A STOCHASTIC MODEL OF LANGUAGE CHANGE THROUGH PREDICTION-DRIVEN INSTABILITY

1. Introduction

Language change is paradoxical: Children acquire their native language accurately, yet over time, the language can change, apparently even without external influence. A possible resolution to this paradox comes from the facts that language is flexible and children do not learn from the entire population equally. Individuals vary in how they use their native language, and chance local fluctuations in such variation might be enough to trigger a language change.

As we will see in Section 2, a mean-field model in which children learn from the entire population equally does not lead to spontaneous change, even in the presence of random variation. However, in Section 3 we will discuss a modification of the model, in which children can detect age-correlated patterns in variation. This does exhibit spontaneous language change: Children can detect accidental correlations between age and speech, predict that the population is about to undergo a language change, and accelerate the process, a process we will call *prediction-driven instability.*

2. A mean field model

Let us suppose, for the sake of simplicity, that individuals have the choice between two grammars G_1 and G_2 when forming sentences, and that each individual has a particular usage rate, that is he uses G_2 in forming a fraction x of sentences, and G_1 for the rest. As a specific example, consider the formation of questions in Late Middle and Early Modern English. We could take G_1 to be verb-raising, and G_2 to be *do*-support. Manuscripts exist that use both at a variety of rates, and this formulation can model such variation. Since verb-raising and *do*-support both exhibit a certain stability, we should seek a model with two stable states, one for a population that prefers G_1 and a second for G_2.

Initially, we might consider a large unstructured population, in which children learn from all individuals equally and therefore hear essentially the mean usage rate. The simplest learning model with the desired bi-stability is a differential equation for the time-dependent mean usage rate $m(t)$ in the population,

$$\dot{m} = q(m) - m \tag{1}$$

where $q(m)$ is the learning function. Specifically, $q(m)$ is the mean usage rate of children learning from a population that uses G_2 with a mean rate m. The first term represents birth and learning, and the second represents death. This model is deterministic and has two stable equilibrium states. There is no way for it to spontaneously switch grammars.

To add random fluctuations, we switch to a discrete Markov chain model. We assume that the population consists of N individuals, each of which is one of $K+1$ types, numbered 0 to K, where state j means that the individual uses G_2 at a rate j/K. The state of the chain at time t is a vector $Y(t)$ whose j-th element $Y_j(t)$ is the number of individuals of type j in the population. The mean usage rate at time t is therefore

$$M(t) = \sum_{j=0}^{K} \left(\frac{j}{K}\right) Y_j(t) \tag{2}$$

The transition process from $Y(t)$ to $Y(t+1)$ is as follows. With some probability, one individual is selected and removed, to simulate death. A replacement individual is created and its type is selected at random based on a discrete distribution vector $Q(M(t))$. That is, $Q_j(m)$ is the probability that a child learning from a population with mean usage rate m is of type j, and therefore uses G_2 at a rate j/K. See Figure 1.

If the population is large, then the behavior of $M(t)$ can be approximated by the solution $m(t)$ to the deterministic differential equation (1). If Q is defined properly, then the resulting Markov chain will be ergodic, meaning that it must visit every possible state eventually. Thus, the model spends most of its time hovering near an equilibrium dominated by one grammar or the other, but it must eventually exhibit spontaneous language change by switching to the other equilibrium.

However, computer experiments show that under this model, a population takes an enormous amount of time to switch dominant grammars. It is therefore unsuitable for modeling language change on historical time scales. A further undesirable property is that if a population does manage to shift to an intermediate state, it is equally likely to return to the original grammar as to complete the shift to the other grammar. Historical studies show that language changes typically run to completion and do not reverse themselves, so again this model is unsuitable.

3. An age-structured model

To remedy the weaknesses of the mean-field model, we introduce social structure into the population. According to sociolinguistics, ongoing language change is reflected in social variation, so there is reason to believe children are aware of socially correlated speech variation and use it during acquisition.

There are many ways to formulate such a model, and not all formulations apply to all societies. For simplicity, we assume that there are two age groups, roughly representing parents and grandparents, and that children can detect systematic differences in their speech. We also assume that there are social forces leading children to avoid sounding out-dated. To represent the population at time t, define $V_j(t)$ to be the number of parents of type j, and define $W_j(t)$ to be the number of grandparents of type j. We also assume that apart from age, children make no distinction among individuals. Thus, they learn essentially from the mean usage rates of the two generations,

$$M_V(t) = \sum_{j=0}^{K} \left(\frac{j}{K}\right) V_j(t) \quad M_W(t) = \sum_{j=0}^{K} \left(\frac{j}{K}\right) W_j(t) \tag{3}$$

The modified transition process from $(V(t), W(t))$ to $(V(t+1), W(t+1))$ is as follows. With some probability, a grandparent is removed to simulate death, and a replacement individual is selected from the parents to simulate aging. A new parent is created based on the discrete probability vector $Q_2(M_V(t), M_W(t))$. Here, $Q_2(v, w)$ represents the acquisition process, together with prediction: Children hear that the younger generation uses G_2 at a rate v, and the older generation uses a rate w. They predict that their generation should use a rate

