Seminar 9
Modelling language change

Contents

A Intro 1
B Preliminaries 2
C A simple model of (linguistic) replacement 7
D Review 12
E Further reading 12

A Intro

§1 In the previous seminar, we talked about S-curves: what they are, how their appearance might be explained, and why, sometimes, an S-curve may be only an approximation of real-life linguistic change. Today, I want to talk about some ways of *modelling* language change — S-curves in particular — mathematically.

§2 Mathematical and computational modelling is a relatively new technique in historical linguistics — the first serious applications of mathematics in this domain date from the mid-1990s. Since then, a lot of progress has been made, in a number of different modelling frameworks. Sometimes, these models are influenced by other domains in which mathematical modelling has a longer history. In evolutionary biology, for example, questions of very similar kinds are asked — how does a mutation spread through a population, for example. Any mathematical model of language change must, however, respect what we know empirically about language, in particular about the processes of language acquisition and language use.

§3 Currently existing mathematical models of language change can be divided, roughly, into two types:

- **Acquisition-based models**: in this approach, the process of language acquisition in individual speakers is modelled explicitly, and predictions of population-level linguistic diachrony (such as the S-curve) are derived from that model of acquisition.
- **Usage-based models**: in this approach, language use is modelled as a sequence of speaker interactions. Speakers are typically assumed to be parts of social networks, and special attention is paid to the way the connectivity

patterns of these networks, as well as "social prestige", influence language change. We will see an example of this kind of approach next week.

Today's content is largely agnostic with respect to this division: we will be looking at an extremely simple model of change which offers neither an explicit view of language acquisition nor an explicit account of language use. Still, the model is able to tell us something interesting about S-curves and, more importantly, to illustrate the fundamental techniques of mathematical modelling.

§4 I have spoken of both "mathematical" and "computational" modelling. These are largely synonymous, though it is possible to make the following distinction:

- A model is a mathematical model if its behaviour can be described using analytical mathematical means (basically: if the relevant mathematical equations can be solved — what this means will soon become clear).
- A model is a computational model if the above is not possible, and one has to resort to simulating the model on a computer. In practice, most models are of this kind: it turns out that only the very simplest models admit an analytical solution.

B Preliminaries

§5 To do mathematical modelling, we of course need some mathematical tools. Here, I will try to make things as simple as possible, using mostly terminology and techniques you will have learned in high-school mathematics. (Also bear in mind that, for our purposes, the important thing is to understand what the models do and why — not the technical ability to solve equations!) Sometimes things get a bit more complicated and I will simply ask you to suspend disbelief and accept that the relevant proof can be carried out.

§6 Figure 1 shows an S-curve. This curve is a **function**, more specifically a function of time: it associates to every point of time t one (and only one) value $p = p(t)$, which gives the probability of one of two competing linguistic variants.

Now consider Figure 2. This figure, too, shows a function, the **parabola** $f(x) = x^2$. In this case, we have an explicit formula for the curve, i.e. given a value of x, we know that the value of the function for this x is x^2. It is then reasonable to ask: can a formula be found for the S-curve, too? And, more importantly, what kind of process of linguistic interaction and change can give rise to that formula?

