

Mathematical Appendix to “Modeling language change: An evaluation of Trudgill’s theory of the emergence of New Zealand English”

GARETH J. BAXTER
Universidade de Aveiro

RICHARD A. BLYTHE
University of Edinburgh

WILLIAM CROFT
University of New Mexico

ALAN J. MCKANE
University of Manchester

In this appendix, we give some additional details of how the major mathematical results presented in the main text are obtained. To prevent overburdening this section with equations, we shall make frequent reference to our earlier work (Baxter et al., 2006) that contains further mathematical details. In footnotes, we place comments of a technical nature that can be skipped by less mathematically inclined readers.

Probability of fixation

As discussed in the main text, the model defined is stochastic. That is, each realization of the sequence of steps described results in a different outcome because the choice of speakers that interact and the tokens they produce is random. However, one can ask how *average* quantities change with time, and these vary in a predictable way. In the present context, we mean an average taken over the distribution generated by many independent realizations of the same process.

Using the results of Baxter et al. (2006), one can show¹ that the average of the lingueme variant frequency x_i stored in speaker i 's grammar, a quantity denoted \bar{x}_i , changes with time according to the equation

$$\frac{d}{dt} \bar{x}_i = h \sum_{j=1}^N G_{ij} (\bar{x}_j - \bar{x}_i) \quad (\text{A1})$$

within a model of neutral interactor selection, where the speaker interaction weights H_{ij} are all equal to the common value λh . In words, this equation says that, at any given instant, the rate at which the average quantity \bar{x}_i changes with time depends on the values of other averages \bar{x}_j through the sum on the right-hand side. In this formulation, one unit of time corresponds to $1/\lambda^2$ interactions between pairs of speakers.

To determine the probability that the variant whose frequency is given by x_i eventually

fixes, that is, becomes the sole remaining variant used in the entire community, it is useful to focus on its total usage frequency across the entire community. This quantity we simply call x (with no subscript) and define as follows:

$$x = \frac{1}{N} \sum_{i=1}^N x_i.$$

This frequency can equal 0 only if no speakers are using this variant (i.e., it has gone extinct), and similarly can equal 1 only if all speakers use only this variant (i.e., it has fixed). The key point here is that, using Equation (A1), one can show that the rate of change of the frequency \bar{x} , that is, the average value of the community frequency x , is 0.² That is, the value \bar{x} does not change with time.

The significance of this result is as follows. Consider a very large number of independent realizations of the process, all starting from the same initial condition, that is, the same set of grammar frequencies x_i . The initial value of the quantity x is then the combined usage frequency across the entire community and has the same (prescribed) value in all realizations. Therefore its average \bar{x} across all realizations is equal to this prescribed value. Now look sufficiently far ahead that fixation or extinction is guaranteed to have occurred. If in each independent realization, fixation occurs with probability p , then in a fraction p of realizations, the value of $x = 1$, and in the remaining fraction $x = 0$. Thus, at very late times, $\bar{x} = p \times 1 + (1 - p) \times 0 = p$. Therefore, the fixation probability p is equal to \bar{x} , which in turn is equal to the variant's initial usage frequency.

Mean time to fixation

In the sections “Propagation by neutral interactor selection” and “Modeling generational replacement” (see print version), we were mostly concerned with the number of interactions between speakers³ that are needed on average until a state of fixation is reached. With only a single generation of speakers, as described in the section “Propagation by neutral interactor selection,” computer simulations suggest (see also Baxter et al., 2006) that fixation is a two-stage process:

1. The average quantity \bar{x}_i , which may initially be different for each speaker, approaches the community average \bar{x} .
2. The x_i values fluctuate until such a time that all speakers all spontaneously fix on—or stop using—the variant.

This behavior can be seen from Figure A1. The average \bar{x}_i for two speakers, one initially always using a variant and one initially never using it, is plotted, along with the variance in a speaker average and the covariance between two of them. In both plots, all speakers interact with each other equally often, and in the upper plot $N = 20$, but in the lower plot $N = 40$. In both cases, half the speakers initially used one variant, and the other half used the other. Superposed on these plots is the probability distribution for the fixation time, and the solid vertical line is the mean fixation time.

