The importance of interlinguistic similarity and stable bilingualism when two languages compete

To cite this article: J Mira *et al* 2011 *New J. Phys.* **13** 033007

View the article online for updates and enhancements.

New Journal of Physics

The importance of interlinguistic similarity and stable bilingualism when two languages compete

J Mira[1,3]**, L F Seoane**[1] **and J J Nieto**[2]

[1] Departamento de Física Aplicada, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[2] Departamento de Análise Matemática and Instituto de Matemáticas, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
E-mail: jorge.mira@usc.es

Abstract. One approach for analyzing the dynamics of two languages in competition is to fit historical data for the number of speakers of each with a mathematical model in which the parameters are interpreted as the similarity between those languages and their relative status. Within this approach, on the basis of a detailed analysis and extensive calculations, we show the outcomes that can emerge for given values of these parameters. In contrast to previous results, it is possible that in the long term both languages may coexist and survive. This happens only where there is a stable bilingual group, and this is possible only if the competing languages are sufficiently similar, in which case its occurrence is favoured by both similarity and status symmetry.

Contents

[3] Author to whom any correspondence should be addressed.

1. Introduction

An aspect of globalization that alarms many is the replacement of local tongues by more hegemonic languages [1], a trend that has been investigated from multiple points of view, including that of physics [2]–[7]. Where two languages compete, it is reasonable to assume that the outcome is influenced by their relative status [6] (i.e. the speakers' perceptions of the social and/or economic advantages each language offers) and their similarity [7] (or the complementary concept, interlinguistic distance [8, 9]), but consideration of these factors has been hindered by difficulties in their operationalization.

One of the earliest and simplest mathematical models for language shift was that of Abrams and Strogatz [6], [10]–[12], who considered a stable population in which two languages with different statuses competed for speakers. This model, which involves analysis of the evolution of the number of speakers over time, predicted that one of the languages would inevitably die out, and was successfully fitted to historical data on the competition between Scottish Gaelic and English, Welsh and English, and Quechua and Spanish, among other language pairings [6]. However, it did not take into account the possibility of bilingual individuals, a possibility that is of course realized in numerous multilingual societies. In Spain, for example, where Castilian Spanish is the official language throughout the state, but in certain regions is co-official with another language (mainly Galician, Basque, Catalan or Valencian), individual bilingualism is common in communities with more than one official language.

We recently showed that the historical evolution of the use of Galician and Castilian in Galicia (NW Spain) can be explained by a modified Abrams–Strogatz model that allows for bilingual as well as monolingual speakers of the competing languages, and that includes a parameter that represents the ease of bilingualism [7] (figure 1). We considered a population in which monolingual speakers of the language X make up a fraction x of the population, monolingual speakers of the language Y account for a fraction y and the bilingual (B) for a fraction b (with $x + y + b = 1$). Throughout this paper, capital letters X and Y denote the two languages spoken in the population; the upper case letter B denotes the group of bilingual speakers; and the lower case letters x, y and b refer to the fraction of speakers of each of the languages in the population and the fraction of bilingual speakers, respectively.

The dynamics of language change are accordingly described by the system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = y P_{\mathrm{YX}} + b P_{\mathrm{BX}} - x(P_{\mathrm{XY}} + P_{\mathrm{XB}}), \tag{1a}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x P_{\mathrm{XY}} + b P_{\mathrm{BY}} - y(P_{\mathrm{YX}} + P_{\mathrm{YB}}), \tag{1b}$$

$$\frac{\mathrm{d}b}{\mathrm{d}t} = x P_{\mathrm{XB}} + y P_{\mathrm{YB}} - b(P_{\mathrm{BY}} + P_{\mathrm{BX}}), \tag{1c}$$

