A STOCHASTIC MODEL OF LANGUAGE CHANGE THROUGH SOCIAL STRUCTURE AND PREDICTION-DRIVEN INSTABILITY

1. Introduction

Language change is paradoxical: Children acquire their native language accurately, yet over time, the language can change. Some changes may be attributed to an external event, such as political upheaval, but not every instance of language change seems to have an external cause. A possible resolution to this paradox comes from the facts that (1) spoken language has inherent variation and (2) children may take social structure into account, giving more or less weight to speech patterns correlated with age and social status. 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 3, 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 4 we will discuss a simple modification of the model, in which children can detect age-correlated patterns in variation. When subject to random fluctuations, this model 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 change, a process I will call *prediction-driven instability.*

2. The example of question syntax in English

Language has both discrete and continuous characteristics. On the discrete side, most sentences are clearly either grammatical or ungrammatical. Most linguistic formalisms, such as government-binding, minimalism, and optimality theory, are designed to describe such idealized grammars. However, true speech has significant variability, and individuals may bend the rules of the underlying idealized grammar in a variety of ways. A spoken language may be described as a collection of similar idealized grammars that speakers draw from at variable rates depending on context.

Let us suppose, for the sake of simplicity, that individuals have the choice between two similar idealized grammars, G_1 and G_2, when forming sentences, and that each individual has particular fixed usage rates, that is he uses G_2 in forming a fraction x of spoken sentences, and G_1 in forming the rest. As a specific example, consider the syntax of questions in Late Middle and Early Modern English. We take G_1 to be idealized English grammar with verb-raising syntax, and G_2 to be a similar grammar but with *do*-support:

(1) Know you what time it is? (verb-raising, G_1)

(2) Do you know what time it is? (*do*-support, G_2)

Manuscripts exist that use both at a variety of rates [1]. The formulation proposed here can represent such variation.

However, language acquisition is now more complicated: Children must learn multiple idealized grammars, plus the usage rates. Since verb-raising and *do*-support both exhibit stability over certain time scale, we should seek a model of learning within a population that has two stable states, one representing populations that prefer G_1 and a second representing populations that prefer G_2. To represent a language change from G_1 to G_2, the model must be able to switch from one stable state to the other over large time scales, while remaining steady over short time scales.

3. A mean field model

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,

$$\dot{m} = q(m) - m \tag{3}$$

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 $q(m)$ term represents birth and learning, and the $-m$ term represents death. The term *mean field* refers to the fact that the population's influence on an individual is represented by a single aggregate property, in this case, the mean usage rate of G_2. This model is deterministic and has two stable equilibrium states. However, there is no way for it to spontaneously switch grammars.

To add the possibility of a language change, we formulate a Markov chain model for a finite population, thereby adding random fluctuations. See Figure 1. We assume that the population consists of N adults, each of which is one of $K + 1$ types, numbered 0 to K, where type 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

$$M(t) = \sum_{j=0}^{K} \left(\frac{j}{K}\right)\left(\frac{Y_j(t)}{N}\right) \tag{4}$$

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 when she grows up. This Markov chain model maintains the mean field assumption because the population influences language acquisition only through the mean usage rate of G_2.

If the population is large, then the behavior of $M(t)$ can be approximated by the solution $m(t)$ to the deterministic differential equation (3). 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. See Figure 1 for a graph of the mean usage rate of G_2 as a function of time for a typical run of this Markov chain. This model is therefore unsuitable for simulating 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 [1, 3] show that language changes typically run to completion and do not reverse themselves, so again this model is unsuitable.

4. An age-structured model

To remedy the weaknesses of these mean-field models, 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 [2].

There are many ways to formulate an age-structured population, 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.

Let us adapt the Markov chain from Section 3 to include age structure. See Figure 2. 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. The total number of parents is N_V and the total number of grandparents is N_W. 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,

$$M_V(t) = \sum_{j=0}^{K} \left(\frac{j}{K}\right)\left(\frac{V_j(t)}{N_V}\right) \quad M_W(t) = \sum_{j=0}^{K} \left(\frac{j}{K}\right)\left(\frac{W_j(t)}{N_W}\right) \tag{5}$$

Here we have modified the mean-field assumption by representing the influence of the population on a child with two aggregate quantities. 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 determined by any trend and learn based on that predicted target value. If the prediction is given by $r(v, w)$, then $Q_2(v, w) = Q(r(v, w))$.

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. See Figure 2 for a graph of the mean usage rate of G_2 among the younger age group as a function of time for a typical run of this Markov chain.

