Modelling endangered languages: The effects of bilingualism and social structure

James W. Minett [*], William S-Y. Wang

*Language Engineering Laboratory, Department of Electronic Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong*

Abstract

The mathematical model for language competition developed by Abrams and Strogatz allows the evolution of the numbers of monolingual speakers of two competing languages to be estimated. In this paper, we extend the model to examine the role of bilingualism and social structure, neither of which are addressed in the previous model. We consider the impact of two strategies for language maintenance: (1) adjusting the status of the endangered language; and (2) adjusting the availability of monolingual and bilingual educational resources. The model allows us to predict for which scenarios of intervention language maintenance is more likely to be achieved. Qualitative analysis of the model indicates a set of intervention strategies by which the likelihood of successful maintenance is expected to increase.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Language competition; Language death; Language maintenance; Dynamical system; Agent-based model

1. Introduction

The 6000 or so languages spoken on our planet today are the products of numerous millennia of cultural evolution. They encapsulate the experience and knowledge of diverse peoples collected in widely different environments, and are a precious part of the human heritage. With the explosive expansion of a few dominant languages in recent decades, at least half of the world's languages are critically endangered in that they will soon have no speakers and become extinct (Krauss, 1992; Crystal, 2000). Pagel (1995) estimates that roughly 140,000 languages have ever existed (median estimate), so it is the fate of the majority of languages to become extinct. Fishman (1991) argues that

---

* Corresponding author. Tel.: +852 3163 4346; fax: +852 2603 5558.
E-mail addresses: jminett@ee.cuhk.edu.hk (J.W. Minett), wsywang@ee.cuhk.edu.hk (W.S-Y. Wang).

death of a language often leads to death of the underlying culture to which it is linked. It is, therefore, an important challenge to understand such situations as precisely as possible, and to recognize whether there are measures that can help us preserve some of this heritage.

Much work has been carried out on both theoretical and empirical issues of achieving language maintenance as evidenced by the numerous recent volumes on the subject (e.g. Fishman, 1991; Grenoble and Whaley, 1998; Crystal, 2000; Nettle and Romaine, 2000; Fishman, 2001; Bradley and Bradley, 2002; Grenoble and Whaley, 2006). Fishman (2001:1) begins his treatise by stating:

> What the smaller and weaker languages (and peoples and cultures) of the world need are not generalized predictions of dire and even terminal illnesses but, rather, the development of therapeutic understandings and approaches that can be adjusted so as to tackle essentially the same illness in patient after patient.

Toward this end, Abrams and Strogatz (2003) have proposed a mathematical model for studying language competition. The model predicts that whenever two languages compete for speakers, one language will eventually become extinct, the language that dies depending on the initial proportions of speakers of each language and their relative status. The model obtains a good fit to a number of empirical data sets, tracing the relative abundance of speakers of several endangered languages that have been undergoing competition with other, more prestigious languages. However, it does not account for either bilingual individuals or the social structure of the population within which the languages compete. The model also does not distinguish the vertical and horizontal transmission of language, and ignores the impact of the behaviour of individual speakers in the population, focusing instead on the expected aggregate behaviour of the population as a whole. Despite these limitations, the model has stimulated a burst of research into the dynamics of language competition and diversity (e.g., Patriarca and Leppänen, 2004; Mira and Paredes, 2005; Wang and Minett, 2005; Schulze and Stauffer, 2006; Stauffer et al., 2006), much of it performed by non-linguists.

In this paper, we extend the Abrams and Strogatz work to model bilingualism explicitly, accounting for the fact that some individuals may speak both of the competing languages. Our first step is to formulate a mathematical model that, like the Abrams and Strogatz model, deals with the expected aggregate behaviour of the whole population—this model also predicts that death of one of the two competing languages is inevitable (although the trajectories that lead to this state differ from those of the previous model). However, in order to investigate the range of possible trajectories that a system of two competing languages can potentially follow from the same initial state, and to devise a method by which the probabilities associated with them can be predicted, we also implement an agent-based model. In particular, we investigate the impact of simple strategies for language maintenance, allowing us to estimate for different scenarios the relative likelihood that a pair of competing languages can be maintained in a population comprising both monolingual speakers of each language and bilingual speakers. We then examine the role of social structure on the probabilities of maintenance, representing the social structure by a local-world network (Li and Chen, 2003) to encapsulate the patterns of sociolinguistic interactions among the individuals comprising the population.

The paper is laid out as follows: In section 2, we discuss deterministic models of language competition, first briefly describing the Abrams and Strogatz model for a population in which two languages compete for speakers, then introducing an extension of the model that incorporates modelling of bilingualism. In section 3, we apply the extended model to investigate the efficacy of simple strategies for language maintenance. The deterministic models that we discuss in sections 2

and 3 allow us to determine the expected final state of a system in which two languages compete for speakers. However, they do not reveal the variety of competing outcomes that are possible, the likelihood that the system converges to each of them, or the impact of population size on the likelihood of maintenance. Therefore, in section 4 we go on to derive an agent-based model for language competition that allows such effects to be assessed. Our simulations suggest some general principles for achieving maintenance. We then consider the impact of social structure on the probability of successful maintenance in section 5. The paper concludes in section 6.

2. Deterministic models of language competition

2.1. The monolingual model of Abrams and Strogatz (2003)

Abrams and Strogatz (2003) (henceforth A&S) have developed a simple mathematical model of language competition for a population in which two languages compete for speakers. The model aims to trace the variation over time of the proportion of speakers of each language, potentially allowing a language that is endangered to be identified at an early stage so that appropriate action to maintain it can be planned.

In deriving their deterministic model, A&S make a number of simplifying assumptions, in particular:

 (i) Each individual is monolingual in one of the two languages, but no individual speaks both languages.
 (ii) Speakers of one language switch to speak the second language according to the 'attractiveness' of that second language.
(iii) The attractiveness of a language increases with both the proportion of speakers of that language and its 'status', a global parameter that measures its inherent utility within the community—we shall discuss status in section 3.
(iv) The population has a uniform social structure: individuals interact with each other at the same rate, and influence the languages that they each use equally—we discuss social structure in section 5.
 (v) The population size remains constant.

Let us consider how these assumptions translate into a mathematical model. We write the proportions of monolingual speakers of X and Y as x and y, respectively. Since each individual is assumed to be monolingual in one of these two languages, the values of x and y must sum to 1 (i.e., $x + y = 1$). The rate at which speakers of one language switch to become speakers of the second language depends on the attractiveness of that second language. In their most general conception of attractiveness, A&S assume that a language has greater attractiveness the more monolingual speakers it has and the greater its status. All such functional forms of attractiveness, they state, give rise to the same qualitative dynamics as that of the following constrained model in which the attractiveness of language X to speakers of Y is given by the power-law

$$P_{YX} = cs_X x^a. \tag{1}$$

In the above formula, s_X denotes the status of language X, and a is a parameter that models how the attractiveness of X scales with the proportion of speakers of X. The attractiveness of Y to speakers of X can be stated similarly. For $a = 1$, the attractiveness of X increases linearly with its proportion of speakers. For a greater 1, doubling the proportion of speakers more than doubles

the attractiveness, and vice versa. The parameter c in (1) indicates the peak rate at which speakers of Y switch to speak X. This can reflect, variously, the rate of contact between pairs of individuals, the propensity for individuals to learn a new language based on their existing linguistic skills, or the provision of language resources to children. The formula (1) therefore provides a simple model for a variety of sociolinguistic factors that influence the attractiveness of a language, and whose qualitative dynamics reflect those of the entire class of models in which attractiveness increases monotonically with both the proportion of monolingual speakers and status.

In this model, Y monolinguals switch language to become monolingual speakers of X at a rate proportional to P_{YX}. Likewise, X monolinguals become Y monolinguals at a rate proportional to P_{XY} (defined similarly to (1)). The rate at which the proportion of monolingual speakers of X changes over time can therefore be written symbolically as

$$\frac{dx}{dt} = yP_{YX} - xP_{XY}. \tag{2}$$

Substituting for P_{YX} and P_{XY} in (2), we obtain the formula

$$\frac{dx}{dt} = c[yx^a s_X - xy^a s_Y]. \tag{3}$$

The rates of change of the system are summarized in Fig. 1. We can make further simplifications to the formula by assuming, as do A&S, that $s_X + s_Y = 1$, and by substituting $(1-x)$ for y to obtain

$$\frac{dx}{dt} = c[(1-x)x^a s_X - x(1-x)^a (1-s_X)]. \tag{4}$$

Because the proportions of speakers of X and Y sum to 1, there is no need to consider separately the rate at which the proportion of Y monolinguals changes with time—this is just minus the rate of change of the X monolinguals.

The dynamics of the model can be analyzed by seeking the values of x at which no change in the proportion of X monolinguals is expected to take place. These points are termed 'equilibria', and are of two types: stable and unstable.[1] In order to locate the equilibria, we calculate the values of x for which the rate of change of x is zero. Stable equilibria occur at $x = 0$ and at $x = 1$. A third equilibrium occurs for an intermediate value of x, corresponding to a situation in which speakers of both languages remain, but this equilibrium is unstable. Therefore, no matter what is the initial state, the system will ultimately end up at one of the two stable equilibria. The significance for language competition is that this model implies that one language will always acquire all the speakers in the population, causing the language with which it competes to become extinct. However, A&S suggest that a third stable equilibrium, corresponding to a state in which the two competing languages are both maintained, can be achieved by appropriate control of the status of the endangered language. The idea of controlling the status, as well as other model parameters, in order to achieve maintenance forms the basis of the language maintenance strategies that we pursue in section 3.

---

[1] Informally, an 'equilibrium', or 'fixed point', of a dynamical system is a state in which the system will remain, once attained. In other words, an equilibrium is a state for which the rate of change is zero, i.e., $dx/dt = 0$. An equilibrium is said to be 'stable' when the system, upon a small perturbation from the equilibrium, returns back to that state. We shall refer to all equilibria that are not stable as 'unstable'. More formal definitions of these and other terms relating to dynamical systems can be found, for example, in Strogatz (1994).
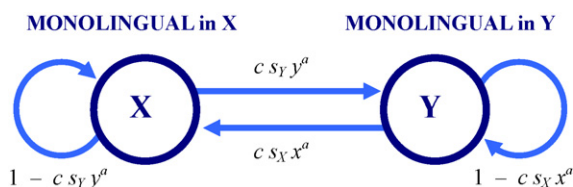
Fig. 1. Transition rates for the monolingual model of Abrams and Strogatz (2003).