where P_{XY} denotes the probability of a monolingual speaker of language X being replaced in the population by a monolingual speaker of language Y, with analogous notation for the other possible replacements. The probability of a monolingual person being replaced by a mono- or bilingual speaker of the other language is assumed to be proportional both to the status of the second language, i.e. the social and/or economic advantages it offers, and to a power of the proportion of the population that speaks it. Thus, denoting by s the relative status of language X
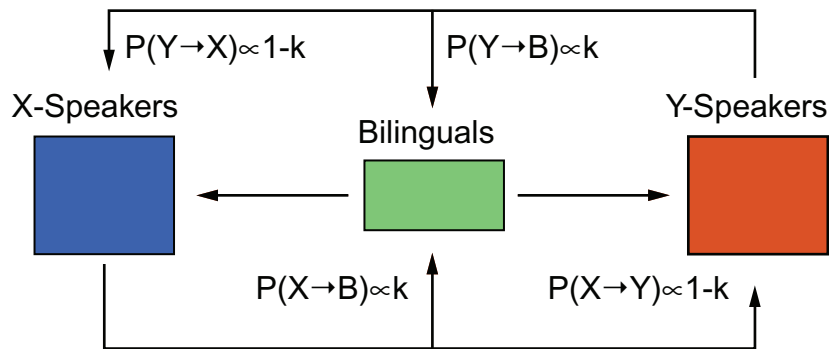
Figure 1. Flux among X-monolingual, Y-monolingual and bilingual groups. According to the model set forth in equations (1a)–(1c) and (2a)–(2d), flux is governed both by the relative statuses of the competing languages (the relative social and/or economic advantages they offer) and by interlinguistic similarity k, the probability that the disappearance of a monolingual speaker of one language will be compensated for by the appearance of a bilingual rather than by a monolingual speaker of the other language.

and by $1 - s$ that of language Y,

$$P_{XB} = c \cdot k(1 - s)(1 - x)^a, \tag{2a}$$

$$P_{YB} = c \cdot ks(1 - y)^a, \tag{2b}$$

$$P_{BX} = P_{YX} = c \cdot (1 - k)s(1 - y)^a, \tag{2c}$$

$$P_{BY} = P_{XY} = c \cdot (1 - k)(1 - s)(1 - x)^a, \tag{2d}$$

where c is a normalization factor related to the time scale, a is the power parameter and k is the probability that the disappearance of a monolingual speaker of language X (respectively Y) will be compensated for by the appearance of a bilingual rather than by a monolingual speaker of language Y (respectively X). We identify interlinguistic similarity with this parameter k.

Note that since people do not actually forget their native tongue(s), the above model is a model of population renewal, rather than of individuals switching from one linguistic practice to another, except that it also allows those who are currently monolingual to become bilingual. When $k = 0$, it reduces to the Abrams–Strogatz model [6] or decays towards this model if b is initially non-zero. Note also that population growth does not invalidate the model, so long as the various linguistic groups are all affected in proportion to their size.

Although the model sketched above adequately accounts for the Galician data up to the present [7], the question arises whether this situation is stable. More generally, what are the possible long-term outcomes of competition between two languages, and under what conditions do they come about? Namely, might two languages coexist stably?

