Macro and Micro Models of Linguistic Evolution

Ted Briscoe
Computer Laboratory, University of Cambridge
ejb@cl.cam.ac.uk

Introduction

Mathematical modelling and computational simulation of dynamical systems are useful tools for refining and checking hypotheses about linguistic evolution. I compare macro-evolutionary models, which typically assume non-overlapping generations and infinite populations to achieve analytic tractability, to stochastic micro-evolutionary models, which typically trade analytic tractability for more realistic demographic assumptions (e.g. Renshaw, 1991). The predictions of a macro and micro model for a simple linguistic example are dramatically different.

The NB Model

Niyogi and Berwick, 1997; Niyogi, 2000; hereafter NB) have developed a model of linguistic evolution based on a macro-evolutionary model in which languages are treated as dynamical systems, the aggregate output of a population of grammars, and evolution of the system corresponds to changes in the distribution of (variant) grammars. This distribution changes as each new generation of language learners each acquire a grammar from the data provided by their speech community (i.e. the previous generation of learners).

The NB model has three main components: a class of grammars, G, from which a learner selects on the basis of data; a learning algorithm, A, used by the learner to choose a grammar, $g \in G$; and a probability distribution, P, with which sentences are presented to the learner. P is defined in terms of the distribution on sentences for each $g \in G$ and the proportions of each g in the population. A dynamical system can now be defined in which each state of the system is represented by a P for state, S_i, and P' for state S_{i+1} can be calculated by an update rule which depends only on P, A and G. Crucially, this deterministic update rule relies on the assumption of non-overlapping generations of learners and speakers, and the abstraction to infinite populations. The former assumption makes the analytic calculation of P for each state of the system tractable and the latter abstraction amounts to the assumption that random sampling effects are irrelevant in the calculation of the proportions of learners who converge to specific grammars given P.

A critical result which follows from one instantiation of the NB model is that the evolution of linguistic systems will be S-shaped or logistic. NB argue that it is a strength of their model that logistic behaviour can be derived analytically from the properties of the update rule, given certain assumptions about G, A and P. Diachronic linguistic work has shown that language change often follows a broadly S-shaped pattern but has not been able to derive this behaviour from more fundamental assumptions (e.g. Lightfoot, 1999:101f). To derive the logistic map, NB assume a two grammar / language system in which A selects between g^1 and g^2 on the basis of 2 example sentences from P. If the last sentence is unambiguously from one grammar, then this grammar is selected. If the first sentence is unambiguously from one grammar and the last is ambiguous, then the learner selects the grammar on the basis of the first example. Otherwise, a random unbiased selection is made. The update rule is defined in terms of the consequent probabilities of A selecting g^1 or g^2 given P. If these probabilities are not equal then the population will converge logistically to the better represented grammar over time. If they are equal then the system is stable and does not evolve.

The critical assumption for the analytic derivation of logistic behaviour lies not in the specific assumptions about G, A or P, but rather in D, the model of a dynamical system that NB adopt. (This is not to say that G, P and particularly A are not important – Robert Clark (1996) demonstrates via simulation that logistic change is the exception rather than rule in the NB model, and NB only derive this behaviour analytically for the specific case of selecting between g^1 and g^2 from 2 examples.) NB characterise the states of the system in terms of the proportion of average or arbitrary learners exposed to P who converge to g^1 (equivalenty g^2). This is a macro-evolutionary model in which what is modelled is the gross statistical behaviour of learners and thus of the linguistic systems, rather than the behaviour of individual learners within the population.

A Simple Stochastic Model

If we replace D with a stochastic micro-evolutionary model, D' in which there is a finite population of non-overlapping generations, and we model the behaviour of each individual learner while keeping assumptions about P, A and G identical, we find very different behaviour – at least until population sizes become very large. The differences are most obvious when we consider the case where each learner has an equal chance of being exposed to an unambiguous sentence from g^1 or g^2. In the NB model this leads to stasis, but in a micro model stasis is extremely improbable.

For simplicity assume a starting point in which there are equal numbers of g^1 and g^2 speakers in the population, $\frac{1}{2}$ of sentences from g^1 and g^2 can distinguish the two grammars (i.e. are unambiguous with respect to the source grammar which generated them), and P is a uniform distribution. The probability that a learner selecting a grammar based on 2 sentences will select on the basis of an initial random unbiased setting, because the two sentences are ambiguous, is $\frac{1}{4}$, because for each independently drawn observation from P the chance of seeing an ambiguous sentence is $\frac{1}{2}$. Therefore, the learner will select g^1 on the basis of data with probability $\frac{3}{8}$ ($P = 0.375$). (NB give equations for calculating such probabilities for A.)