A&S have tested the accuracy of their model by fitting it to diachronic data collected for three endangered languages: Scottish Gaelic, Welsh and Quechua. The former two data sets are based on census data, while the latter data derive from records of the languages used in religious services in Peru. In each case, the set of values of the model parameters for which the corresponding trajectory most closely matches the diachronic data can be calculated. In this way, the status of Scottish Gaelic was estimated to be 0.33, half that of English with which it continues to compete. Welsh was estimated to have a slightly higher status with respect to English, 0.4, but the predicted trajectory for the proportion of speakers of Welsh throughout Wales nevertheless shows a strong downward trend that fits the empirical data well. A&S also estimated the value of parameter a for each data set, and found a to cluster about the mean value 1.31. This value of a exceeds 1, indicating that the attractiveness of a language more than doubles as its proportion of speakers doubles. We have therefore assumed that the parameter a takes a value no less than 1 in the course of development of our own model.

The model of A&S appears to work well in modelling changes in the patterns of language usage within a population in which two languages compete. However, the model deals only with monolingual speakers. In practice, we observe that typically a speaker does not suddenly give up one language completely in favour of another—it is extremely rare, for example, for children to lose the ability to communicate with their parents. Almost always, speakers will maintain the language acquired from their parents and, perhaps, learn additional languages to various degrees, particularly while young. Such speakers may switch languages back and forth, depending on the context, whether it be home, school, or workplace. The nature and extent of the bilingualism will depend on a variety of societal factors. We therefore believe that the incorporation of bilingualism is essential for realistic modelling of language death.

2.2. A bilingual model incorporating vertical and horizontal transmission

We here extend the A&S model by explicitly modelling bilingualism, which we accommodate by introducing a third class of speakers, Z, who speak both X and Y. Whereas A&S model only two types of transition (X → Y and Y → X), there are potentially six types of transition possible among monolingual and bilingual speakers of two languages (X → Y, Y → X, X → Z, Z → X, Y → Z, and Z → Y). The transitions X → Y and Y → X are exceedingly rare in practice, whether as a result of vertical transmission or horizontal transmission. One would not expect a child of monolingual speakers of X, say, to acquire only language Y. Nor would one expect an adult who previously spoke only language X to then acquire Y and simultaneously forget how to speak X. The other four types of transition, however, all occur frequently in practice. We therefore model only transitions of the four types X → Z, Z → X, Y → Z, and Z → Y, as suggested by Wang and Minett (2005).

Consider first vertical transmission. Children of monolingual parents necessarily acquire the language of their parents as their first language. However, children of bilinguals may acquire

either or both of the competing languages. For simplicity, we adopt a uniparental model of transmission. We therefore adopt the following model for vertical transmission (V-Model):

- All children of monolingual parents acquire the language of their parents, that is $X \to X$ and $Y \to Y$.
- Children of bilingual parents acquire only language X from their parents, that is $Z \to X$, at a rate proportional to the attractiveness of X.
- Children of bilingual parents acquire only language Y from their parents, that is $Z \to Y$, at a rate proportional to the attractiveness of Y.
- All other children of bilingual parents acquire both languages from their parents so that they too become bilingual, that is $Z \to Z$.

We assume that the attractiveness of language X to a child of a bilingual parent increases with both the status of X and the proportion of monolingual speakers of X. For a given status, X is maximally attractive when the entire population is monolingual in X, and minimally attractive when none of the population is monolingual in X. Given this, it follows that the attractiveness of acquiring both languages to a child of a bilingual parent is maximal when the entire population is bilingual, and minimal when the entire population (other than the parent) is monolingual. Adopting the same functional form for attractiveness as in (1) above, we write the attractiveness of being monolingual in X to children of bilingual parents as

$$P_{ZX} = c_{ZX} s_X x^a. \tag{5}$$

A similar formula holds for the attractiveness of Y. Notice that in (5) we have defined a parameter, c_{ZX}, which allows the peak attractiveness of X to be modelled independently from that of Y, reflecting, for example, the different availability of educational resources in each language. Control of these parameters will form the basis of a proposed strategy for language maintenance that we discuss in section 3.

Symbolically, the V-Model for language competition can be written as

$$\begin{aligned} \frac{dx}{dt} &= z P_{ZX} \\ \frac{dy}{dt} &= z P_{ZY} \end{aligned}, \tag{6}$$

where z denotes the proportion of bilingual speakers. Note that, for the bilingual model, $x + y + z = 1$. The rate of change of the proportion of bilinguals is therefore simply minus the sum of the rates of change of the proportions of monolingual speakers of X and Y. By substituting for P_{ZX} and P_{ZY} in (6), we obtain the V-Model

$$\begin{aligned} \frac{dx}{dt} &= c_{ZX} s_X (1 - x - y) x^a \\ \frac{dy}{dt} &= c_{ZY} s_Y (1 - x - y) y^a \end{aligned}. \tag{7}$$

What then of horizontal transmission? We have assumed that adults retain sufficient knowledge of a language, once acquired, for that language to be available for transmission to any offspring they might produce. That being the case, bilingual adults are assumed to remain

bilingual throughout their lifetimes. Monolingual adults, however, may either remain monolingual or subsequently become bilingual by acquiring the second language. We therefore adopt the following model for horizontal transmission (H-Model):

- All bilingual adults remain bilingual, that is Z → Z.
- Adults speaking X only subsequently acquire language Y, that is X → Z, at a rate proportional to the attractiveness of Y.
- Adults speaking Y only subsequently acquire language X, that is Y → Z, at a rate proportional to the attractiveness of X.
- All other monolingual adults remain monolingual, that is X → X or Y → Y.

In addition to being able to communicate with both X monolinguals and bilinguals, the X monolingual adult who subsequently acquires language Y can additionally communicate with Y monolinguals. Therefore, we assume that the attractiveness of acquiring language Y, so becoming bilingual, to monolingual speakers of X increases with both the status of Y and the proportion of monolingual speakers of Y:

$$P_{XZ} = c_{XZ}s_Y y^a, \tag{8}$$

a functional form consistent with (5). The attractiveness, P_{YZ}, of becoming bilingual to monolingual speakers of Y is defined similarly.

The H-Model can be written as

$$\begin{aligned} \frac{dx}{dt} &= -xP_{XZ} \\ \frac{dy}{dt} &= -yP_{YZ} \end{aligned}, \tag{9}$$

which reduces to the following system after substituting for P_{XZ} and P_{YZ}:

$$\begin{aligned} \frac{dx}{dt} &= -c_{XZ}s_Y xy^a \\ \frac{dy}{dt} &= -c_{YZ}s_X yx^a \end{aligned}. \tag{10}$$

To develop a unified bilingual model for language competition, encompassing both vertical and horizontal transmission, we model the rates at which individuals follow the V-Model and H-Model. We do so by defining a mortality rate, μ, at which adults are replaced by children. Children acquire languages from their parent according to the V-Model; surviving adults all acquire languages according to the H-Model. The unified bilingual model can therefore be stated explicitly as

$$\begin{aligned} \frac{dx}{dt} &= \mu c_{ZX}s_X(1 - x - y)x^a - (1 - \mu)c_{XZ}s_Y xy^a \\ \frac{dy}{dt} &= \mu c_{ZY}s_Y(1 - x - y)y^a - (1 - \mu)c_{YZ}s_X yx^a \end{aligned}. \tag{11}$$

The rates of change of the unified bilingual model are summarized in Fig. 2. For convenience, throughout the remainder of this paper, we will assume, without loss of generality, that $s_X + s_Y = 1$, allowing us to substitute s_Y with $(1 - s_X)$.
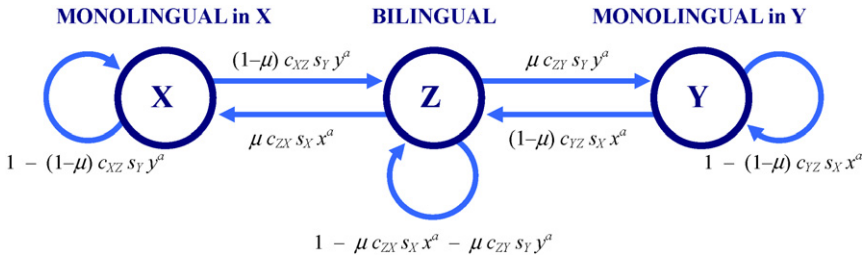
Fig. 2. Transition rates for the unified bilingual model. (a) Transition rates for the V-Model. (b) Transition rates for the H-Model.

In order to determine the dynamics of the unified bilingual model, and thereby draw conclusions regarding the competition of two languages, we study the 'direction field'[2] of the system (11) for different sets of values of the model parameters s_X, c_{XZ}, c_{YZ}, c_{ZX}, and c_{ZY}. Fig. 3 shows the direction field for two values of the status of language X ($s_X = 0.5$ and $s_X = 0.3$). The system has four equilibria: X^* and Y^* correspond to all individuals being monolingual in X and Y, respectively; Z^* corresponds to all individuals being bilingual; and U^* corresponds to a state in which there are some speakers of each type. It can easily be shown (by linearizing (11) about each equilibrium) that X^* and Y^* are both stable, meaning that once the system has approached either of these two states, it will remain nearby. However, both Z^* and U^* are unstable, meaning that even if these states are approached, the system will subsequently tend to move away from them.

The position of U^* shifts according to the values of the model parameters. For X having the same status as Y ($s_X = 0.5$, shown in panel a), the position of U^* corresponds to a state in which X and Y have the same numbers of monolingual speakers. However, as the status of X is reduced, so the position of U^* moves towards the stable equilibrium X^*. As a result, fewer trajectories converge on X^*, meaning that language X will become extinct for a greater range of initial states. In terms of the language competition, the model predicts that one of the two competing languages will eventually acquire all the speakers, regardless of the initial conditions, resulting in a monolingual system in which only one language is spoken.

2.3. The bilingual model of Mira and Paredes (2005)

A different approach to modelling bilingualism, in which the two languages that compete for speakers are partially mutually intelligible, has been suggested by Mira and Paredes (2005). When two languages are partially mutual intelligible, monolingual speakers of one language can sometimes communicate effectively with monolingual speakers of the competing language. Such communications potentially allow monolingual speakers to become bilingual. Monolingual speakers of a language X, say, are therefore assumed to become bilingual at a rate that is proportional both to the proportion of speakers of the competing language Y (as in our extension of the A&S model) and to the degree of mutual intelligibility. Correspondingly, the greater the mutual intelligibility, the lesser is the proportion of monolingual speakers of X attracted to become monolingual in Y.