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

$$\begin{aligned} \dot{v} &= q(r(v, w)) - v \\ \dot{w} &= v - w \end{aligned} \tag{6}$$

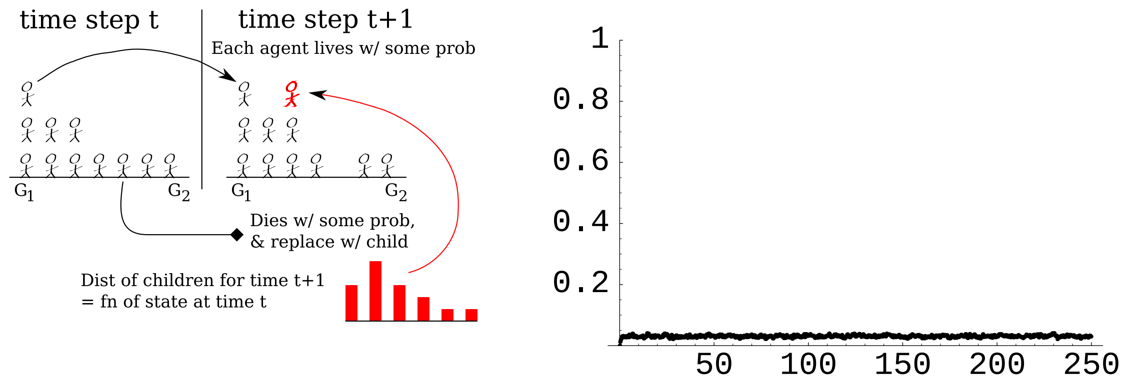

FIGURE 1. The basic Markov chain. Left: Diagram of the transition function. Right: A plot of the mean usage rate $M(t)$ of G_2 from a sample path.

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 stray across the separatrix, where it will be blown toward the other equilibrium.

5. Discussion and conclusion

We set out to build a mathematical model that can represent language learning in a population. The model was required to have two semi-stable states, representing populations dominated by one idealized grammar or another. To represent language change on historical time scales, the model was required to hover near one stable state on short time scales, but to spontaneously switch to the other after a reasonable amount of time. Language is represented as a mixture of the idealized grammars to reflect the variety seen in manuscripts and social data.

The Markov chain mean-field model of language learning in a population turns out not to have all of the desired properties. Although it can exhibit the required bi-stability, the stable states are too stable, and the simulated population cannot switch from one to the other in a reasonable amount of time. A further shortcoming is the fact that once a change begins, it can reverse itself.

A more complex Markov chain model that includes age structure does have all the desired properties. The population can switch spontaneously from one language to the other. 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.

This research suggests that some social structure is necessary in a model so that it may accurately represent the qualitative features of spontaneous language change. A further project would be to fit the parameters of the age-structured Markov chain to manuscript data and obtain quantitative results as well.

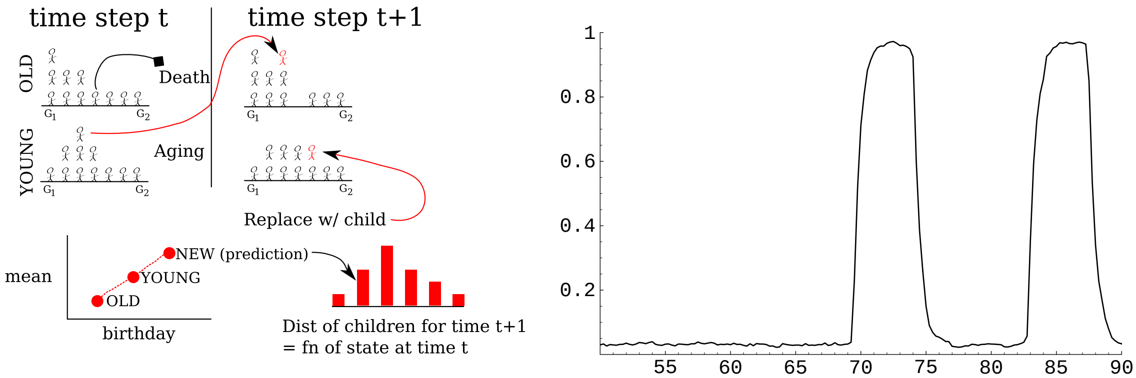

FIGURE 2. The age-structured Markov chain. Left: Diagram of the transition function. Right: A plot of the mean usage rate $M_V(t)$ of G_2 in the younger generation from a sample path.

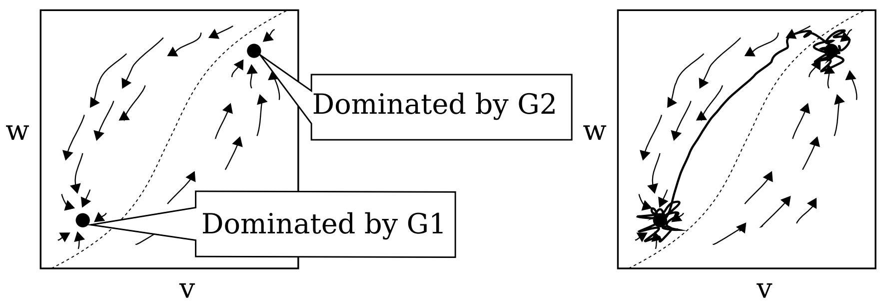

FIGURE 3. Phase portrait for (6). The two dots represent stable equilibria, and the dashed curve is the separatrix between their basins of attraction. The arrows indicate the direction of the vector field, as given by (6). The picture on the right illustrates how a sample path of the Markov chain can hover around one stable state, then eventually cross the separatrix and be blown toward the other.

References

[1] Anthony Kroch. Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1:199–244, 1989.

[2] William Labov. *Principles of Linguistic Change: Internal Factors*. Blackwell, Cambridge, MA, 1994.

[3] Charles D. Yang. *Knowledge and Learning in Natural Language*. Oxford University Press, Oxford, 2002.