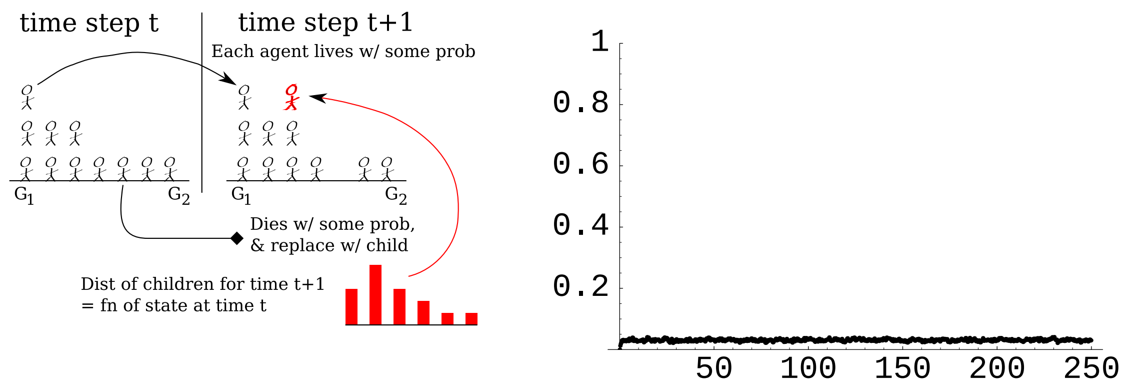

FIGURE 1. The basic Markov chain. Left: Diagram of the transition function. Right: A plot of the mean usage rate $M(t)$ of G_2 from a sample path.

determined by any trend and learn based on that predicted target value. If the prediction is given by $r(v, w)$, then $Q_2(v, w) = Q(r(v, w))$. See Figure 2.

This model turns out to exhibit the desired properties. The population can spontaneously change from one language to the other and back within a reasonable amount of time, and once initiated the change runs to completion without turning back.

To understand why this happens, we approximate the Markov chain by a system of deterministic differential equations governing the mean usage rates v and w of the two generations,

$$\begin{aligned} \dot{v} &= q(r(v, w)) - v \\ \dot{w} &= v - w \end{aligned} \tag{4}$$

where $q(r)$ is the mean of $Q(r)$. The phase space of this dynamical system is a square, and it happens to have two stable equilibria representing populations where both generations are dominated by one grammar or the other. Each such equilibrium has a basin of attraction. Populations in the basin flow toward the equilibrium and settle there. The boundary between the two basins is called the separatrix, and in this case, the separatrix passes very close to the stable equilibria. See Figure 3. The Markov chain model will hover near one equilibrium or the other, but since it incorporates random fluctuations, it is possible for the population state to spontaneously stray across the separatrix, where it will be blown toward the other equilibrium.

Intuitively, the mechanism of these spontaneous changes is that every so often, children pick up on an accidental correlation between age and speech. The prediction step in the acquisition process amplifies the correlation, and moves the population away from equilibrium. We therefore coin the term *prediction-driven instability* for this effect.

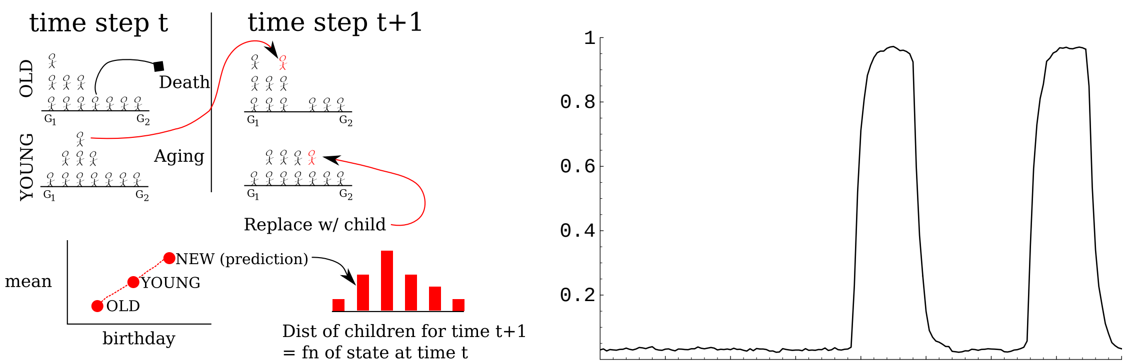

FIGURE 2. The age-structured Markov chain. Left: Diagram of the transition function. Right: A plot of the mean usage rate $M_V(t)$ of G_2 in the younger generation from a sample path.

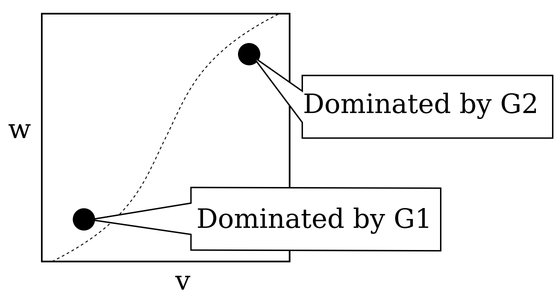

FIGURE 3. Phase portrait for (4). The two dots represent stable equilibria, and the dashed curve is the separatrix between their basins of attraction.