---

[2] The 'direction field' of a dynamical system is a diagram that shows, for a representative set of states of that system, the direction in which it is expected to change state, the direction of change being indicated by an arrow. A similar diagram, called a 'phase portrait', additionally represents the rate of change at each point by the length of the arrow (longer arrows indicate a faster rate of change).
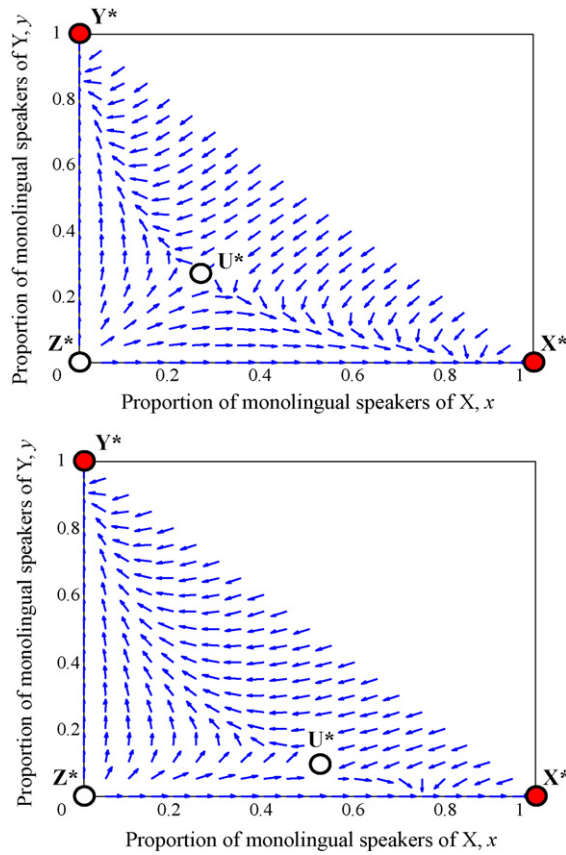
Fig. 3. Direction field for the unified bilingual model for two values of status, s_X. Each arrow shows the direction of change of the system at that state, allowing trajectories and the stability of equilibria to be inferred. Stable equilibria are marked by filled circles; unstable equilibria are marked by unfilled circles ($a = 1.0$; $\mu = 2\%$; $c_{XZ} = c_{YZ} = 0.035$; $c_{ZX} = c_{ZY} = 1.0$; $s_X = 0.3, 0.5$).

Based on these amendments to the A&S model, Mira and Paredes derive the system

$$
\begin{aligned}
\frac{dx}{dt} &= c[(1-k)(1-x)(1-y)^a s_X - x(1-x)^a(1-s_X)] \\
\frac{dy}{dt} &= c[(1-k)(1-y)(1-x)^a(1-s_X) - y(1-y)^a s_X]
\end{aligned}, \tag{12}
$$

where k, with value between 0 and 1, denotes the degree of mutual intelligibility between languages X and Y. This system is noteworthy since, for sufficiently large k, a stable equilibrium is introduced in which both bilingual speakers and monolingual speakers of the more prestigious language survive. Although no monolingual speakers of the endangered language remain, that language is preserved among the bilingual members of the population.

The model provides a good fit to data for the competition between the Castilian and Galician dialects of Spanish during the period 1875–1975 (Mira and Paredes, 2005), but we doubt that it can be usefully applied in its current form to general situations of language competition. Often, competing languages are mutually unintelligible ($k \approx 0$), in which case the model reduces to the

monolingual model of A&S, which, we have argued, is unsuited to modelling language competition due to the lack of bilingual modelling. Also, even when the competing languages are mutually intelligible dialects or closely genetically related languages, it is not clear that individual speakers make the transition directly between monolingual use of one language to monolingual use of the competing language (X → Y or Y → X) in any but the rarest of situations. Moreover, there is no linguistic advantage conferred to a bilingual individual when no monolingual speakers of one of the competing languages remain: in such a situation, one would expect the bilinguals to be quickly replaced by monolinguals, resulting in the extinction of the endangered language, behaviour that is reproduced by the model we have proposed in section 2.2. Nevertheless, tweaking the model by constraining the transitions that are likely to occur in practice, as we have done in the previous section, and by considering the impact of social structure, as we will do in section 5, may result in a model that allows the dynamics of competition between mutually intelligible, related languages to be better understood.

3. Language maintenance

Fishman has proposed the 8-stage Graded Intergenerational Disruption Scale (GIDS) by which the prospects for the continuing usage of a particular language in a community can be assessed (Fishman, 1991, 2001). The GIDS can be used to identify the contexts in which the language is spoken, ranging from nationwide usage throughout the mass media and governmental operations, to the language being taught as an option in literacy schools, to sparse usage by socially isolated elderly people. By using this scale, languages that are threatened with extinction can be detected.

Once a language has been identified as endangered, it must be decided whether or not an attempt will be made to maintain it. It is well known that maintenance of an endangered language can sometimes be achieved by top down processes like legislation or by bottom up movements like ethnic pride. Sometimes, however, the socio-economic factors that led to a language becoming endangered in the first place might discourage its speakers—and linguists—from seeking to maintain it. As Mufwene writes, ''Linguists concerned with rights of languages must ask themselves whether these rights prevail over the rights of speakers to adapt competitively to their new socioeconomic ecologies'' (Mufwene, 2004:219). Our focus in this paper is not to advocate language maintenance in all situations. Rather, we aim to develop quantitative models that can assist in the identification of maintenance strategies that can be applied in certain situations of language endangerment. It is for the community whose indigenous language is endangered to decide whether maintenance should be undertaken.

Crystal (2000) has identified six main mechanisms of intervention by which maintenance may be attempted:

 (i) Increasing the prestige of its speakers.
 (ii) Increasing the wealth of its speakers.
(iii) Increasing the power of its speakers.
(iv) Improving its presence in the educational system.
 (v) Ensuring that the language can be written down.
(vi) Providing access to electronic technology to its speakers.

The first three mechanisms relate to the 'status' of the language, an assessment of the socioeconomic advantages conferred to members of the community speaking that language.

Bringing about an increase in the status of a language may serve to maintain it. The latter three mechanisms relate to people's 'access' to the language, the ease with which they can come to make productive use of the language. For example, promoting the teaching of an endangered language at school may enhance the fluency of its existing speakers and encourage bilingualism among speakers of the language with which it competes, both of which may have impact on its viability.

3.1. Modelling language maintenance

We model the effects of adopting such mechanisms of intervention as adjustments to the parameters s_X, c_{XZ}, c_{YZ}, c_{ZX}, and c_{ZY} of the system (11). In particular, we seek strategies for controlling these parameters such that an endangered language, X, and its competitor, Y, can both be maintained. We make no attempt here to determine optimal strategies for language maintenance—there is little value in doing so since, at this time, the precise quantitative relationships among the various maintenance mechanisms and the parameters of models such as the one proposed here are still only poorly understood. Rather, we investigate conditions under which intervention that brings about significant changes to the model parameters of an endangered language can lead to that language being maintained.

We suppose that the value of an arbitrary parameter, θ, of system (11) can be represented as a function, $\theta(x)$, of the proportion of monolingual speakers of the endangered language X. In order to obtain a model that is amenable to analysis, we assume that a community can bring about a change in the value of the parameter whenever the proportion of monolingual speakers of X falls below some threshold, $x < th_X$. Symbolically, we write this as

$$\theta(x) = \begin{cases} \theta : & x \geq th_X, \\ \theta' : & x < th_X \end{cases}, \tag{13}$$

where θ and θ' are both constants, θ representing the value of the parameter prior to intervention, and θ' representing its value after intervention. A graph of $\theta(x)$ is shown in Fig. 4. In practice, the extent to which a community can bring about a change in a parameter is likely to be limited, and the speed with which it can do so constrained. However, a more realistic intervention model, having less abrupt onset and offset (indicated in the figure by the dotted lines), generally gives rise to the same qualitative behaviour as (13), which we now describe.

We assume that the system (11) is valid both pre- and post-intervention, but with different sets of values of the parameters s_X, c_{ZX}, c_{ZY}, c_{XZ}, and c_{YZ}. Fig. 5 illustrates the impact of an intervention that brings about an increase in the status of a language. The first panel shows the direction field for the system without intervention when the status of the language X is 0.4. The only stable states correspond to death of either language X or language Y. The second panel shows the direction field corresponding to language X having status 0.6. Again, the only stable states correspond to death of one of the two competing languages. Panel (c), however, shows the direction field for the system in which the pre-intervention status of X is increased from 0.4 to the post-intervention status 0.6 whenever the proportion of monolingual speakers of X falls below 30%. By intervening in this way, an additional stable fixed point, S^*, is created between the two unstable fixed points, U_1^* and U_2^*.

We can better understand which strategies give rise to the stable fixed point S^*, and so lead to the potential maintenance of the two competing languages, by studying the 'nullclines' of the system (11). The nullclines are lines along which the rate of change of either x or y is zero: that is, $dx/dt = 0$ or $dy/dt = 0$. The system without intervention has two nullclines, shown superimposed on the direction field in Fig. 6a. The nullclines intersect at the unstable equilibrium U^* and
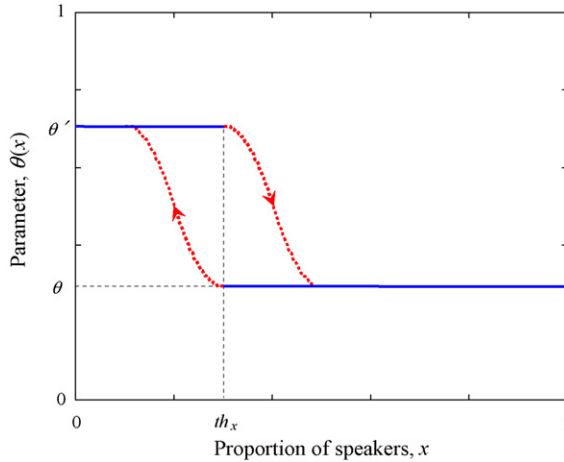
Fig. 4. Functional relationship between the proportion, x, of speakers of an endangered language X and an arbitrary sociolinguistic parameter, $\theta(x)$. The solid line models a population intervening to bring about a discrete increase in the value of the parameter, from θ to θ', as its proportion of speakers falls below a threshold, th_X—this functional relationship is assumed in the experiments throughout this paper. The dashed lines model a more gradual increase in the value of the parameter at the onset of intervention, followed by a more gradual decrease in its value at the offset of intervention after maintenance has been achieved.

partition the state space into four regions, each of which is characterized by a direction of change of x (increasing or decreasing) and a direction of change of y (increasing or decreasing), as shown in Fig. 6b. Trajectories that start in two of these regions —$\mathbf{R_U}$ and $\mathbf{R_Z}$—initially move towards the equilibrium U^* but, because this equilibrium is unstable, then enter one of the two remaining regions—$\mathbf{R_X}$ and $\mathbf{R_Y}$—after which the system converges to one of the two stable states, X^* or Y^*. The nullclines thus provide a summary of the qualitative dynamics of the system.

