Seminar 11
Modelling language change, part 2

Contents

A Intro

§1 Two weeks ago, we studied a very simple model of language change in a population of speakers with two linguistic variants. Although this model was unrealistic in many ways, it turned out to predict the S-curve and also worked to illustrate some of the most basic concepts in the mathematical and computational modelling of language change. This week, I want to keep exploring this model. Specifically, we will see how varying **population size** affects the evolution of the two linguistic variants.

§2 Before turning to population size, however, I want to make a very small modification to the model in order to make it a bit more interesting. First of all, let's recall the basic assumptions of the model:

1. two linguistic variants, A and B
2. $p(t)$ denotes the probability of coming to contact with an A speaker (and, hence, $1 - p(t)$ is the probability of encountering a B speaker)
3. two speakers come into contact at regular intervals
4. if both speakers are A or B, nothing happens
5. if one is A and the other B, then with probability α the B speaker turns into an A speaker

(Note that this model is not particularly realistic: in reality, people don't acquire variants in one-to-one linguistic interactions; in reality, people age; in reality, people may use one variant with some probability and the other with the remaining probability, rather than being categorical; and so on. But we have to set such refinements aside for now.)

So, we have the following kinds of "linguistic reactions" (compare to chemistry) in the model:

$$
\begin{aligned}
&\mathrm{A} + \mathrm{A} \to \mathrm{A} + \mathrm{A} \\
&\mathrm{B} + \mathrm{B} \to \mathrm{B} + \mathrm{B} \\
&\mathrm{A} + \mathrm{B} \to \begin{cases} \mathrm{A} + \mathrm{A} & \text{with probability } \alpha \\ \mathrm{A} + \mathrm{B} & \text{with probability } 1 - \alpha \end{cases}
\end{aligned}
\tag{1}
$$

Now, an obvious extension is to allow A speakers to become B speakers, too. Let's suppose that this occurs with probability β in an interaction between an A speaker and a B speaker. We then get the following reactions:

$$
\begin{aligned}
&\mathrm{A} + \mathrm{A} \to \mathrm{A} + \mathrm{A} \\
&\mathrm{B} + \mathrm{B} \to \mathrm{B} + \mathrm{B} \\
&\mathrm{A} + \mathrm{B} \to \begin{cases} \mathrm{A} + \mathrm{A} & \text{with probability } \alpha \\ \mathrm{B} + \mathrm{B} & \text{with probability } \beta \\ \mathrm{A} + \mathrm{B} & \text{with probability } 1 - \alpha - \beta \end{cases}
\end{aligned}
\tag{2}
$$

How does this affect the dynamics of the model? Last time, we derived the following equation for the model without β (recall that h is the probability that an interaction of *any* kind occurs between the time points t and $t + h$):

new value of p — probability that an interaction of *any* kind occurs — probability that second speaker is B

$$p(t+h) = p(t) + hp(t)[1 - p(t)]\alpha$$

old value of p — probability that first speaker is A — probability that second speaker turns from B to A

We can now update this with the β-process:

$$p(t+h) = p(t) + h\big[p(t)[1 - p(t)]\alpha - p(t)[1 - p(t)]\beta\big] \tag{3}$$

(Note that whereas the α-process *adds* something to the value of p, the β-process *subtracts* from it. This is because, in a β-interaction, one A speaker is lost.) We can rearrange this equation using standard rules of algebra:

$$p(t+h) = p(t) + hp(t)[1 - p(t)](\alpha - \beta) \tag{4}$$

And rearrange some more:

$$\frac{p(t+h) - p(t)}{h} = p(t)[1 - p(t)](\alpha - \beta) \tag{5}$$

Letting $h \to 0$, we get the derivative of p just as before:

$$\frac{d}{dt}p(t) = p(t)[1 - p(t)](\alpha - \beta) \tag{6}$$

The only difference is that, this time, we have the term $(\alpha - \beta)$ instead of just α. That is, we still get logistic (S-curve) change, but with a different rate of change parameter.

What does this mean? Well, notice that in the new model with the β-process, the rate of change is the *difference* of the two probabilities, α and β. This implies that the rate of change can be negative, too:

- If $\alpha > \beta$, then $\alpha - \beta > 0$ and the derivative of p is positive; the value of p *grows* over time.
- If $\alpha < \beta$, then $\alpha - \beta < 0$ and the derivative of p is negative; the value of p *diminishes* over time.
- If $\alpha = \beta$, then $\alpha - \beta = 0$ and the derivative of p is zero; the value of p does not change over time.