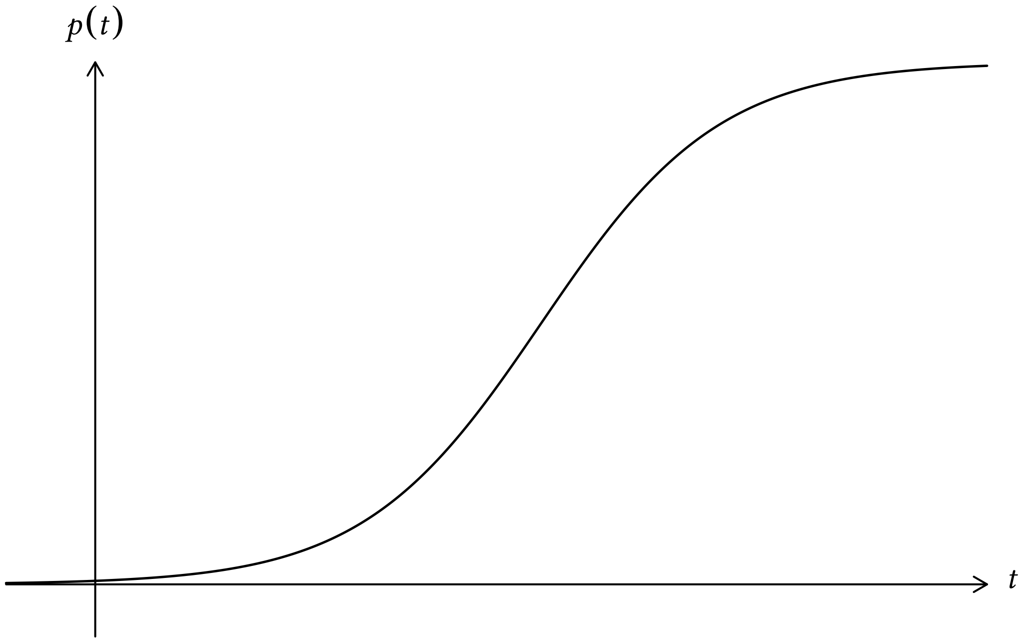


Figure 1: An S-curve.

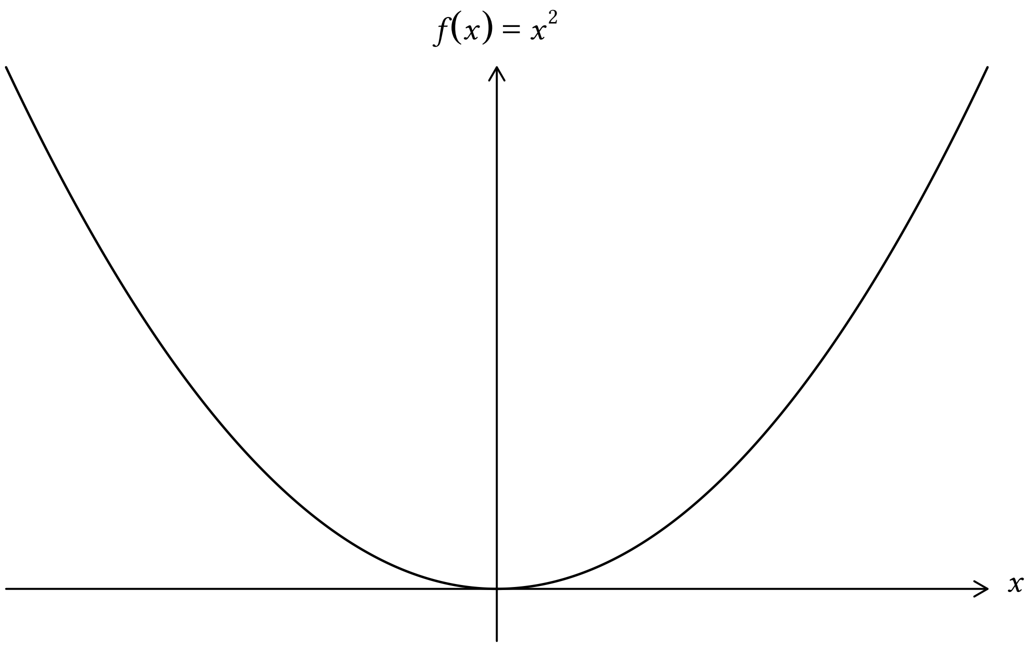
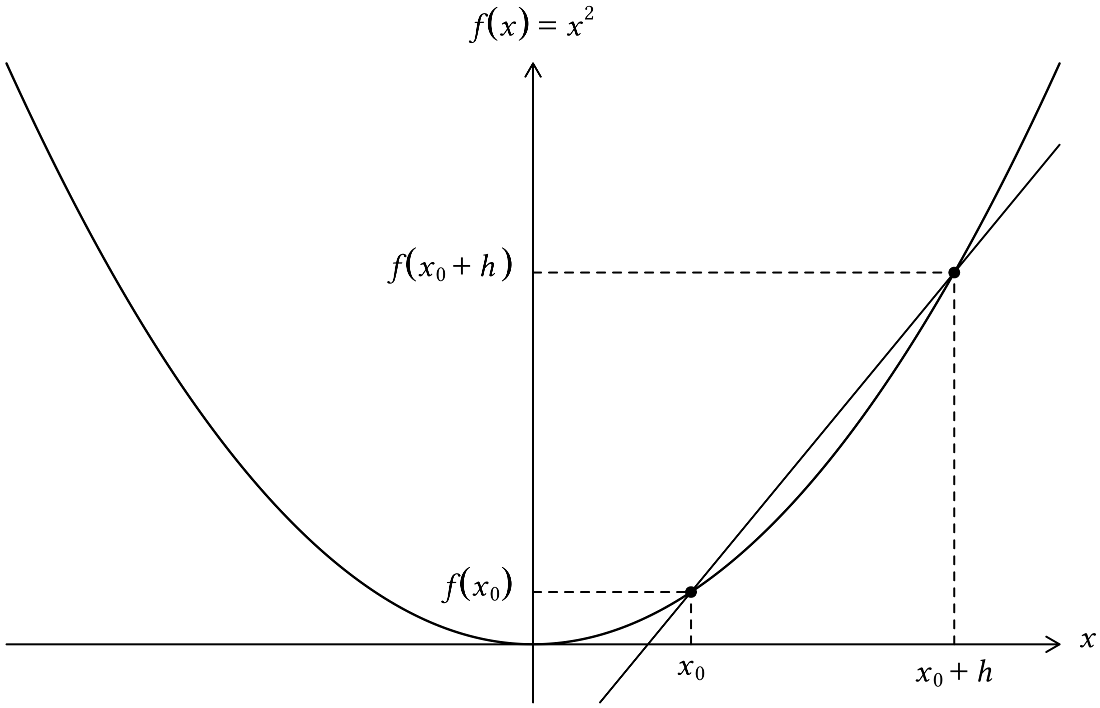