Suppose that the proportion of monolingual speakers of an endangered language X having status 0.2 has fallen to 30%, and that the community in which it is spoken is to intervene in order to bring about its maintenance. How should the community intervene and what is the corresponding effect on the nullclines? Fig. 7a shows the direction field for the system without intervention. For $x \leq 0.3$, almost all trajectories converge to Y^* (i.e., language X dies). Fig. 7b shows the impact of increasing the status of X from 0.2 to 0.4 whenever the proportion of monolingual speakers of X falls below 30%. This adjustment of the status of X leads to an increase in the slopes of both of the post-intervention nullclines, bringing about a change in the post-intervention dynamics. However, the post-intervention nullclines do not intersect, so no stable equilibrium is introduced. Thus, intervening to increase the status of the endangered language from 0.2 to 0.4 fails to achieve maintenance. Fig. 7c shows the impact of an alternative strategy for intervention: increasing the peak rate, c'_{ZY}, at which children of bilingual speakers acquire only language Y. Increasing the value of this parameter has the effect of increasing the proportion of monolingual speakers of Y. For this strategy, the slope of only one nullcline is increased ($dy/dt = 0$). Again, the post-intervention nullclines do not intersect, and intervention fails to bring about a stable state in which both languages are maintained.

Fig. 7d, however, shows the situation in which both the status of X and the peak transition rate, c'_{ZY}, are increased. In this case, the post-intervention nullclines are increased in slope sufficiently that they intersect at $x < 0.3$. Region $\mathbf{R_Y}$ (see Fig. 6b) of the pre-intervention state space abuts region $\mathbf{R_X}$ of the post-intervention state space. This is sufficient to introduce a stable equilibrium,
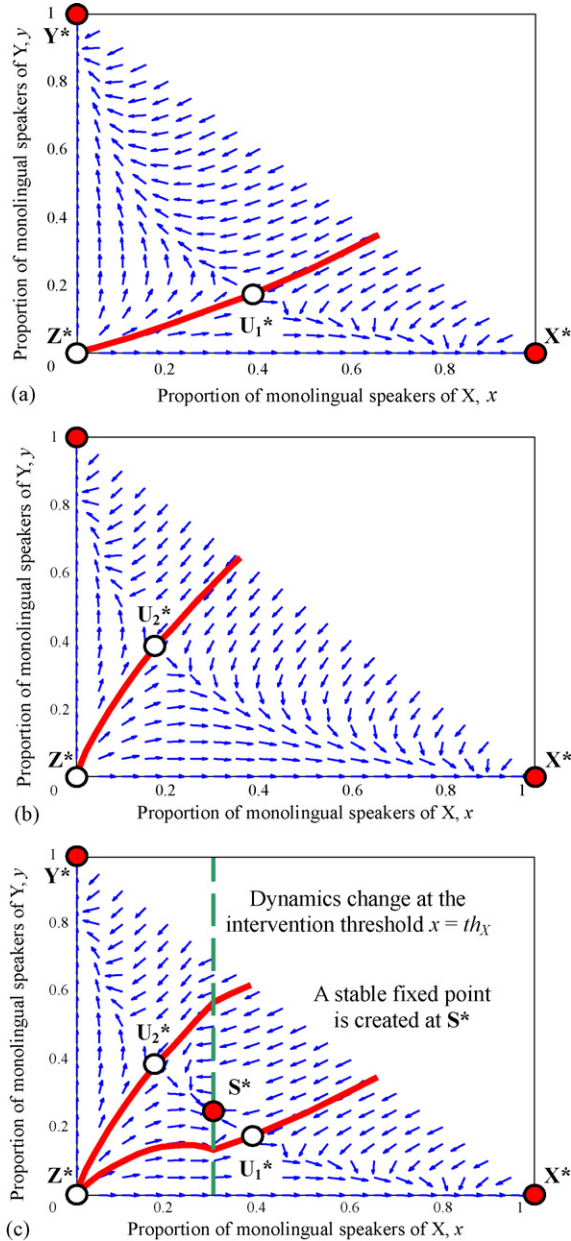
Fig. 5. Direction field of the system with intervention. Panels (a) and (b) show that, without intervention, language death is inevitable. Panel (c) shows the creation, as a result of intervention, of a new stable equilibrium, S^*, in which both languages are spoken. When the proportion of speakers of X exceeds the threshold, $x > th_X$, the dynamics follow that of the system without intervention with $s_X = 0.4$; panel (a). When the proportion of speakers of X falls below the threshold, $x < th_X$, the dynamics follow that of the system without intervention with $s_X = 0.6$; panel (b). Thick solid lines indicate the borders of the 'basin of attraction' of the stable equilibrium S^*—all initial states within a particular basin of attraction tend towards the associated equilibrium ($a = 1.0$; $\mu = 2\%$; $c_{XZ} = c_{YZ} = 0.035$; $c_{ZX} = c_{ZY} = 1.0$): (a) without intervention ($s_X = 0.4$); (b) without intervention ($s_X = 0.6$); and (c) with intervention ($s_X = 0.4$; $s'_X = 0.6$; $th_X = 0.3$).
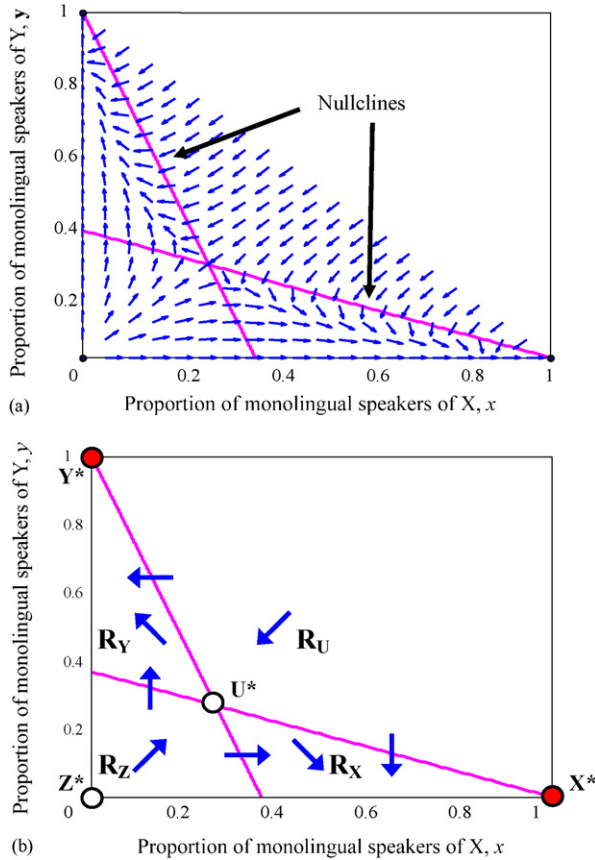
Fig. 6. Nullclines for the unified bilingual model. The nullclines segment the phase space into four regions, $\mathbf{R_X}$, $\mathbf{R_Y}$, $\mathbf{R_U}$ and $\mathbf{R_Z}$, each characterized by a different direction of change. Calculating the nullclines provides a summary of the qualitative dynamics of the system ($a = 1.0$; $\mu = 2\%$; $c_{XZ} = c_{YZ} = 0.035$; $c_{ZX} = c_{ZY} = 1.0$; $s_X = 0.5$): (a) nullclines and direction field and (b) nullclines and qualitative dynamics.

S^*, in which both languages are spoken, at $x = 0.3$, $y \approx 0.15$, $z \approx 0.55$—trajectories to the left of S^* lie in region $\mathbf{R_X}$ of the post-intervention state space and therefore approach S^* from the left, and trajectories to the right of S^* lie in region $\mathbf{R_Y}$ of the pre-intervention state space and therefore approach S^* from the right. The greater are the slopes of the nullclines, the larger are the corresponding regions $\mathbf{R_X}$ and $\mathbf{R_Y}$, which, in turn, enlarges the 'basin of attraction' of S^*, i.e., the range of initial states from which the system converges to S^*.

This behaviour points to a set of general principles that can guide strategies for language maintenance. Table 1 lists a set of mechanisms by which the slope of the post-intervention nullclines can be increased in order that they intersect, so enlarging the basin of attraction of the stable equilibrium, and thereby increasing the likelihood of successful maintenance. For a given intervention threshold, th_X, the likelihood of successful maintenance is increased by increasing the post-intervention values of parameters s'_X, c'_{ZX} and c'_{ZY}, and by decreasing the post-intervention values of parameters c'_{XZ} and c'_{YZ}. In other words, the status of the endangered language should be increased and the two languages isolated by encouraging monolingual education of children.
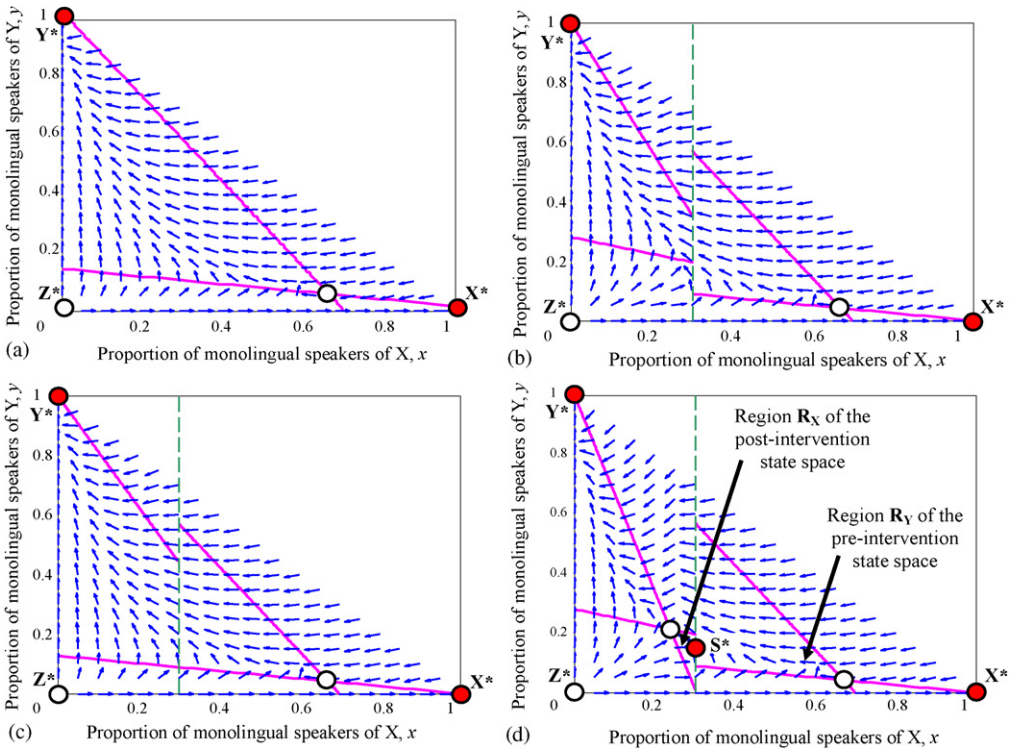
Fig. 7. Comparison of the impact of different strategies for language maintenance. Panel a shows a baseline situation with no intervention. In this example, controlling only the status of an endangered language (panel b: $s'_X = 0.4$) or only the peak transition rates (panel c: $c'_{ZY} = 0.070$) fails to bring about language maintenance, whereas controlling both the status and the peak transition rates can succeed (panel d: $s'_X = 0.4$; $c'_{ZY} = 0.070$). Regions $\mathbf{R_Y}$ of the pre-intervention state space and $\mathbf{R_X}$ of the post-intervention state space abut, creating a 'trapping region' that contains a stable equilibrium, S^*, at which both languages are spoken ($a = 1.0$; $\mu = 2\%$): (a) no intervention ($s_X = 0.2$; $c_{ZY} = 0.035$); (b) controlling status s_X only ($s_X = 0.2$; $s'_X = 0.4$; $th_X = 0.3$); (c) controlling peak transition rate c_{ZY} only ($c_{ZY} = 0.035$; $c'_{ZY} = 0.070$; $th_X = 0.3$); and (d) controlling both status and peak transition rate ($s_X = 0.2$; $s'_X = 0.4$; $c_{ZY} = 0.035$; $c'_{ZY} = 0.070$; $th_X = 0.3$).