2. Methods and results

To investigate these issues we have carried out extensive calculations, systematically varying s and k, to determine population states (x, y) that act as point attractors for the coupled system
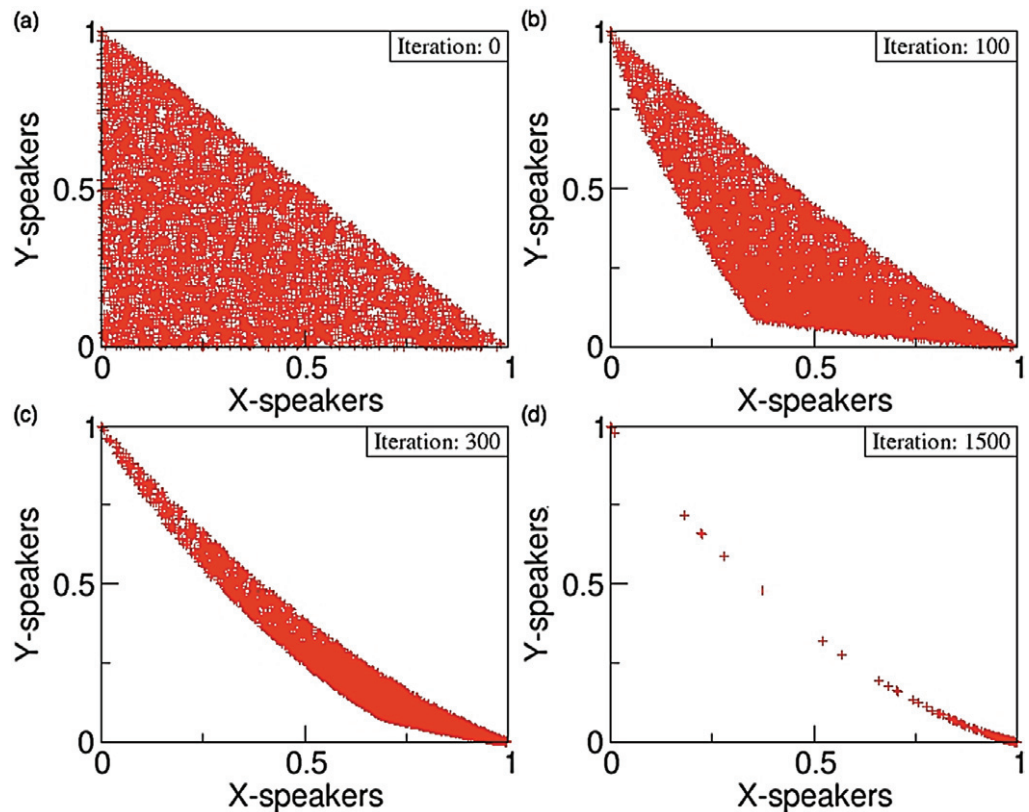
Figure 2. Evolution of language dominance when languages X and Y compete. (a) The proportion of initial speakers of the languages X and Y is represented by a point in the graph, each point in this panel being a possible initial distribution generated at random. (b–d) Each initial distribution of speakers evolves according to the system of equations (1*a*)–(1*c*). The evolution of the proportion of speakers for a given initial situation is tracked down by the trajectory of the points through the phase space. Here we present the evolutions after 100, 300 and 1500 steps of the computation, respectively. In the example $s = 0.75$ and $k = 0.3$. With these parameters, the final distribution of speakers will be that with all the population being monolingual with language X.

defined by equations (1*a*)–(1*c*) and (2*a*)–(2*d*) with $x + y + b = 1$; the problem is well posed because, as is easily shown, the velocity field on the boundary of the set of possible states (defined by $x \geqslant 0$, $y \geqslant 0$, $x + y \leqslant 1$) never takes the system outside this set, whatever the values of s and k.