The main point to notice is that the initial relaxation of \bar{x}_i to the community mean of 0.5

happens very quickly compared with the mean fixation time (or the peak in the fixation time distribution). Furthermore, the duration of the initial relaxation relative to the mean fixation time (or the peak of the fixation time distribution) is shorter in the system with the larger number of speakers. In fact, it turns out that, under neutral interactor selection, the first stage (relaxation) lasts for a time that—for the vast majority of networks—is proportional to the number of speakers N . Meanwhile, on all networks, the second stage lasts for a time proportional to N^2 . Therefore, if the number of speakers is large, the second stage typically lasts very much longer than the first, and so one can—to a good approximation—ignore the expected length of the first stage and focus on that of the second. There are some networks where the first stage lasts much longer than the second does—for example, if the community comprises a very long “chain” of speakers, each talking only to its neighbors. Such networks lack the “small-world” property, whereby the shortest path between any pair of speakers is much less than the total number N , which is believed to be present in real social networks (Watts, 1999). We will not discuss these unusual cases further here.

The lifetime of the second stage is estimated by making two assumptions on the variation in grammar frequencies x_i across the community at the start of the second stage. First it is assumed that the difference between x_i and the community average \bar{x} is uncorrelated with the difference for some other speaker j . In other words, the covariance of x_i and x_j is assumed to vanish. Support for this assumption is given in Figure A1, which shows this covariance to be small at the onset of the second stage. Second, it is assumed that at this time, the rate of change of the variance is so slow that it can be neglected—this can also be seen from Figure A1, where the rate of growth of the variance slows dramatically at the start of the second stage.

Following a procedure similar to that used to obtain Equation (A1) from the results of Baxter et al. (2006), one can show that under neutral interactor selection, a time-independent variance in x_i is obtained if the condition

$$G_i \left[\bar{x}_i (1 - \bar{x}_i) - (2h + \frac{1}{T}) \text{Var}(x_i) \right] = 0$$

is satisfied. In this equation, $\text{Var}(x_i)$ denotes the variance of x_i and G_i is the total fraction of all interactions in which speaker i participates. The main point to notice about this equation is that it can be satisfied only if $G_i = 0$ (which is never true as all speakers are involved in some fraction of interactions) or the quantity in square brackets is 0. Because this latter quantity does not involve the interaction frequencies G_{ij} , it follows that the variance of x_i is independent of the network structure. If we further assume that the lifetime of the second stage depends only on the mean and variance of the grammar frequencies at the start of the second stage, it follows that this lifetime is also independent of the network structure under neutral interactor selection.

Specifically, this means that the fixation time given for the case of a network in which all speakers interact equally often with each other (Baxter et al., 2006) applies for any network structure where the duration of Stage I (described in the print version) is much shorter than that of Stage II. Expressed in terms of the total number of interactions that need to have taken place in the whole community, this fixation time is⁴

$$I_{fix} = \frac{N^2 T}{\lambda^2} \left(1 + \frac{1}{2Th} \right) \frac{(x-1)\ln(1-x)}{x}. \quad (\text{A2})$$

In Figure A2, we demonstrate the fact that the fixation time is independent of the network structure by plotting its value (divided by the number of speakers) obtained by computer simulation for a range of structures (see the description of the model in the print version). These are a flat society, where all speakers interact equally with all other speakers; a hub-and-spoke society divided into groups of 10 speakers, one of which is a central hub whose speakers interact with those in the remaining spoke groups, between which there are no interactions; an equal group network, where pairs of speakers within the groups of 10 interact 10 times more frequently than pairs drawn from 2 different groups; a ring of speakers with 10% of the possible remaining long-range connections between them established; and a network in which only 20% of all possible interactions occur. We see that for any given system size, the fixation times across all structures are in agreement with each other and the theoretical prediction given previously (at least within the errors that arise from statistical fluctuations inherent in a finite sample of simulated realizations of the process).

We finally remark in this section that the fixation time predicted by Equation (A2) can actually be shown by a more rigorous argument to underestimate its true value under neutral interactor selection. The details of this calculation were presented elsewhere (Baxter et al., 2008).

