PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2010) **277**, 429–436
doi:10.1098/rspb.2009.1513
Published online 7 October 2009

Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift

Florencia Reali* and Thomas L. Griffiths

Department of Psychology, 3210 Tolman Hall, MC 1650, University of California at Berkeley, Berkeley, CA 94720-1650, USA

Scientists studying how languages change over time often make an analogy between biological and cultural evolution, with words or grammars behaving like traits subject to natural selection. Recent work has exploited this analogy by using models of biological evolution to explain the properties of languages and other cultural artefacts. However, the mechanisms of biological and cultural evolution are very different: biological traits are passed between generations by genes, while languages and concepts are transmitted through learning. Here we show that these different mechanisms can have the same results, demonstrating that the transmission of frequency distributions over variants of linguistic forms by Bayesian learners is equivalent to the Wright–Fisher model of genetic drift. This simple learning mechanism thus provides a justification for the use of models of genetic drift in studying language evolution. In addition to providing an explicit connection between biological and cultural evolution, this allows us to define a 'neutral' model that indicates how languages can change in the absence of selection at the level of linguistic variants. We demonstrate that this neutral model can account for three phenomena: the s-shaped curve of language change, the distribution of word frequencies, and the relationship between word frequencies and extinction rates.

Keywords: language evolution; genetic drift; Bayesian inference; neutral models

1. INTRODUCTION

Natural languages, like species, evolve over time. The mechanisms of language evolution are quite different from those underlying biological evolution, with learning being the primary mechanism by which languages are transmitted between people. However, accounts of language evolution often appeal to forces that have analogues in biological evolution, such as selection or directed mutation. Recent computational work has emphasized the role of selective forces by focusing on the consequences of a language for the 'fitness' of its speakers in terms of communication success (Cavalli-Sforza & Feldman 1983; Hurford 1989; Oliphant 1994; Komarova & Nowak 2001). Other studies have emphasized the effects of differential learnability of competing linguistic variants, with selection or directed mutation operating at the level of sounds, words or grammatical structures (Batali 1998; Pearl & Weinberg 2007; Christiansen & Chater 2008). These functional explanations provide an intuitive and appealing account of language evolution. However, it is possible that the changes we see in languages over time could be explained without appealing to such factors, resulting from processes analogous to genetic drift.

Evaluating the role of selective forces in language evolution requires developing *neutral models* for language evolution, characterizing how languages can be expected to change simply as a consequence of being passed from one learner to another in the absence of selection or directed mutation. Neutral models have come to play a significant role in the modern theory of biological evolution, where they account for variation seen at the molecular level and provide a tool for testing for the presence of selection (Kimura 1983). The work mentioned in the previous paragraph illustrates that there are at least two levels at which evolutionary forces can operate in language evolution: at the level of entire languages (through the fitness of speakers or directed mutation when languages are passed from one speaker to another), and at the level of individual linguistic variants (with particular sounds, words or grammatical structures being favoured over others by learners). In this paper, we define a model that is neutral at the level of linguistic variants, indicating how languages can change in the absence of selection for particular variants.

Defining a model of language evolution that is neutral at the level of linguistic variants requires an account of learning that is explicit about the inductive biases of learners—those factors that make some variants easier to learn than others—so that it is clear that these biases do not favour particular variants. We model learning as statistical inference, with learners using Bayes' rule to combine the clues provided by a set of utterances with inductive biases expressed through a prior distribution over languages. We define a neutral model by using a prior that assigns equal probability to different variants of a linguistic form. While it is neutral at the level of variants, this approach allows for the possibility that learners have more general expectations about the structure of a language—such as the amount of probabilistic variation in the language, and the tendency for new

* Author for correspondence (florencia.reali@gmail.com).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2009.1513 or via http://rspb.royalsocietypublishing.org.

Received 21 August 2009
Accepted 21 September 2009