Initially, our calculations were performed as follows. For all (s, k) of the form $(0.05n_s, 0.05n_k)$ $(0 \leqslant n_s, n_k \leqslant 0)$, 10 000 starting states (x_0, y_0) were randomly selected in the set of possible states, and the discretized form of the above system (1*a*)–(1*c*) was solved numerically for each, with $a = 1.31$, using a time step $\Delta t = 0.01$ (the value $a = 1.31$ was chosen because Abrams and Strogatz [6] found that this parameter was surprisingly constant, 1.31 ± 0.25, when their model was fitted to 42 data sets concerning dissimilar language pairs).
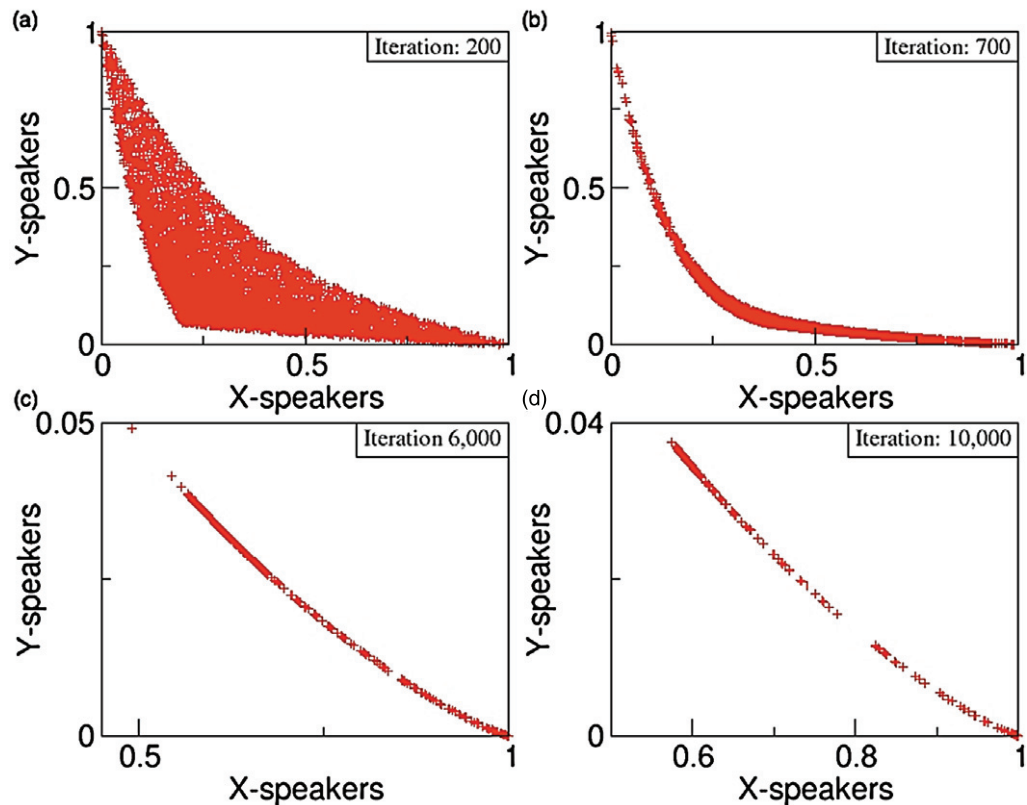
Figure 3. A case with two different attractors. Again, random proportions of speakers of each language were provided as initial conditions for equations (1a)–(1c), this time with parameters $s = 0.67$ and $k = 0.77$. The sampled points evolve according to these equations and snapshots of the phase space are taken after (a) 200, (b) 700, (c) 6000 and (d) 10 000 computation steps. In this case, all the initial distributions of speakers evolved towards a small region, which is zoomed-in in panels (c) and (d). For these parameters, and depending on the initial proportions of the X-monolingual, the Y-monolingual and the bilingual, the language Y either dies out completely or survives, in the latter case mostly among bilinguals. Note the opening of a gap in the points of panel (d), because the states evolve towards different attractors ($x = 1$, $y = 0$) and ($x = 0.6$, $y = 0.04$).