Figure 2: A parabola.

§7 First, to introduce some fundamental notions, let's think about functions in more abstract terms. The first and perhaps the most important one of these notions I want to discuss is that of a function's **derivative**, which roughly means the rate of change of the function. To formally define this notion, we first look at the difference between two points, as illustrated here for the parabola $f(x) = x^2$ and two arbitrary points x_0 and $x_0 + h$:



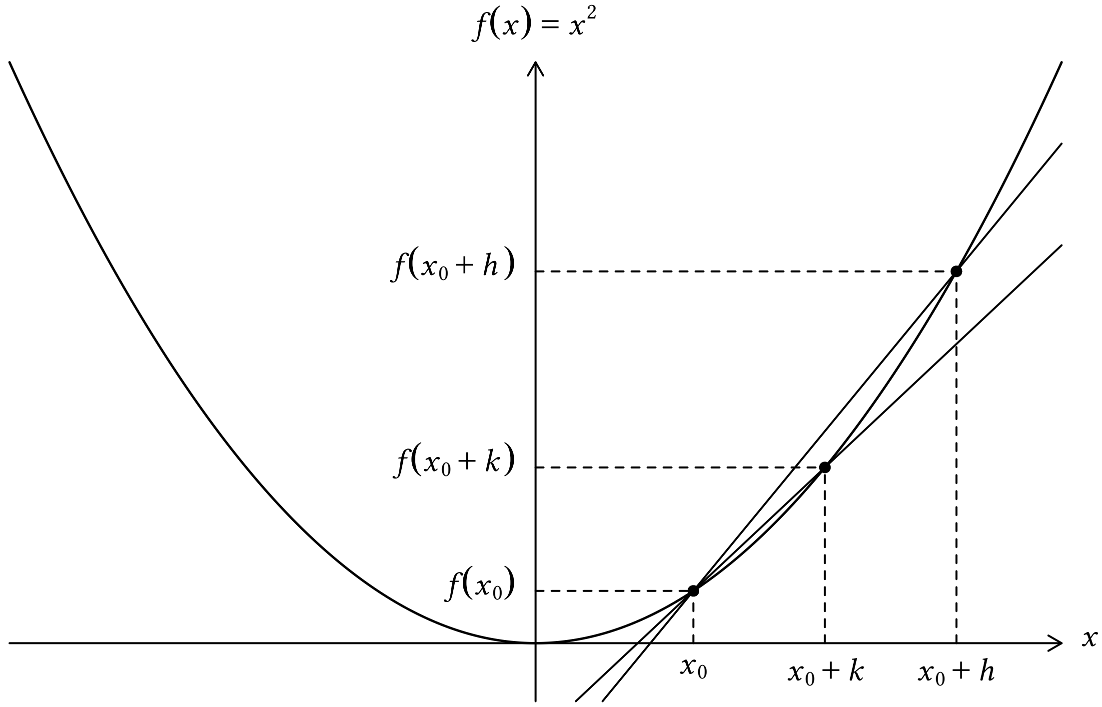
To get an estimate of how fast the value of the function (on the vertical axis) changes as one moves from x_0 to $x_0 + h$ (on the horizontal axis), we can take the difference $f(x_0 + h) - f(x_0)$ and divide it by the difference $x_0 + h - x_0$:

$$\frac{f(x_0 + h) - f(x_0)}{x_0 + h - x_0} = \frac{(x_0 + h)^2 - x_0^2}{h} \tag{1}$$

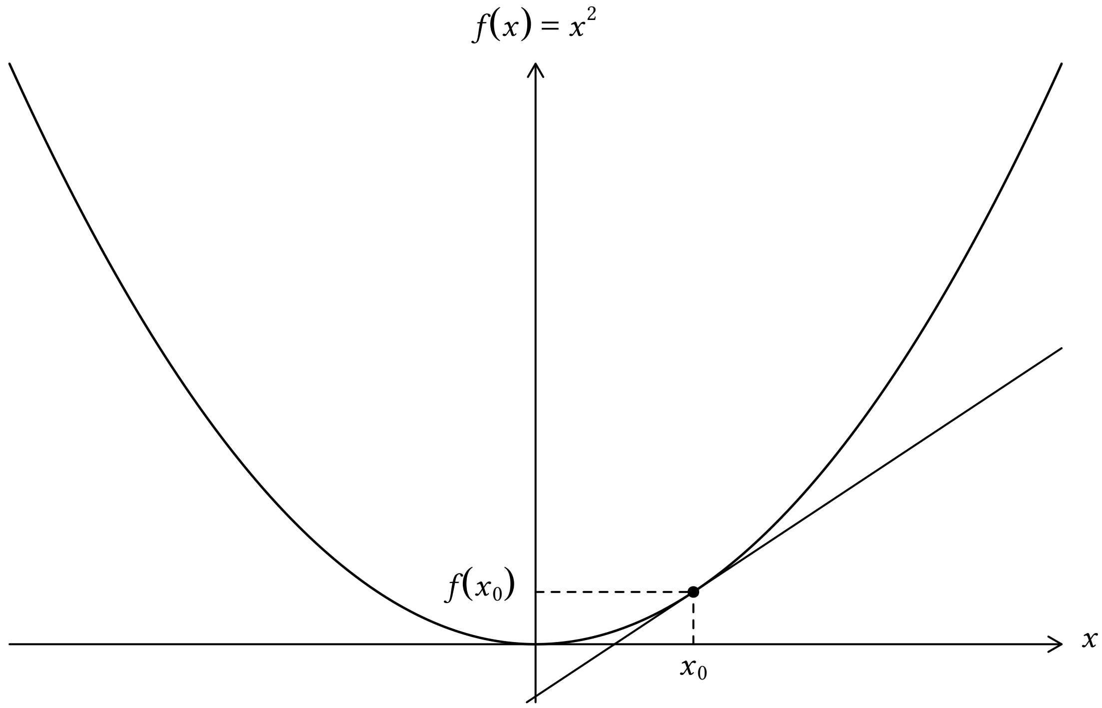
Geometrically, this ratio is the slope of the **secant** line drawn through the points $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$.