This should be intuitive: if $\alpha > \beta$, then more B speakers are being turned into A speakers than the other way around, and if $\alpha < \beta$, the reverse situation obtains. The change observed in both cases is an S-curve, since integrating Equation (6) again yields the logistic function (Figure 1).

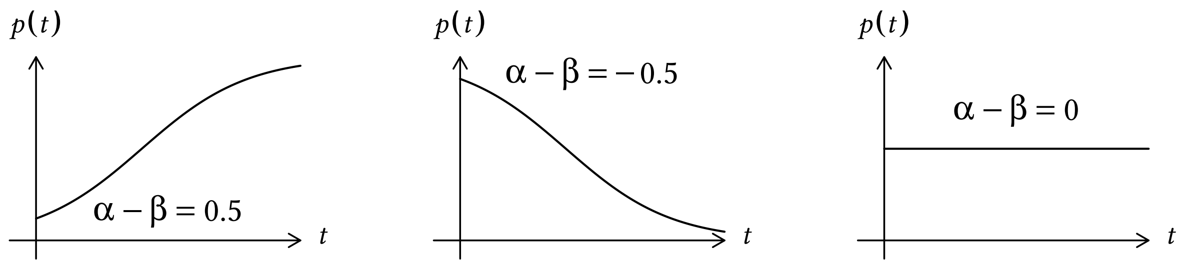

Figure 1: Evolution of p for positive, negative and zero values of $\alpha - \beta$, from different values of $p(t)$ at time $t = 0$.

B Finite vs. infinite populations

§3 I now want to draw your attention to assumption 2 under §2, reproduced here:

> $p(t)$ denotes the probability of coming to contact with an A speaker (and, hence, $1 - p(t)$ is the probability of encountering a B speaker)

The assumption here is that $p(t)$ gives this probability *irrespective of who you are (in the population)*. Technically, modellers say that a system of this kind is **well-mixing**, meaning that every speaker comes into contact with every other speaker with equal probability. Then, since we assume that $p(t)$ is the overall occurrence of variant A in the population, it follows that with probability $p(t)$ the encountered speaker is an A speaker.

This is a significant idealization, for a number of reasons. Firstly, in reality people are part of social networks, and things such as age, social class and profession affect how likely it is for a person to meet an A (or B) speaker. Secondly, even if age, class and other such effects were not in operation, it is the case that finite populations can never be truly well-mixing.

§4 To understand this, let's first consider a non-linguistic example. Tossing an unbiased coin, we say that the probability of the coin landing heads (or tails) is $1/2 = 0.5$. But what does this statement mean? Quite simply, the following: if you toss the coin 10 times, you ought to get $0.5 \times 10 = 5$ heads; if you toss the coin 100 times, you ought to get $0.5 \times 100 = 50$ heads, and so on. But now notice that, for any finite number of tosses, *it is possible* (though unlikely) that the coin always lands heads. For a series of 10 tosses, for example, this probability is

$$0.5^{10} = \underbrace{0.5 \times \cdots \times 0.5}_{10 \text{ times}} \approx 0.00098 \tag{7}$$

This is true for any sequence of coin tosses of finite length, and the probability of such extreme occurrences only tends to zero as the sequence tends to infinity (i.e. theoretically, if you toss the coin infinitely many times, then the probability that it always lands heads is zero).

Similarly, in a finite population of speakers, it is possible that in any given finite sequence of interactions, you only meet A speakers (or B speakers). This is true even if speakers come into contact with each other at random (i.e. not because of age, class and so on). Again, the probability of such extreme occurrences only tends to zero in the theoretical limit of an **infinite population**. Thus, the well-mixing assumption requires that our population is infinite. In practice, this means that a well-mixing model can only correctly capture the behaviour of large populations, as we will soon see. (Another way of putting this is to say that small populations are subject to **stochastic (or random) effects**.)

§5 The analysis of finite-population models is, however, technically much more challenging than the analysis of infinite-population models. This is because it is impossible to write down an equation like (3) for an infinite-population model (because of the aforementioned stochastic effects). While other kinds of analytical techniques exist for finite populations, they are much more complicated. It often makes more sense to simulate the model on a computer instead — this doesn't require one to solve (or even write down) any equations.