4. An agent-based model of language competition

We now examine the dynamics of language competition from a different perspective, introducing an agent-based model that allows us to estimate the likelihood that the system

Table 1

Mechanisms to increase the slopes of the post-intervention nullclines, and thereby increase the likelihood of language maintenance

- *Increasing* the post-intervention status, s'_X, of endangered language X increases the slopes of the post-maintenance nullclines $dx/dt = 0$ and $dy/dt = 0$
- *Increasing* the post-intervention peak transition rate c'_{ZX} increases the slope of the post-maintenance nullcline $dx/dt = 0$
- *Increasing* the post-intervention peak transition rate c'_{ZY} increases the slope of the post-maintenance nullcline $dy/dt = 0$
- *Decreasing* the post-intervention peak transition rate c'_{XZ} increases the slope of the post-maintenance nullcline $dx/dt = 0$
- *Decreasing* the post-intervention peak transition rate c'_{YZ} increases the slope of the post-maintenance nullcline $dy/dt = 0$

converges to each stable state. The model that we have introduced in sections 2 and 3 is deterministic: given the initial state, its state at any future time is uniquely determined. The system allows, in principle, the proportion of speakers in each state to be calculated (even if one is not always able to establish an analytical expression for it). However, which particular speakers are bilingual at some time and which of them produce monolingual offspring, for example, are not addressed. The model does not trace the states of every single speaker, only the proportions of speakers having each state. Thus, it specifies a model of the *expected* behaviour of the competition, but not the *range* of behaviours that can result from a given initial state or their relative likelihoods.

Typically, the behaviour of an appropriately defined system of differential equations approaches the actual behaviour of the underlying system being modelled when the population size is large. However, the population sizes that are relevant in the context of language endangerment and maintenance are often of the order of hundreds or thousands of individuals. For such small populations, fluctuations in the language usage patterns of certain individuals may lead to dynamics that diverge significantly from the expected behaviour. Even when we consider the maintenance of an endangered or minority language having a considerable number of speakers, it may often be the case that many of the speakers live in small, relatively isolated communities or else form cliques within larger communities with which they have comparatively little interaction. Deterministic models based on systems of differential equations, such as (11), might be unable to capture the full range of possible behaviours of the underlying system.

In order to encapsulate variation caused by such factors, we adapt our model to investigate the stochastic nature of the dynamics of language competition. To do so, we implement the system as an agent-based simulation model, an approach that has found frequent application in language evolution studies (e.g. Hurford, 1989; Nowak et al., 1999; Wang et al., 2004). An agent-based simulation is a model in which discrete elements, called 'agents', represent selected entities or groups of entities of the underlying system that is being modelled. Unlike dynamical systems such as (11) in which the global behaviour of the entire system is determined by a single set of differential equations, the dynamics of agent-based models are described locally in terms of how individual agents interact with each other.

In the agent-based model of language competition that we introduce here, the agents correspond to the individual speakers that comprise the population being modelled. Each of the n agents adopts one of three possible states: monolingual in language X (state X), monolingual in language Y (state Y), or bilingual (state Z). Whereas the deterministic model is described in terms of formulae that specify the rates of change of the proportions of individuals having certain states, we design the agent-based model to use those same formulae to specify the probabilities with which each agent makes the transition from state to state. For example, in the deterministic model, Z-bilinguals following the V-Model produce X-monolingual offspring at rate $c_{ZX}\, s_X\, x^a$ (by Eq. (7)). In the agent-based model we re-interpret this to mean that Z-bilingual agents produce X-monolingual offspring with probability $c_{ZX}\, s_X\, x^a$, where x now denotes the proportion of an agent's neighbours who are X-monolingual. Throughout the current section we will assume that the agents are all neighbours of each other, modelling a population having a fully connected social network. However, the re-interpretation of the transition rates as probabilities allows us to model the impact of other social structures on the dynamics of language competition, an idea that we pursue in section 5.

The transition probabilities of the V-Model are given by:

$$\Pr(X \rightarrow X) = 1, \tag{14a}$$

$$\Pr(Y \rightarrow Y) = 1, \tag{14b}$$

$$\Pr(Z \to X) = c_{ZX} s_X x^a, \tag{14c}$$

$$\Pr(Z \to Y) = c_{ZY} s_Y y^a, \tag{14d}$$

$$\Pr(Z \to Z) = 1 - c_{ZX} s_X x^a - c_{ZY} s_Y y^a \tag{14e}$$

where x, y and z denote, respectively, the proportions of an agent's neighbours who are X-monolingual, Y-monolingual and Z-bilingual. All other transitions, e.g., X → Z and X → Y, have probability zero. The transition probabilities of the H-Model are:

$$\Pr(X \to X) = 1 - c_{XZ} s_Y y^a, \tag{15a}$$

$$\Pr(X \to Z) = c_{XZ} s_Y y^a, \tag{15b}$$

$$\Pr(Y \to Y) = 1 - c_{YZ} s_X x^a, \tag{15c}$$

$$\Pr(Y \to Z) = c_{YZ} s_X x^a \tag{15d}$$

$$\Pr(Z \to Z) = 1. \tag{15e}$$

All other transitions have probability zero. Agents undergo vertical transmission with probability μ, the mortality rate defined in section 2.2; otherwise, they undergo horizontal transmission.

The simulation is run as follows: The n agents that comprise the population are assigned initial states according to the selected initial proportions of speakers of each type. We denote the initial proportions of monolingual speakers of X and Y by x_0 and y_0, respectively. Having also specified the values of the other parameters, s_X, c_{ZX}, c_{ZY}, c_{XZ}, c_{YZ}, and μ, we set the simulation running iteratively. At each iteration, each agent samples the states of all its neighbours to determine its transition probabilities according to equations (14) and (15). Its state is then randomly updated accordingly using the roulette wheel procedure (Goldberg, 1989). Once the simulation has run for a specified number of iterations, we identify the global state that emerges: X-Monolingual (95% or more agents monolingual in X), Y-Monolingual (95% or more agents monolingual in Y), or Z-Bilingual (a mixture of both monolingual and bilingual agents).

4.1. Dynamics of the agent-based model

We now explain our approach to estimating the probability that the system converges to each global state by means of an example. Fig. 8a shows the evolution of the system during one run of the simulation for a population of 1,000 agents of whom 35% are initially X-monolingual and 60% Y-monolingual, the remainder being bilingual. In this run, the system quickly converges to the stable equilibrium S^* at $x \approx 0.30$, $y \approx 0.45$, $z \approx 0.25$ about which it then oscillates. This represents an endangered language X being maintained with about 30% monolingual and 25% bilingual speakers. Notice that the trajectory, also shown in the figure, follows the phase portrait of the deterministic system (11), on which it is superimposed, to a large degree, indicating that the behaviour of the stochastic, agent-based model has not diverged significantly from that of the deterministic system.
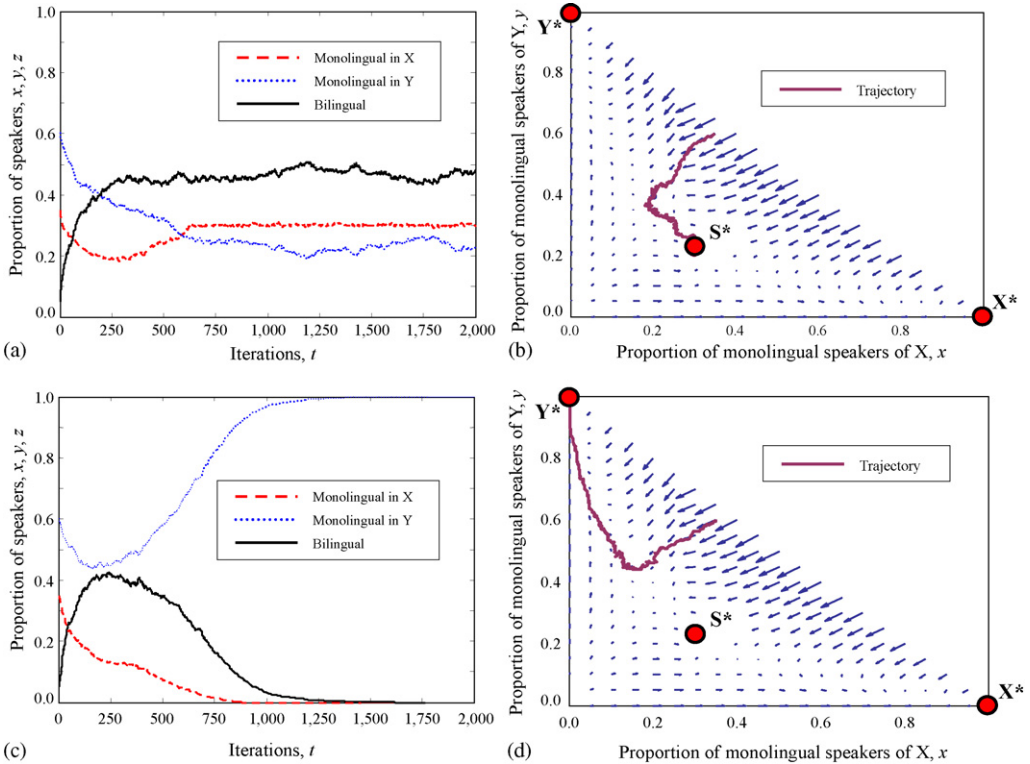
Fig. 8. Each set of initial conditions produces multiple possible outcomes. The figure shows the trajectories that ensue from the same set of initial conditions in two runs of the simulation. Not only do the trajectories differ, the system converges to different stable states ($a = 1.0$; $\mu = 2\%$; $c_{XZ} = c_{YZ} = 0.035$; $c_{ZX} = c_{ZY} = 1.0$; $s_X = 0.4$; $s'_X = 0.6$; $th_X = 0.3$). (a and b) In one run of the simulation, the system converges to a stable state at $x \approx 0.30$, $y \approx 0.45$, $z \approx 0.25$ after 600 iterations. Thereafter, the system oscillates about this state in which both languages are maintained. (c and d) In a second run, the system converges to a different stable state at $x = 0$, $y = 1$, $z = 0$. Language Y quickly acquires all speakers, and the endangered language X dies.

