: Blackwell, 2011, pp. 549-563.

Smith BD, The emergence of agriculture. New York: Scientific American Library, 1995.

Wang WS-Y, Ke J and Minett JW, “Computational studies of language evolution.” In: Huang CR and Lenders W, editors. Computational linguistics and beyond: Perspectives at the beginning of the 21st century. Taipei: Institute of Linguistics, Academia Sinica, 2004.