The 10 000 calculations proceeded concurrently and were halted as soon as all the state points had converged to within a circle of radius 10^{-5}. Figures 2(a)–(d) show selected stages of this process for the case ($s = 0.75$, $k = 0.3$). However, it was soon found that there were values of (s, k) for which the state points did not converge, and examination of their distribution indicated that this was due to the existence of more than one point attractor, the attractor to which any particular state point tended, depending on its starting state. In these cases, the calculations were allowed to proceed for times several orders of magnitude longer than the average single-attractor convergence time, until all the state points were within two or three circles of radius 10^{-5} (figure 3 shows the result for the case $s = 0.67$, $k = 0.77$).
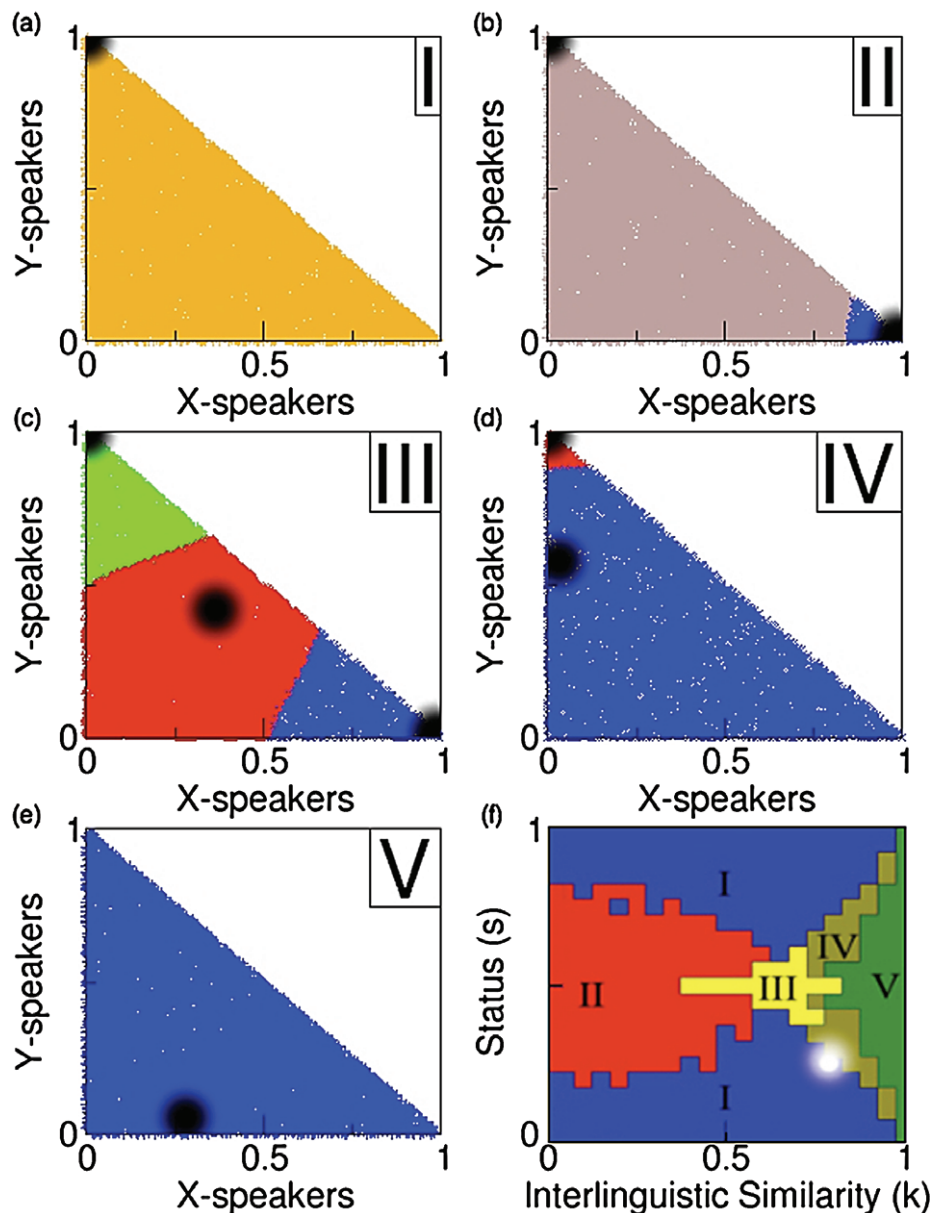
Figure 4. The five possible stable situations determined by s and k. In panels (a–e), the points representing the initial random conditions of five simulations are classified according to the attractor towards which they converge. The different attractors are represented by a huge black spot. All the initial distributions of speakers that collapse to the same stable situation share the same colour. The parameters k and s determine the number and position of the attractors and allow a sorting of the overall system into five topologically dissimilar types (see the text for their description). In panel (a), an example of type I is shown ($s = 0.80$, $k = 0.65$); (b) type II ($s = 0.40$, $k = 0.20$); (c) type III ($s = 0.50$, $k = 0.65$); (d) type (IV) ($s = 0.35$, $k = 0.75$); and (e) type (V). In (f) this sorting is made explicit. All pairs of parameters leading to the same topological distribution of attractors are coloured together. A white spot localizes parameters k and s of the system Galician–Castilian.

Depending on s and k, the following five situations were observed to emerge.

(I) There is just a single stable state at $x = 1$ or $y = 1$, i.e. one of the languages becomes extinct even though it may initially have been dominant. This is the behaviour illustrated in figures 2(a)–(d) and in figure 4(a).

(II) There are stable states at both $x = 1$ and $y = 1$; one of the languages dies out, but which one depends on the initial distribution (x_0, y_0), as illustrated in figure 4(b).

(III) There are stable states at $x = 1$ and $y = 1$, together with a third one that lies below the line $x + y = 1$ and thus corresponds to the presence of a stable bilingual group (figure 4(c)).