The same initial conditions, however, sometimes lead to the system converging to the state Y^* in which only language Y has any speakers, as shown for a second run in Fig. 8b. Despite converging to a different stable state, the trajectory again follows the direction field closely. This variation in the final state of the system is due to the fact that the initial state ($x_0 = 35\%$, $y_0 = 60\%$) lies close to the boundary between the basins of attraction of the stable equilibria Y^* and S^*. Slight perturbations away from the expected trajectory, indicated by the phase portrait, can lead to the system converging to stable equilibria that differ from that predicted by the deterministic system. By calculating the relative frequency of convergence to each stable equilibrium over many runs of the simulation, the likelihood that the system converges to each equilibrium can be estimated.

Our primary concern is to determine not only what intervention should be undertaken but also when intervention should be undertaken in order that an endangered language be maintained. Consider Fig. 9a, which shows the impact on a population of 1,000 agents, initially consisting of 80% X-monolinguals and 20% Y-monolinguals, of intervening to increase the status of X from $s_X = 0.2$ to $s'_X = 0.3$. The figure indicates that if the intervention is made after the proportion of
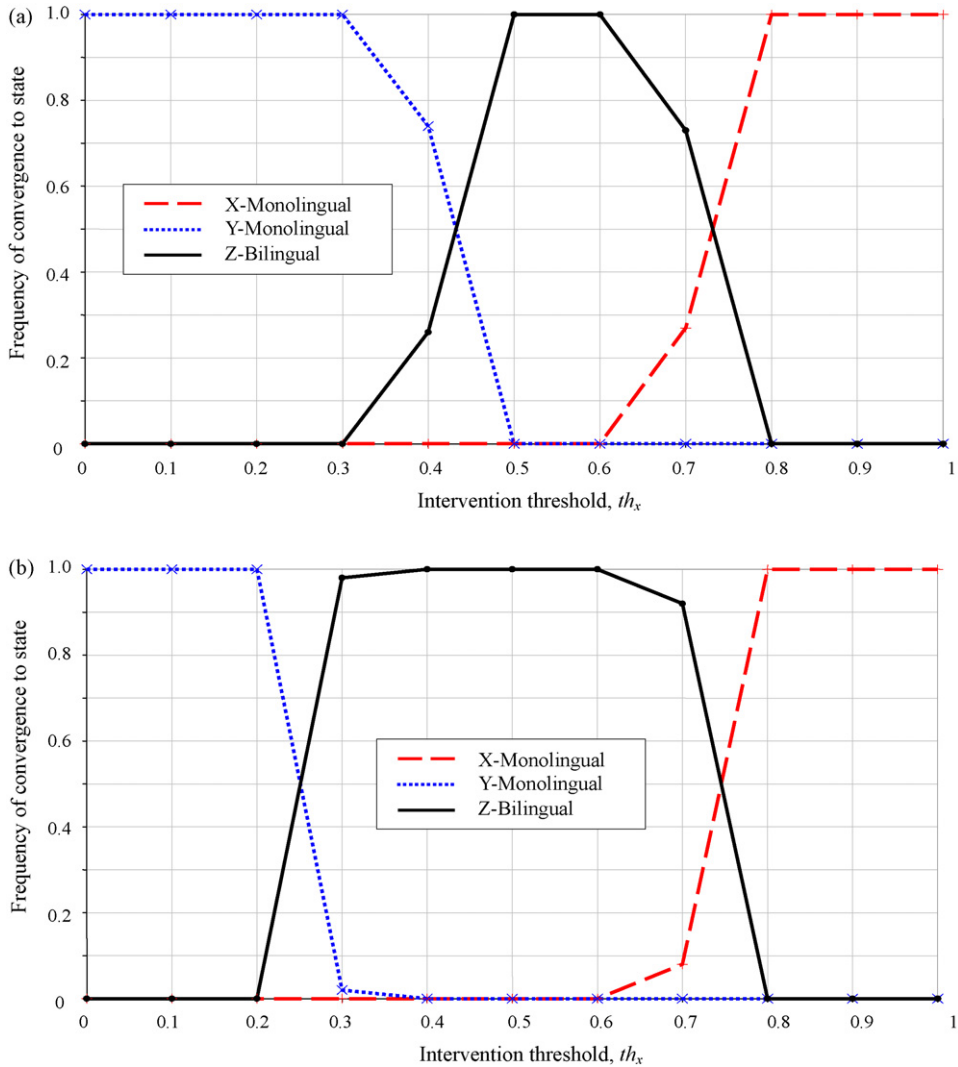
Fig. 9. The impact of the intervention threshold, th_X, and status, s'_X, on the frequencies of convergence. Increasing the post-intervention status, s'_X, of an endangered language X broadens the range of values of the intervention threshold for which a bilingual state is maintained with high frequency ($n = 1000$; $x_0 = 80\%$; $y_0 = 20\%$; $a = 1$; $\mu = 2\%$; $c_{XZ} = c_{YZ} = 0.035$; $c_{ZX} = c_{ZY} = 1.0$; 100 runs per point): (a) status: $s_X = 0.2$; $s'_X = 0.3$ and (b) status: $s_X = 0.2$; $s'_X = 0.4$.

monolingual speakers of X has fallen to 0.3 or lower ($th_X \leq 0.3$), then only language Y can be maintained. Maintenance of the two competing languages is possible if the intervention is made on the interval $0.3 < th_X < 0.8$, and is inevitable only for intervention on the interval $0.5 \leq th_X \leq 0.6$. Fig. 9b, however, shows the effects of intervention whereby the status of X is enhanced from $s_X = 0.2$ to $s'_X = 0.4$. In this case, both languages can be maintained with non-negligible probability on the interval $0.2 < th_X < 0.8$, the probability rising to 1 on the interval $0.4 \leq th_X \leq 0.6$.

We observe that increasing the post-intervention status of an endangered language broadens the range of values of the intervention threshold, th_X, for which the endangered language and the
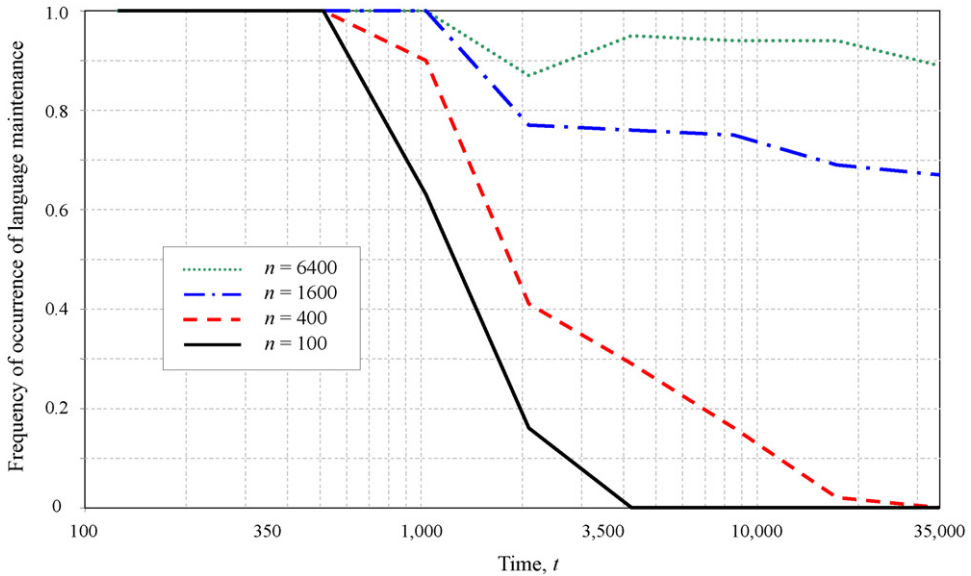
Fig. 10. Frequency of occurrence of language maintenance as a function of time for different population sizes ($x_0 = 80\%$; $y_0 = 20\%$; $a = 1$; $\mu = 2\%$; $c_{XZ} = c_{YZ} = c'_{XZ} = 0.035$; $c'_{YZ} = 0.070$; $c_{ZX} = c_{ZY} = 1.0$; $s_X = 0.2$; $s'_X = 0.4$; $th_X = 0.3$; 100 runs per point).

language with which it competes are both maintained. In particular, the lower bound on th_X is decreased but the upper bound undergoes no significant change. We observe the same qualitative behaviour for other sets of parameter values. This suggests that the more efficiently a community can bring about an increase in the status of an endangered language—or, more generally, implement any of the strategies proposed earlier in Table 1 —the later such intervention may take place in order that maintenance be achieved.

Even when such intervention is expected ultimately to fail, the extinction of an endangered language is not immediate. The rate at which a language loses speakers depends on the values of the system parameters. Fig. 10 exemplifies the impact of population size, n, plotting the frequency of language maintenance as a function of time for four different values of population size. For all populations sizes, both languages are maintained within 500 time steps. As t is increased beyond 500 time steps, however, the frequency of maintenance diminishes for all population sizes, the fall in frequency being greater for smaller populations. We observe the same qualitative behavior for other sets of parameter values. We therefore conclude that the probability of maintenance over a fixed duration for a fully connected population increases with population size. This result does not imply, however, that the probability of maintenance necessarily increases with population size for arbitrary social structures. The behavior of the system for qualitatively different social structures must be analyzed case by case. In the following section, we introduce the method by which social structures other than the fully connected can be modelled.