C Simulating the AB model in a finite population

§6 I have written a simple computer program that simulates the AB model (for want of a better name). This program has a population of N speakers (this is a **model parameter** that can be varied from simulation to simulation), and the program keeps track of which variant (A or B) each of the N speakers has at any given moment in time. It is then possible to calculate, for any time t, the **relative frequency** of variant A in the population, by simply counting the number of A speakers and dividing by N. I will denote this by $f(t)$. Symmetrically, the relative frequency of the variant B will be $1 - f(t)$. (Note that, theoretically, if we let the population size $N \to \infty$, the relative frequency $f(t)$ tends to the probability $p(t)$ for the corresponding infinite-population model.)

§7 Now let's see how finite-population stochastic effects can affect the dynamics of the two linguistic variants in a small population of speakers. First, note that in the infinite-population model the variant A cannot die out if $\alpha > \beta$, and, symmetrically, variant B cannot die out if $\beta > \alpha$:

Theorem 1. *Assume an infinite population. Then, if $\alpha > \beta$, variant B can never overtake variant A. Conversely, if $\beta > \alpha$, variant A can never overtake variant B.*

Proof. If $\alpha > \beta$, then $\alpha - \beta > 0$. It follows that the derivative of p is always positive:

$$\frac{d}{dt}p(t) = \underbrace{p(t)}_{>0}\underbrace{[1 - p(t)]}_{>0}\underbrace{(\alpha - \beta)}_{>0} > 0 \tag{8}$$

Thus, the value of p always increases, meaning that the abundance of A in the population always increases. Thus B cannot overtake A. An entirely symmetric argument shows the converse claim for $\beta > \alpha$. □

Thus, in an infinite population, we can demonstrate with mathematical certainty that $\alpha > \beta$ implies that variant A will always win.

§8 This, however, does not hold in a finite population! The reason is similar to the coin-tossing example from above: namely, even if $\alpha > \beta$, it is possible to have a sequence of $A + B \to B + B$ interactions such that all A speakers get wiped out from the population. Suppose, for example, that there are just two A speakers in a population of 100 speakers. Then, with probability $2/100 \times 98/100$ a B speaker meets one of the A speakers. Moreover, with probability β the A speaker turns into a B speaker. Then there is just one A speaker left. With probability $1/100 \times 99/100$ another A + B interaction occurs, and again with probability β the A speaker is turned into a B speaker. The combined probability of the whole sequence is

$$\frac{2}{100} \times \frac{98}{100} \times \beta \times \frac{1}{100} \times \frac{99}{100} \times \beta \approx 0.0002 \times \beta^2 \tag{9}$$

At this point, the population has run out of A speakers. Crucially, in a finite population, this can happen even if $\alpha > \beta$.

> **?**
>
> Why does it not happen in an infinite population, you may ask? Although Theorem 1 has already demonstrated this, it may be more insightful to consider what happens to the probability of such interaction sequences as the population size is scaled up. Notice that the fractions in Equation (9) are of the form $1/N$. If we let $N \to \infty$, then $1/N \to 0$, so the whole combined probability tends to zero.

§9 I illustrate this in Figure 2 for two different population sizes, $N = 50$ and $N = 100$, with the α and β parameters set to the values $\alpha = 0.1$ and $\beta = 0.05$ (i.e. $\alpha > \beta$). In each figure, I have run the simulation 20 times from a starting point of 5 A speakers and 45 B speakers (the $N = 50$ case) or 10 A speakers and 90 B speakers (the $N = 100$ case). In the $N = 50$ population, variant A is victorious in most simulations, but in a few simulations A gets lost because of the aforementioned stochastic effects. In the larger $N = 100$ population, the role of stochastic effects is much weaker. In fact, variant A fails to propagate in none of the 20 simulations. (With a larger number of repetitions, it would, however.)

§10 The lesson to be learned from this is: in a small population, random effects may sometimes serve to alter the diachronic outcome of linguistic variants. Infinite-population modelling predicts that whenever $\alpha > \beta$, the A variant ought to win. In a small population, however, B sometimes wins due to stochastic drift. This is the more likely the smaller the population. As N approaches infinity, the finite-population model begins to resemble the infinite-population model.