(IV) There is a stable state lying below the line $x + y = 1$ (i.e. with a non-empty bilingual group), together with just one stable monolingual state at $x = 1$ (figures 3(a)–(d)) or $y = 1$ (figure 4(d)).

(V) There is just one stable state, and it includes a stable bilingual group (figure 4(e)).

Note that in cases (II), (III) and (IV) there are sharp boundaries between the initial state zones leading to different outcomes; at these boundaries, an exogenous injection of just a few speakers into one group or another can determine whether a language lives or dies, as is illustrated in figures 5(a) and (b).

Figure 4(f) shows which of the above five situations each (s, k) value led to. The salient aspects of this map are that the stable existence of a bilingual group requires that k exceed a minimum value of about 0.35; that if $k \geqslant 0.6$, then the less symmetric languages X and Y are status-wise, the larger k must be for stable bilingualism; and that when the two languages are moderately symmetric status-wise, which of them disappears depends on the initial sizes of the linguistic groups if they are essentially dissimilar ($k \leqslant 0.4$), but not if they are more similar ($0.4 < k < 0.6$).

Returning to the case of Galician and Castilian in Galicia, the inclusion of recently published data [13] in the analysis (figure 5(c)) corroborates our previous estimates [7] of s_{Galician} (0.26) and k (0.80), in fairly good agreement with lexicostatistics-based estimates of similarity among closely related Romance languages [14]. These values place Galicia in zone I of figure 4(f), and accordingly predict the eventual extinction of Galician. However, the close proximity of zone IV suggests, and figure 5(c) shows, that extinction is not imminent: in fact, it is predicted that by the end of the century the population will be roughly equally divided between monolingual Castilian speakers and the bilingual.

3. Final remarks and future work

After exhaustive research in respect of the proposed model, we can conclude that mathematical solutions showing the survival of bilingual speakers are possible, and that these solutions are also linked to the survival of groups of monolingual speakers of both competing languages. This steady coexistence of all the linguistic groups can depend on the global parameters of the system (both languages being more likely to survive where there is large interlinguistic similitude and a status close to 0.5), but also on the initial population supporting each tongue when the system is governed by certain sets of parameters (k, s pairs in zones III and IV in figure 4).

The model used in this paper has two evident limitations. Firstly, the model does not consider possible alterations of the relative proportions of the linguistic groups due to immigration, emigration or differential birth and/or death rates. Secondly, partially related to
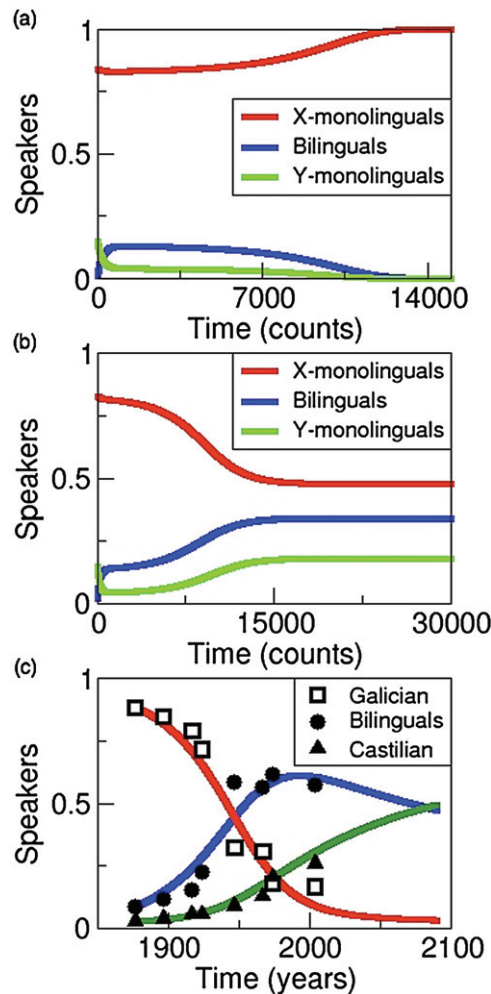
Figure 5. Time evolution of single cases. Time dependence of the proportions of the X-monolingual (red), the Y-monolingual (green) and the bilingual (blue). Panels (a) and (b) illustrate for a type II system ($s = 0.53$, $k = 0.55$) how the fate of the system can depend critically on initial conditions, language Y becoming extinct if initially $x = 0.84$ and $y = 0.15$ (a), but not if $x = 0.83$ and $y = 0.15$ (b). (c) Results of fitting the model to historical data for Galician (X, red), Castilian (Y, green) and the bilingual (B, blue) in Galicia (NW Spain). Although the fitted values of s and k, 0.26 and 0.80, place Galicia in zone I of figure 4(f), thus predicting the eventual extinction of Galician, those bilingual in Galician and Castilian are not expected to disappear within this century or the next.