> To make things more concrete, think of $f(x_0 + h) - f(x_0)$ as the distance travelled by a car in a time period of length h. Then Equation (1) gives the average velocity of the car in that time period (velocity equals distance travelled divided by travel time).

So, Equation (1) gives the average rate of change of the function $f(x) = x^2$ in the interval ranging from our arbitrarily chosen point x_0 to our point $x_0 + h$. We can of course repeat this exercise for different selections of h. For example, take some k with $0 < k < h$, and we get the following picture and the average rate of change between x_0 and $x_0 + k$:



We can, in fact, take the increment h to be as small as we like. Mathematicians say that we let h **tend** to 0, written in symbols as $h \to 0$. Intuitively, this should give us the *instantaneous* rate of change of $f(x) = x^2$ at the point x_0:



Note, however, that if we simply substitute $h = 0$ in Equation (1), we will not get anything useful (division by zero is not allowed). To get around this, we apply some simple algebra to the equation to bring it to a manageable form:

$$\begin{aligned} \frac{f(x_0+h)-f(x_0)}{x_0+h-x_0} &= \frac{(x_0+h)^2-x_0^2}{h} \\ &= \frac{x_0^2+2x_0h+h^2-x_0^2}{h} \\ &= \frac{2x_0h+h^2}{h} \\ &= 2x_0+h. \end{aligned} \tag{2}$$

If we now let h tend to 0, the original expression tends to $2x_0$, since the h term simply vanishes from the last row of the above calculation. Mathematicians say that $2x_0$ is the **limit** of the original expression as h tends to 0.

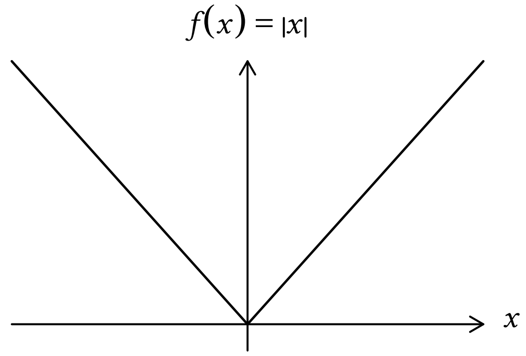
This means that we have figured out the instantaneous rate of change of the function at the point x_0, and we have found it to be equal to $2x_0$. This is known as the **derivative** of the function at the point x_0. Geometrically, it is the slope of the **tangent** drawn to the function at the point x_0 (note how we arrive at the tangent by letting $h \to 0$: the secant line basically "turns" to become the tangent at x_0).

§8 Next, notice that in our above calculations the choice of the point x_0 was arbitrary. You can replace this x_0 with any value of x and the calculations will go through. This means that $f(x) = x^2$ has a derivative at every point, and that the value of this derivative is always $2x$. Mathematicians say that the function is **everywhere differentiable** (**differentiation** is the name of the process whereby the derivative of a function is figured out). For an everywhere differentiable function f, the derivative is also a function and is denoted by $\frac{d}{dx}f(x)$. For our parabola,

$$\frac{d}{dx}f(x) = \frac{d}{dx}x^2 = 2x. \tag{3}$$

❓

The above discussion may have given the impression that every function has a derivative at every point at which the function is defined. This is not the case. For example, the function $f(x) = |x|$ (the absolute value of x) is not differentiable at the point $x = 0$. Intuitively, this is because there is no unique tangent to this point:

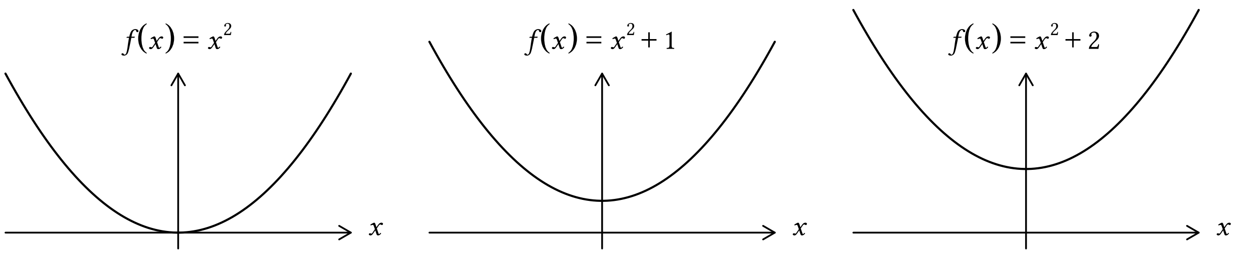


In fact, a derivative only exists at a point x if the same tangent is found when the point is approached from both left and right, a technical detail that I have glossed over in the above discussion.

There are also functions which are *nowhere* differentiable, but these generally play no role in applications such as modelling language change.

§9 Another basic notion I need (briefly) to discuss is **integration**. Integration can be viewed as the reverse of differentiation: given a derivative function, integration can be used to recover the original function. (This is guaranteed by the so-called Fundamental Theorem of Calculus.)

Short of noting its existence, I will not discuss integration in any more detail — we will only need it once, and the technical details are more complicated than those to do with differentiation. I will mention one important detail, however, which is that integration does not yield a unique outcome. To illustrate this, consider these three functions:



The derivative of each of these is intuitively the same, since the curves only differ in their "vertical" placement and not in their shape, i.e. their rates of change must be the same. Indeed, it can be checked that the derivative of each of these functions is $2x$. When this derivative is integrated, we therefore do not know which original function to recover. The solution is to say that what integration yields is an expression of the form $x^2 + C$, where C is a so-called **constant of integration**.

C A simple model of (linguistic) replacement

§10 Back to linguistics — now suppose we have a population of speakers (a language community) and that each speaker in this population entertains one of two possible linguistic variants, A or B. By **linguistic variants**, I mean different ways of "saying the same thing". For concreteness, you can think A = VO order and B = OV order, but these two variants could be anything, also different phonological feature specifications, different morphological systems, and so on. For the sake of simplicity, we here assume competition between two variants only, but the models can be generalized (with more mathematical difficulty) to competition between n (where $n > 2$) variants.

§11 Suppose, moreover, that our speakers are constantly interacting with each other. More concretely, suppose that in any given time interval, a number of events of the following kind occur:

> Two speakers come into contact. If both are A, nothing happens (both speakers exit the interaction with their linguistic knowledge unchanged). If both are B, again nothing happens. However, if one speaker is A and the other is B, then with some probability α the B speaker becomes an A speaker.

More specifically, assume that if $0 < h < 1$, then the probability of *one* such interaction occurring between the times t and $t + h$ is equal to h (i.e. the smaller the time interval, the less likely that two speakers come into contact in the first place).

Now let $p = p(t)$ represent the frequency of variant A in the population at time t. In other words, this is the probability with which you come into contact with an A speaker at time t, if you select one speaker at random. Now suppose an interaction between an A speaker and a B speaker occurs in the time interval from t to $t + h$. What is the value of p after the interaction, i.e. what is $p(t + h)$? Here is the answer:

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

Next, we rearrange this equation a bit:

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

and divide by h:

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

This should start looking similar to something we've seen before, namely the process whereby we arrived at the derivative of the function $f(x) = x^2$! Only now our function is different and the variable is indicated by the letter t. The crucial point: if we let $h \to 0$, then the left hand side of Equation (5) becomes $\frac{d}{dt}p(t)$, i.e. the derivative of p. Thus:

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

Equation (6) turns out to be an important one — so important, in fact, that it has its own name, **Verhulst's Differential Equation**. It is named after Pierre-François Verhulst, who in the 19th century described the equation as a model of population growth in ecology. (In this application, $p(t)$ is interpreted as the population size relative to an ideal maximal population size ($p = 1$) defined by the so-called carrying capacity of the environment in which the species lives — but this is a side point.)