§11 A more general lesson to be learned here is that in a small population, stochastic effects add **noise** around trajectories that would be expected in a theoretically infinite population. This is illustrated in Figure 3 for three different population sizes, $N = 100$, $N = 1{,}000$ and $N = 10{,}000$. Note how, with increasing N,

1. the curve begins to resemble the S-curve (the logistic function) we earlier derived analytically for the infinite-population model, and
2. different simulations begin showing more of a uniform character (cp. with the $N = 100$ case especially).

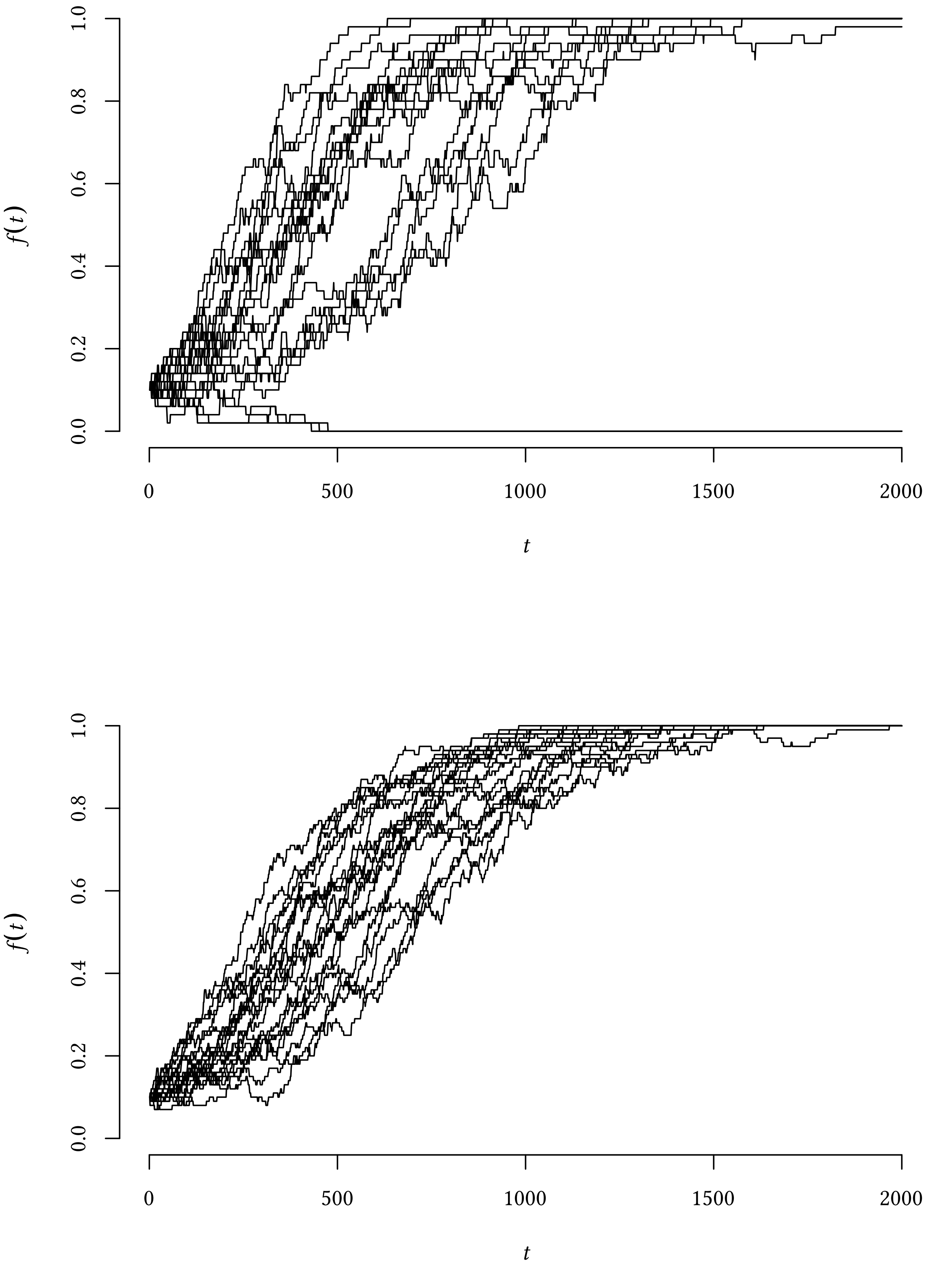

Figure 2: 20 simulations in populations of size $N = 50$ (top) and $N = 100$ (bottom).