Memory lifetimes implied by the model

For a given set of parameters, the model predicts (under neutral interactor selection) that the total number of interactions between all speakers that take place before fixation occurs is

$$I_{fix} = \frac{N^2 T}{\lambda^2} f(hT)\omega(x)$$

where $f(hT) = 1 + \frac{1}{2hT}$ and $\omega(x) = (x-1)\ln(1-x)/x$ (see Equation (A2)). We will assume that this number of interactions takes place over a single human's lifetime and ask what value the parameter λ would need to be to achieve this. In turn, this needs to be converted to a real time so that it may be compared against empirically determined memory time windows.

The strategy is the following. We focus on speaker i who participates in a fraction G_i of all interactions. Therefore, by the time fixation has occurred (on average), she has uttered $G_i I_{fix} T$ tokens. If fixation occurs within a single speaker's lifetime, this number must be equal to T^* . Substituting in the full expression for I_{fix} , we find that

$$\frac{G_i N^2 T^2}{\lambda^2} f(hT)\omega(x) = T^*.$$

Rearrangement of this expression tells us the value of λ that is needed to achieve fixation in speaker i 's lifetime:

$$\lambda = \sqrt{\frac{G_i f(hT) \omega(x)}{T^*}} NT. \quad (\text{A3})$$

The parameter λ controls what fraction of a speaker's grammar contains information about the most recent interaction. The intensity of this information thus decays by a factor of approximately $1 - \lambda$ in each interaction. Therefore, the larger the value of λ that is imposed, the more quickly a speaker forgets information about earlier utterances. We see that λ is large when N is large and/or T^* is small. As fixation times grow with N , it follows that to allow fixation to occur in a (fixed) human lifetime as N is increased, correspondingly faster turnover of the grammar is required to achieve this. On the other hand, if T^* is small, the time between successive tokens heard by a speaker is long, and λ has to be increased so that the tokens heard near the start of a speaker's life are forgotten before the end.

To convert λ to a fraction of a human lifetime, we need to introduce a threshold intensity of the memory of a token at which it is considered forgotten. This we set at a fraction ε of its initial intensity. Because this intensity decreases by a factor $1 - \lambda$ at each interaction, the number of interactions n until the threshold ε is reached is found from

$$(1 - \lambda)^n = \varepsilon.$$

To rearrange this expression to give n as a function of λ , we take the logarithm of both sides

$$n \ln(1 - \lambda) = \ln \varepsilon.$$

We use the fact that, for small λ , $\ln(1 - \lambda) \approx -\lambda$ to obtain the simpler, approximate expression

$$n = -\frac{\ln \varepsilon}{\lambda}.$$

Note that since $\varepsilon < 1$, its logarithm is negative and hence n is a positive number as required.

It now remains to write this number of interactions as a fraction of speaker i 's lifetime. This is achieved by dividing it by the number of interactions in which a speaker participates in her lifetime, which is simply T^*/T . Thus the memory time, as a fraction of a speaker's lifetime, is

$$t_{mem} = -\frac{\ln \varepsilon}{\lambda} \frac{T}{T^*}.$$

Substituting in the value of λ from Equation (A3) we obtain

$$t_{mem} = -\frac{\ln \varepsilon}{\sqrt{G_i N^2 T^* f(hT) \omega(x)}}.$$

An inequality that holds for any network is obtained by noticing that the smallest t_{mem} applies to the most frequently interacting speaker: that with the largest G_i . In turn, the largest G_i must be at least $2/N$, because all the G_i are required to sum to 2. Hence, by putting $G_i = 2/N$ in the previous expression, we find an upper bound on the shortest memory time imposed on any speaker in the model. This bound is presented as Equation (2) in the “Propagation by neutral interactor selection” section of the print version of this paper.

NOTES

1. Specifically, the Fokker-Planck equation is multiplied by x_i and all x coordinates are integrated over to obtain the average.
2. This is achieved by summing Equation (A1) over all speakers i .
3. Here, and in the section “Propagation by neutral interactor selection” (see print version), we mean by interactions the *total* number of interactions that take place in the whole community; to convert this to the number of interactions experienced on average by a single speaker, one should divide by N .
4. This differs from the expression in Baxter et al. (2006) by a factor $1/\lambda^2$, which is due to a conversion between continuous time units and interactions; furthermore, we have only kept terms proportional to N^2 here, consistent with our having neglected the lifetime of Stage I.