§12 Now, importantly, Equation (6) can be integrated. This yields a definite formula for $p(t)$, the frequency of variant A as a function of time. I will skip the technical details and simply give the integral:

$$p(t) = \frac{1}{1 + e^{-\alpha(t-C)}} \tag{7}$$

Here, $e \approx 2.72$ is a mathematical constant (known as **Euler's number**) and C is a constant of integration.

Equation (7) is known as the **logistic function**; it is the **solution** of Verhulst's Differential Equation. Verhulst gave it its name, though no-one knows exactly why he chose "logistic".

Figure 3 shows the logistic function for one choice of α and three different choices of the integration constant C. Remarkably, what we have produced is an S-curve! (The constant of integration only serves to translate the curve along the time axis, i.e. it is related to the question of *when* the change takes place, but not to the form the change assumes.)

In Figure 4, on the other hand, I keep C constant but vary the value of α. It turns out that this parameter determines the overall rate of change, i.e. the overall time it takes for the change to go from beginning to completion.

§13 To summarize: formulating a very simple model of competition between two linguistic variants in a population of speakers allowed us to write down the derivative of p (the probability of one of the variants) as Equation (6). Integrating this equation yields the logistic function (7), which expresses p as a function of time t. This function turns out to be an S-curve. Thus, the simple model is one (but not the only one!) possible mechanism for generating S-curves of linguistic change.

§14 So *why* did we get an S-curve — a "slow–fast–slow" curve with symmetry about its midpoint? To understand this, it pays to examine Equation (6) in a bit

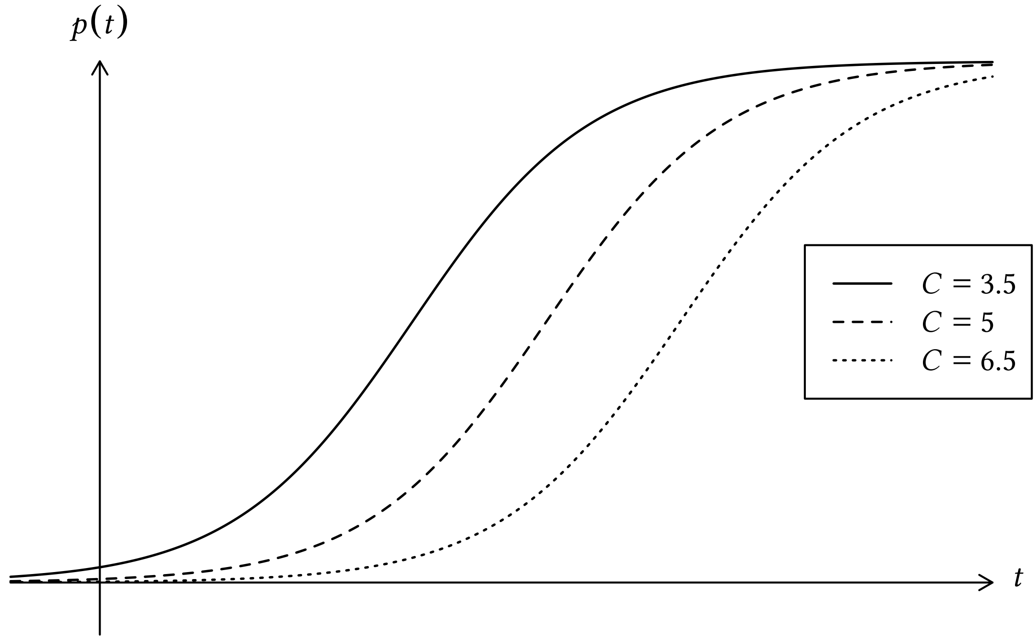


Figure 3: The logistic function for three different choices of the constant of integration *C*.