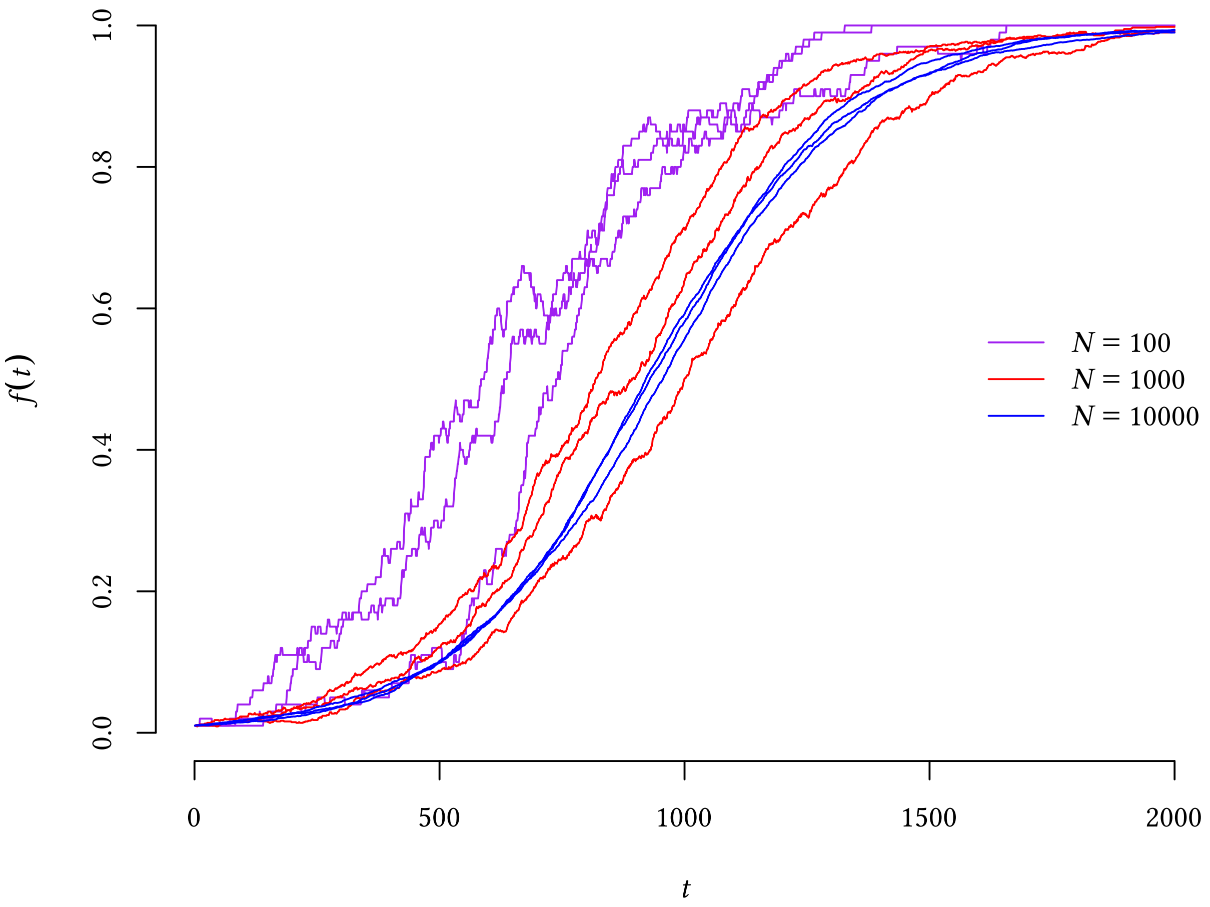

Figure 3: Evolution of a system with $\alpha > \beta$ in populations of three different sizes; 3 simulations per each population size. In each population, 1% of the speakers have variant A at time $t = 0$, and 10% of the speakers are updated at each time step of the simulation.

D Review

§12 After this seminar, you should be able to explain what the following terms mean:

finite population	well-mixing population	model parameter
infinite population	stochastic effects	noise