5. The impact of social structure on the probability of language maintenance

Our simulations in the previous section were based on the assumption that each agent has complete knowledge of, and is influenced by, the states of all other speakers in the population

undergoing language competition. In effect, this is equivalent to assuming that the underlying social structure is fully connected. Here, we investigate the impact of other selected social structures on the probability of language maintenance. We represent social structure as a network whose nodes represent the individual speakers comprising the population. The edges of the network connect individuals who communicate with each other. In particular, we model the social structure as a local-world network (Li and Chen, 2003). Local-world networks integrate into a single paradigm both the random networks of Erdős and Rényi (1959) and the scale-free networks of Barabási and Albert (1999), which, together with 'small-world' (Watts and Strogatz, 1998) and other network structures, are now finding application in studies of social and sociolinguistic systems (e.g., Moody, 2001; Smith, 2002; Ohtsuki et al., 2006; Castelló et al., 2006).

Local-world networks are constructed recursively, adding nodes to the network one at a time. As with scale-free networks, they are constructed using preferential attachment: when a node is added to the network, it is assigned a greater probability of being connected to extant nodes having numerous connections than to nodes having few connections. This reflects an assumption that individuals prefer to interact with those speakers who themselves interact with many other speakers. Unlike in scale-free networks, however, when a node is added to a local-world network it is connected preferentially only to nodes within a randomly selected subset of all extant nodes, its 'local-world'. Thus individuals have local, rather than global, knowledge of the language usage patterns of other speakers in the population and only interact with a fraction of the other speakers comprising the population. The procedure we use for constructing local-world networks is described in Appendix A.

We now analyze the effect of the initial proportions of monolingual speakers on the probability of maintenance, starting with fully connected social structures. Fig. 11 shows the behaviour of the system for a population of 1,000 agents, with parameter values set to $a = 1.0$ and $s_X = 0.4$, with no intervention. The figure plots the estimated probability of convergence to each state as a function of the initial proportion of monolingual speakers of X; the remaining speakers
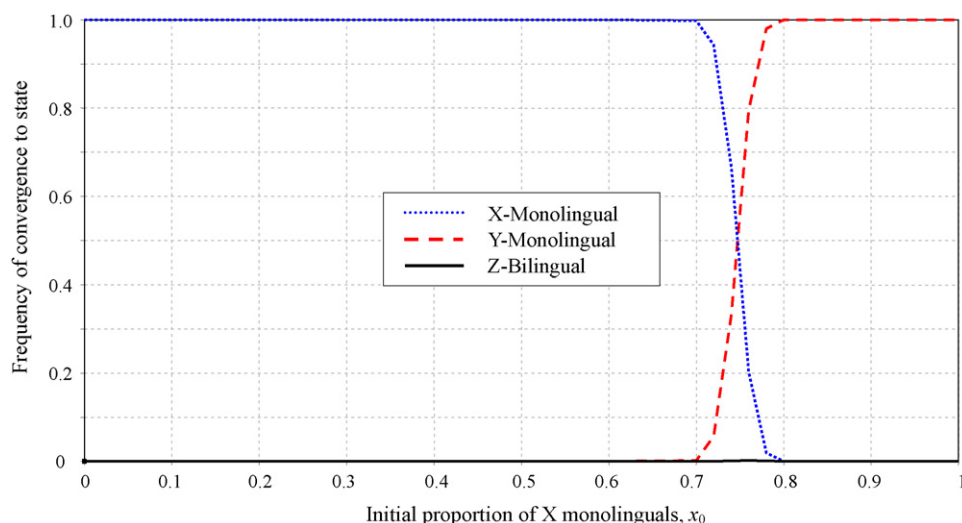

Fig. 11. The impact of the initial proportion of endangered language monolinguals on the frequencies of convergence for a fully connected population without intervention ($n = 1,000$; $a = 1$; $s_X = 0.4$; 100 runs per point).

initially all speak Y. The figure clearly indicates that a state in which both languages are maintained can be achieved only with negligible probability. Furthermore, for most initial proportions of speakers, $x \leq 0.7$ or $x \geq 0.8$, the system behaves in the same manner as the deterministic system (11), with just one language acquiring all speakers with probability 1. However, when the initial proportion of monolingual speakers of X lies in the range $0.7 \leq x \leq 0.8$, there is a gradual transition in the probabilities. This transition reflects the stochastic aspect of the interactions among the agents, this effect being particularly prominent when the initial conditions of the system are located close to the boundary between two basins of attraction (as was the case for the simulation illustrated previously in Fig. 8). As the population size increases, so the transition zone contracts, and the behaviour of the system converges to that of the deterministic system (11) for all initial conditions.

For a locally connected population, the behaviour is indistinguishable from that of the fully connected network shown in Fig. 11. We observe the same lack of impact of social structure on the probabilities of convergence for other values of parameters n, a and s_0. We therefore conclude that, in the absence of intervention, local-world social structure has no significant influence on which language is maintained and which language dies.

When the population intervenes to attempt to maintain both competing languages, however, we find that the underlying social structure does affect the behaviour. Fig. 12 shows examples of the behaviour for a population of 1,000 agents that intervenes to increase the status of the endangered language X from 0.4 to 0.6 whenever the proportion of monolingual speakers of X falls below 0.5. Fig. 12a highlights the behaviour for a fully connected population, clearly indicating the range of initial proportions of monolingual speakers of the endangered language that allow both languages to be maintained, $0.2 \leq x_0 \leq 0.8$, with maintenance being virtually certain for $0.3 \leq x_0 \leq 0.7$. The same qualitative behaviour, in which a single stable state emerges with probability ~1 over a broad range of initial conditions, is observed for other values of parameters n, a and s_0 for a fully connected population.

For a locally connected population, however, the behaviour is less regular. Fig. 12b shows the graph for a local-world size of 50 agents, with the number of connections between each new node and its local-world set to 20. We observe that the range of values for which both languages can possibly be maintained is the same as for the fully connected population: $0.2 \leq x_0 \leq 0.8$. However, the peak probability is somewhat less than 90% (for $x_0 = 0.3$). The probability of maintenance decays gradually for larger initial proportions of monolingual speakers of the endangered language to about 50% for $x_0 = 0.7$. The probability then decays rapidly to zero as x_0 approaches 0.8, as is the case for the fully connected population. From this behaviour we infer that maintenance is more difficult to achieve within societies having an underlying local-world structure. Furthermore, the probability of maintenance appears to be maximal when intervention is undertaken "at the last moment", but not so late that the opportunity is missed. Intervention is best implemented when the state of the system is closest to the position of the stable equilibrium S^* that would be introduced by such enhancement; doing otherwise increases the probability that the system diverges from this equilibrium.

These results suggest that for initial conditions in which death of the endangered language is not inevitable, the probability of maintenance is greater for fully connected populations than for locally connected populations. One reason why this is so may be that for the fully connected population, the impact of a randomly selected individual on the evolution of the entire population is generally the same as any other. But for a locally connected population, there are a few particular individuals, called 'hubs', that connect to many more other individuals than is typical, and so exert greater influence over the evolution of the entire population. Once these hubs have
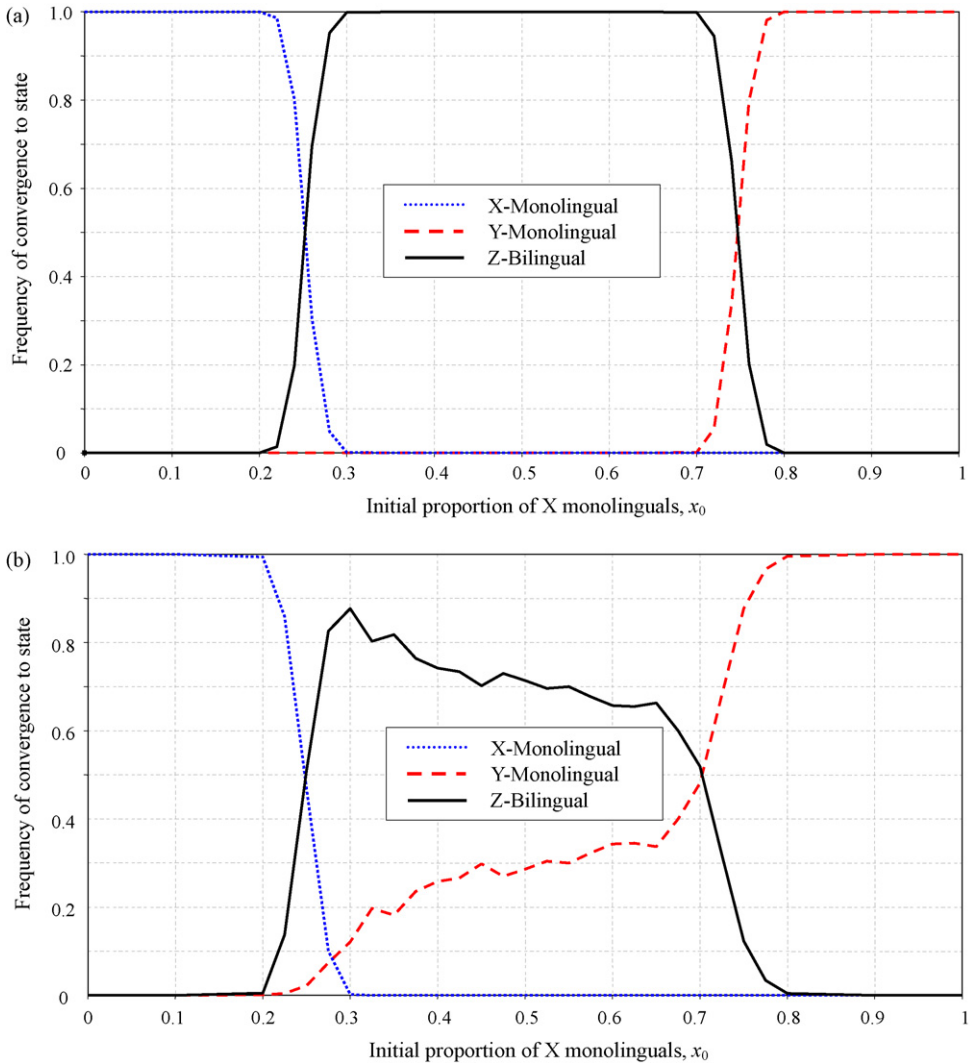
Fig. 12. The impact of the initial proportion of endangered language monolinguals and the social structure on the frequencies of convergence for a population with intervention ($n = 1000$; $a = 1$; $s_X = 0.4$; $s'_X = 0.6$; $th_X = 0.5$; 100 runs per point): (a) fully connected social structure. (b) locally connected social structure ($n_{LW} = 50$; $e_{LW} = 20$).

acquired a monolingual state, the convergence to a global monolingual state over the entire population is hastened. Another reason may be that individuals within a fully connected population sample the states of a greater number of neighbouring individuals than within a locally connected population. As a result, variance in the behaviour of the locally connected population is greater, resulting in rapid transitions toward the death of the endangered language occurring more frequently for locally connected population than for fully connected populations.