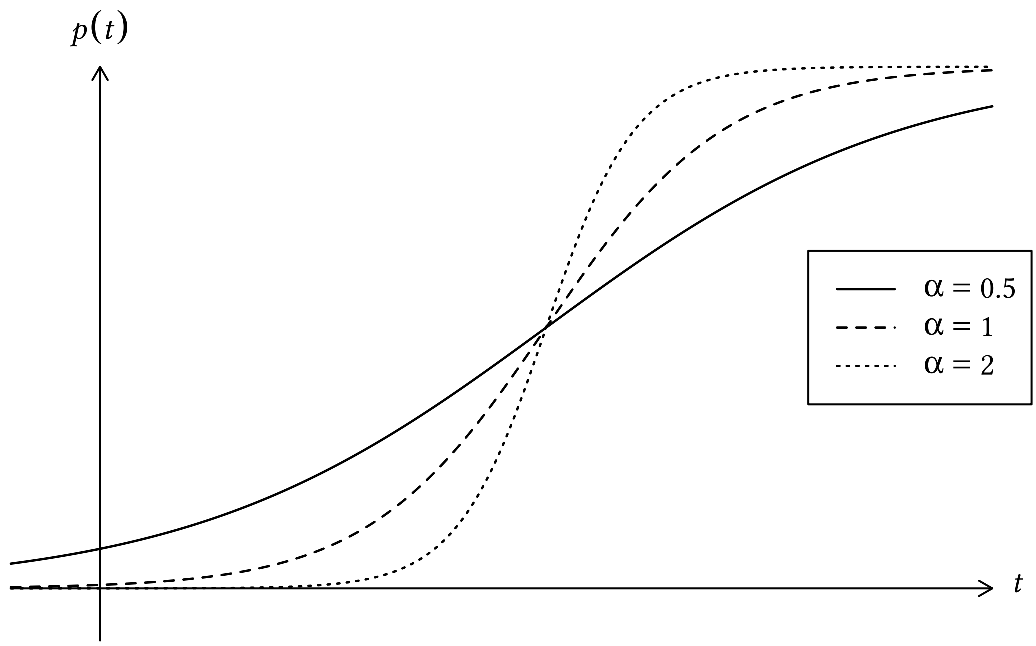


Figure 4: The logistic function for three different choices of α, the probability of a B speaker turning into an A speaker.

Figure 5: The logistic function and its derivative.

more detail. I reproduce the equation here:

$$\frac{d}{dt}p(t) = p(t)[1 - p(t)]\alpha. \tag{8}$$

The key to both the slow–fast–slow pattern and the symmetry lies in the form of this equation. Note that initially, when $p(t)$ is small, the product

$$p(t)[1 - p(t)]\alpha \tag{9}$$

is also small; hence, the value of the derivative is small. (This is the first slow phase of the curve.) But, equally, when $p(t)$ is large, the value of $1 - p(t)$ is small, so the product

$$p(t)[1 - p(t)]\alpha \tag{10}$$

is again small. (This is the second slow phase.) The fast phase occurs for intermediate values of $p(t)$. In fact, the derivative (8) reaches its maximum value when $p(t) = 0.5$, exactly mid-way through the change. In Figure 5 I illustrate this by plotting the logistic function and its derivative simultaneously.

D Review

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

acquisition-based model
usage-based model
function
derivative
limit
differentiation
integration
linguistic variant
Verhulst's Differential Equation
logistic function

E Further reading

§16 No textbook exists (yet) on the mathematical modelling of language change. Seminal research articles include Niyogi and Berwick (1997), Nettle (1999), Yang (2000), Baxter et al. (2006) and Blythe and Croft (2012).

References

Baxter, G. J., R. A. Blythe, W. Croft, and A. J. McKane (2006). Utterance selection model of language change. *Physical Review E 73*, 046118.

Blythe, R. A. and W. Croft (2012). S-curves and the mechanisms of propagation in language change. *Language 88*(2), 269–304.

Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua 108*, 95–117.

Niyogi, P. and R. C. Berwick (1997). A dynamical systems model for language change. *Complex Systems 11*, 161–204.

Yang, C. D. (2000). Internal and external forces in language change. *Language Variation and Change 12*, 231–250.