this is the consideration that the relative status of the two languages may well vary in time. Also, of course, it remains to be seen whether the definition of linguistic similarity in terms of the dynamics of the population renewal process corresponds to any purely linguistic concept; the assumption that similarity is symmetric, i.e. that $P_{XB}/[(1-s)(1-x)^a] = P_{YB}/[s(1-y)^a]$, is clearly an idealization; and there is the issue of whether two very similar languages really have separate identities, especially when they have a similar status and one or both only survive

among *bilingual* speakers. The present results nevertheless suggest that the competition between two languages does not inevitably lead to the extinction of one of them.

Future work on the topic should reduce the limitations outlined earlier. In addition, a complementary analytical procedure should be possible, which could suggest new ways of overcoming these limitations. From such an approach, further analysis of bifurcation and/or phase transitions could enrich the model.

Continuing the numerical perspective, a modelling of the system of two competing languages incorporating migrational (or any other of the previous) phenomena could bring about new dynamics worthy of study. Allowing for varying statuses and interlinguistic similarity could suggest further and more precise political guidelines for protecting endangered tongues, as well as illuminating the evolution of the language entities themselves.

Another interesting approach would be the introduction of a spatial structure—a strategy already tried for previous models [11]—or of a social one. Such research would require modelling based on agents, rather than on mean-field solutions, and the introduction of complex networks. From this new paradigm, the possibility of fairly different results cannot be discarded, as reported in [15, 16] for evolutionary games, which bear a considerable similarity to the current system. Notwithstanding the great impact that an underlying social network could have on the interaction, the mean-field approach presented in this paper should be obtained within a certain limit in any future modelling.

Acknowledgment

References

[1] Sutherland W J 2003 *Nature* **423** 276
[2] Hawkins J A and Gell-Mann M 1992 *The Evolution of Human Languages* (Reading, MA: Addison-Wesley)
[3] Schulze C, Stauffer D and Wichmann S 2008 *Commun. Comput. Phys.* **3** 271–94
[4] Castelló X, Eguiluz V M and San Miguel M 2006 *New J. Phys.* **8** 308
[5] Crystal D 2000 *Language Death* (Cambridge: Cambridge University Press)
[6] Abrams D M and Strogatz S H 2003 *Nature* **424** 900
[7] Mira J and Paredes A 2005 *Europhys. Lett.* **69** 1031–4
[8] Nowak M A, Komarova N L and Niyogi P 2002 *Nature* **417** 611–7
[9] Petroni F and Serva M 2008 *J. Stat. Mech.* P08012
[10] Patriarca M and Leppanen T 2004 *Physica* A **338** 296–9
[11] Patriarca M and Heinsalu E 2009 *Physica* A **388** 174–86
[12] Stauffer D, Castelló X, Eguíluz V M and Miguel M S 2007 *Physica* A **374** 835–42
[13] González-González M, Rodriguez-Neira M A, Fernández-Salgado A, Loredo-Gutiérrez X and Suárez-Fernandez I 2004 *Mapa Sociolingüístico de Galicia* (A Coruna: Real Academia Galega) http://www.udc.es/snl/documentospdf/LinguaInicial.pdf
[14] Dyen I, Kruskal J B and Black P 1992 *Trans. Am. Phil. Soc.* **82** 5
[15] Szabó G and Fáth G 2007 *Phys. Rep.* **446** 97–216
[16] Helbing D, Szolnoki A, Pere M and Szabó G 2010 *PLoS Comput. Biol.* **6** e1000758