For stasis we require exactly half of the learners to acquire g^1. Suppose there are 100 learners; what is the probability that exactly half will select g^1 in the first generation? The data provided to each learner is stochastically independent so this is equivalent to asking how probable is it that in 100 tosses of an unbiased coin exactly 50 will come up heads, and is given by the binomial theorem: $P = 0.0795$ (e.g. McColl, 1995). Therefore, it is very improbable that the distribution P will remain unaltered, and unbiased between g^1 and g^2, for the next generation of learners. This result is in marked contrast from that of NB and follows directly from modelling the fact that each individual learner will be exposed to a different (random) sample of sentences.

To see how likely it is that, given a biased distribution, P, on g^1 and g^2, the dominant grammar will spread logistically through the population given D', we need to consider the shape of the skewed binomial distribution arising from the bias. For example, if we minimally modify the example above by assuming that $\frac{3}{4}$ of the adult population speak g^1, the probability that a learner will acquire g^1 given 2 sentences is now $\frac{11}{16}$ ($P = 0.687$). (Note that it is not $\frac{12}{16}$ ($P = 0.75$) because of the possibility of selection according to the initial unbiased setting when the data seen is ambiguous.) Consequently, the probability that more than 75 learners will acquire g^1 is only $P = 0.070$, though the probability that more than 50 will acquire g^1 is $P > 0.999$. In fact, the distribution peaks at 69 learners predicting not logistic growth but rather a probable slight decline in the number of g^1 speakers in the next generation. In the limit, if the whole population speak g^1, the probability that a learner will select g^1 is $\frac{7}{8}$ ($P = 0.875$) because there remains a $\frac{1}{8}$ chance that a learner will see 2 ambiguous sentences and select g^2 on the basis of a random initial setting. Therefore, even in this case the resulting binomial distribution peaks with 88 learners acquiring g^1 in the next generation. Therefore, given these assumptions for P, the micro model predicts endless random drift in the proportions of g^1 and g^2 speakers.

It might be objected that this result follows primarily from choosing G and P with a high proportion of ambiguous sentences, so that learners frequently select grammars on the basis of random (unbiased) initial settings (though G and P here are in this respect similar to several of the more realistic examples NB consider, derived from Gibson and Wexler, 1994). If we assume, that g^1 and g^2 are as highly differentiated as possible and share no sentences, then a learner will select between them with probability directly correlated with the proportions of g^1 and g^2 speakers in the adult population. In the case of equal proportions, the probability that exactly half the population of learners will acquire g^1 (equivalently g^2) is still given by the unbiased binomial distribution in D', and thus remains low. The binomial distributions for each generation of learners will now peak at exactly the point predicted by the proportion of adult speakers, but this still only allows us to predict that +/–13 learners around this peak will acquire g^1 with $P > 0.99$ for a population of 100 learners. Therefore, we can still expect to see an oscillating pattern of random drift prior to eventual fixation on one variant.

Realistic Stochastic Models

For some types of language change the idealisation of D to infinite populations may not be harmful; for example, diffusion through American English within the last 50 years might be such a case. However, even then we would need to be clear that there is an analytic and thus predictive advantage to macro modelling for *realistic* versions of G, A and P, and this has not been demonstrated as yet. In all cases where evolution of a linguistic system is likely to have taken place in small relatively isolated speech communities – for example, modelling of prehistoric development or of a process like creolisation, where the relevant populations are likely to have been at most in the low hundreds – abstracting away from sampling issues is dangerous.

Furthermore, the specific behaviour which we want to derive, such as logistic change in the system, may simply follow directly from more realistic demographic assumptions than are possible with macro models. For example, population movement, birthrate, the proportion of language learners in the population and the resultant linguistic mix of the population are critical factors in understanding creolisation. Briscoe (2000) discusses several models enriching D' with more realistic demography and realistic accounts of A and G.

Briscoe, Ted (2000) 'Evolutionary perspectives on diachronic syntax' in Susan Pintzuk, George Tsoulas and Anthony Warner (ed.), *Diachronic Syntax: Models and Mechanisms,* Oxford: Oxford University Press.

Clark, R.A.J. (1996) *Internal and External Factors Affecting language Change: A Computational Model,* MSc Dissertation, University of Edinburgh.

Gibson, E. and Wexler, K. (1994) 'Triggers', *Linguistic Inquiry, vol.25.3,* 407–454.

Lightfoot, D. (1999) *The Development of Language: Acquisition, Change, and Evolution,* Blackwell, Oxford.

McColl, J.H. (1995) *Probability,* Edward Arnold, London.

Niyogi, P. (2000, in press) 'Theories of Cultural Change and their Application to Language Evolution' in Briscoe, E.J. (ed.), *Language Acquisition and Linguistic Evolution: Formal and Computational Approaches,* Cambridge University Press, Cambridge.

Niyogi, P. and Berwick, R. (697–719) 'Evolutionary consequences of language learning', *Linguistics and Philosophy, vol.20.* 1997

Renshaw, E. (1991) *Modelling Biological Populations in Space and Time,* Cambridge University Press, Cambridge.