We have yet to investigate language maintenance on social structures that differ qualitatively from the fully connected or locally connected network models discussed above. However, the model imposes no restrictions on the social structure other than that it can be represented as a

network of nodes, representing the members of the population, and edges, representing the communicative interactions among members. It is possible, therefore, to model language competition on a population comprised of two distinct, weakly interacting sub-populations, the sub-populations modelled as two separate networks having sparse inter-connectivity. Doing so may shed light on the empirical observation that small, culturally isolated populations can often maintain a minority language for many generations, one example being the maintenance of Pennsylvania German among the Mennonite (Burridge, 1997), Amish (Huffines, 1980) and other Anabaptist groups in North America since their arrival from Europe in the 17th century.

Until the predictions of the model have been fitted to empirical data, we hesitate to claim that the probabilities quoted here indicate precisely the likelihood of language maintenance being achieved within an actual community. However, we do maintain that comparison of the estimated probabilities of convergence for different sets of parameter values and initial conditions reveal the qualitative behaviour of the system, and so inform us how better to intervene in order that an endangered language be maintained.

6. Discussion

We have introduced an extension of the Abrams and Strogatz (2003) model for language competition in which we explicitly model bilingualism and social structure. In the absence of intervention, the qualitative dynamics of the system are, in most cases, identical to those of the Abrams and Strogatz system—when two languages compete for speakers, eventual extinction of one language is inevitable. However, we have shown that by appropriate increase in the status of an endangered language or by adjusting the availability of educational resources in each of the competing languages, for example, the dynamics can be altered such that both languages are maintained with non-negligible probability. Such intervention should be undertaken within a certain time window, enhancing the viability of the endangered language before it becomes moribund. For all but the simplest social structures that we have modelled, the peak probability of successful maintenance is obtained by implementing the maintenance strategy as late as possible. We have yet to demonstrate, however, that the dynamics of the model presented here fit the empirical data any better than those of others models, such as those developed by Abrams and Strogatz (2003), and Mira and Paredes (2005). This task remains a necessary future step.

Nevertheless, the model introduced here does account for a number of factors that influence language competition and maintenance, including language status, proportions of speakers, population size, and the availability of monolingual and bilingual educational resources. The model is sufficiently flexible that the sociolinguistic particulars of the environment in which language competition takes place can be modelled by appropriate selection of the model parameters and network structure. For example, situations in which two or more communities interact, each initially having distinct language patterns of its own, may be conveniently modelled by merging multiple networks, one for each community. One could then examine how the degree of isolation of an endangered language impacts upon its viability. In this way, the competition between Pennsylvania German and English among Anabaptist communities within North America could be modelled by considering the impact of communicative interactions between one sub-population comprising both German monolinguals and German-English bilinguals, representing an Anabaptist community, and another sub-population comprising mainly English monolinguals, representing that community's neighbours, between which there is only sparse interconnectivity.

A number of aspects of the model proposed here can be refined. We have not explicitly modelled the geographic distribution of the individuals comprising a population. Network

models have been proposed that account for the distances between the nodes making up the network (e.g., Kleinberg, 2000), an approach that could be applied to model the geographic distribution of speakers. However, the geographic distribution of speakers within a population is not necessarily an indication of the likely social connections that obtain among them, particularly in urban environments. A recently proposed modification of the local-world paradigm (Gong et al., 2004), in which the local-world of each node is re-assessed at each time step, may serve to model the dynamic aspects of interaction among speakers within a community, obviating the need to explicitly model geographic distribution.

We have made no attempt to model code-switching, which often leads to an endangered language adopting features of the language with which it competes. Rather, we have assumed implicitly that such language shift, if it occurs, does not impact the attractiveness of a language. Code-switching and language shift might be incorporated into the model by treating the languages as consisting of multiple components, e.g. the lexicon and the syntax, each having its own status and attractiveness, and each of which may be learned independently by each speaker. The loss of individual components may then be traced as the system evolves. It is likely that such a model would predict the fall of an endangered language, slowly at first as the first few components of the endangered language come to be replaced by their more prestigious counterparts, and then increasingly rapidly as the remnants of the dying language vanish. As the two competing languages come to resemble each other ever more closely, their degree of mutual intelligibility would increase. A first step toward modelling the impact of code switching and language shift might therefore be to introduce into the system (11) an additional variable that measures the degree of mutual intelligibility (cf. parameter k in the model of Mira and Paredes, 2005), and which varies as a function of the proportion of components that the competing languages have in common.

Acknowledgements

Appendix A. Algorithm for constructing local-world networks

We use the following recursive algorithm to construct local-world networks (Li and Chen, 2003). The parameters of the algorithm comprise:

- the population size, n, which corresponds to the number of nodes in the network;
- the initial proportions of agents monolingual in X, x_0, and Y, y_0;
- the size of the local-world, n_{LW};
- the number of nodes in the local world to which each new node is connected, e_{LW}.
The algorithm proceeds as follows:

1. Build a random network consisting of n_{LW} nodes, starting with a network consisting of a single initial node.
   a. Add a new node.
   b. Connect the new node to one randomly selected existing node.
   c. Repeat steps a–b until the network consists of n_{LW} nodes.

2. Add the remaining $(n - n_{LW})$ nodes to the network by preferential attachment:
   a. Add a new node.
   b. Randomly select n_{LW} existing nodes to be the local-world of the new node.
   c. Connect the new node to eLW nodes within its local-world by preferential attachment, i.e., the node connects to eLW nodes in its local-world with probability proportional to their degree (i.e., the number of nodes to which they are already connected).
   d. Repeat steps a–c until the network consists of n nodes.
3. Randomly assign the state of each node according to the initial proportions of monolingual speakers of each language, x_0 and y_0.

   Scale-free networks can be constructed using the same algorithm by setting the size of the local-world to the total number of nodes that exist in the network at each step.

References

Abrams, D.M., Strogatz, S.H., 2003. Modeling the dynamics of language death. Nature 424, 900.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 285 (5439), 509–512.

Bradley, D., Bradley, M. (Eds.), 2002. Language Endangerment and Language Maintenance, Routledge Curzon, London/New York.

Burridge, K., 1997. Separate and peculiar—the survival of Pennsylvania 'Dutch' in Ontario, Canada. In: Foster, L., Burridge, K., McNamara, L., Turcotte, G. (Eds.), Canada-Australia, 1895–1995: Towards a Second Century of Partnership. Carleton University Press, Ottawa.

Castelló, X., Eguíluz, V. M., San Miguel, M., 2006. Ordering dynamics with two non-excluding options: Bilingualism in language competition. Published online at arXiv.org (http://arxiv.org/abs/physics/0609079).

Crystal, D., 2000. Language Death. Cambridge University Press, Cambridge.

Erdős, P., Rényi, A., 1959. On random graphs. Publicationes Mathematicae Debrecen 6, 290–297.

Fishman, J.A., 1991. Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages. Multilingual Matters, Clevedon, Philadelphia.

Fishman, J.A. (Ed.), 2001. Can Threatened Languages Be Saved? Reversing Language Shift Revisited: A 21st Century Perspective. Multilingual Matters, Clevedon; Buffalo.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.

Gong, T., Ke, J., Minett, J.W., Wang, W.S.-Y., 2004. A computational framework to simulate the coevolution of language and social structure. In: Pollack, J., Bedau, M., Husbands, P., Ikegami, T., Watson, R.A. (Eds.), Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems. MIT Press, pp. 158–163.

Grenoble, L.A., Whaley, L.J. (Eds.), 1998. Endangered Languages: Current Issues and Future Prospects. Cambridge University Press, Cambridge/New York.

Grenoble, L.A., Whaley, L.J., 2006. Saving Languages: An Introduction to Language Revitalization. Cambridge University Press, Cambridge/New York.

Huffines, M.L., 1980. Pennsylvania German: maintenance and shift. International Journal for the Sociology of Language 25, 43–57.

Hurford, J.R., 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. Lingua 77, 187–222.

Kleinberg, J.M., 2000. Navigation in a small world: It is easier to find short chains between points in some networks than others. Nature 406, 845.

Krauss, M., 1992. The world's languages in crisis. Language 68 (1), 4–10.

Li, X., Chen, G., 2003. A local-world evolving network model. Physica A 328 (1–2), 274–286.

Mira, J.B., Paredes, Á., 2005. Interlinguistic similarity and language death dynamics. Europhysics Letters 69 (6), 1031–1034.

Moody, J., 2001. Race, school integration, and friendship segregation in America. American Journal of Sociology 107, 679–716.

Mufwene, S., 2004. Language birth and death. Annual Review of Anthropology 33, 201–222.

Nettle, D., Romaine, S., 2000. Vanishing Voices: The Extinction of the World's Languages. Oxford University Press, Oxford/New York.

Nowak, M.A., Plotkin, J.B., Krakauer, D.C., 1999. The evolutionary language game. Journal of Theoretical Biology 200, 147–162.

Ohtsuki, H., Hauert, C., Lieberman, E., Nowak, M.A., 2006. A simple rule for the evolution of cooperation on graphs and social networks. Nature 441, 502–505.

Pagel, M., 1995. Contribution to the Conservation of Endangered Languages seminar, University of Bristol, 21 April 1995. Iatiku 1, 6.

Patriarca, M., Leppänen, T., 2004. Modeling language competition. Physica A 338 (1–2), 296–299.

Schulze, C., Stauffer, D., 2006. Recent developments in computer simulations of language competition. Computing in Science and Engineering 8 (3), 60–67.

Smith, L.R., 2002. The social architecture of communicative competence: a methodology for social-network research in sociolinguistics. International Journal of the Sociology of Language 153, 133–160.

Stauffer, D., Castello, X., Eguíluz, V. M., San Miguel, M., 2006. Microscopic Abrams-Strogatz model of language competition. Available online at arXiv.org (http://arxiv.org/abs/physics/0603042).

Strogatz, S.H., 1994. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Addison-Wesley, Reading, MA.

Wang, W.S.-Y., Minett, J.W., 2005. The invasion of language: emergence, change and death. Trends in Ecology and Evolution 20 (5), 263–269.

Wang, W.S.-Y., Ke, J., Minett, J.W., 2004. Computational studies of language evolution. In: Huang, C.-R., Lenders, W. (Eds.), Computational Linguistics and Beyond: Perspectives at the beginning of the 21st Century. Frontiers in Linguistics, vol. 1. Language and Linguistics, Nankang, pp. 65–108.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world networks. Nature 393, 440–442.

Further reading

De Oliveira, S.M., Sá Martins, J.S., de Oliveira, P.M.C., Luz-Burgoa, K., Ticona, A., Penna, T.J.P., 2004. The Penna model for biological aging and speciation. Computing in Science and Engineering 6 (3), 74–81.