Glossogeny and phylogeny: cultural evolution meets genetic evolution

W. Tecumseh Fitch

School of Psychology, University of St. Andrews, St. Andrews, Fife KY169JP, Scotland

Evolutionary theorists since Darwin have been interested in the parallels and interactions between biological and cultural evolution. Recent applications of empirical techniques originally developed to analyze molecular genetic data to linguistic data offer new insights into the historical evolution of language, revealing fascinating parallels between language change and biological evolution. This work offers considerable potential toward unified theories of genetic and cultural change.

Darwinism and cultural change

A suggestive parallel between genetic evolution and cultural change has been recognized for 150 years. Cultural variants – phenotypic behavioural units such as new words, styles, customs or beliefs – can be 'inherited' by a process of imitation and learning and thus passed down the generations [1]. Since Darwin's time, it has been realized that this analogy between cultural inheritance and biological (genetic) inheritance is particularly good in the case of language change. Language change, for example the process by which Latin transformed over centuries into Spanish, French and Italian, is sometimes termed 'language evolution', but this invites confusion between distinct levels of evolutionary change: biological and cultural. Students of language evolution use the term 'glossogeny' to distinguish historical linguistic change from phylogeny (biological evolution: change in gene frequencies) [2].

Historical linguistics was already a mature discipline in 1850, and glossogeny provided an important source of inspiration for Darwin as he pondered how small changes could accumulate over time to lead to large differences in the contemporary descendents, and he proposed that 'if we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world.' Further pushing the analogy with evolution, Darwin suggested that 'some ancient languages had altered very little and had given rise to few new languages, while others had … given rise to many new dialects and languages' [3].

Language goes molecular

At the dawn of the molecular era, Darwin's analogy was placed on a firmer empirical ground when geneticists began to compare patterns of genetic relatedness along with linguistic classifications of the languages of different cultures. Initially using blood group polymorphisms as an index of genetic relationships, and later adding further enzymes and proteins, Cavalli-Sforza et al. [4,5] analyzed genetic relationships among ?42 cultural groups. Adopting the comprehensive but controversial classification of human languages developed by comparative linguist Joseph Greenberg, they found that genetic trees and linguistic family trees show a very high concurrence. This might not seem surprising: after all, children generally inherit both their genes and their language from their parents, and so we might predict a perfect isomorphism between phylogenetic trees built from either genetic or linguistic data. However, as every parent knows, children learn some words from other children or even unrelated strangers – analogous to horizontal gene transfer in bacteria. Over many generations, we might expect that the linguistic family tree would become obscured by such extralineal 'borrowing'. The close parallels between language and genes found by Cavalli-Sforza et al. showed that, in traditional cultures (thus excluding recent adoptions of nonlocal language, such as English as spoken by Australian aboriginals or Native Americans), the gene–language analogy can be rigorously and quantitatively analyzed, and the relationship is a strong and statistically significant one (reviewed in Ref. [6]). These and other papers led to a rapidly increasing interest in exploring the parallels and interactions between genetic and cultural evolution [7–10].

One problem with this early pioneering work comparing genetics and language was its reliance on Greenberg's classification of the world's languages, which remains controversial. Recently, this line of research has been extended in a series of papers that subject the linguistic data to a direct computational analysis, using sophisticated techniques stemming from modern phylogenetics and genomics. Evolutionary biologist Mark Pagel et al. [11] have applied tree-building techniques originally developed to analyze molecular phylogenetic data to a list of 200 'core vocabulary' items – frequently used meanings that are resistant to change and thus tend to be cognate in related languages. First, and reassuringly, these authors found that the resulting trees map quite nicely onto to classifications inferred by linguists. For example, computer analysis of the Indo-European language family (which includes English, German, Gaelic and Spanish, as well as Hindi, Latin, Greek, Russian and many other Eurasian languages) nicely duplicates the 'family tree' of Indo-European developed by linguists over more than a century. However, these computer techniques allow more than just

the topology of the tree to be analyzed: they also allow the rates of glossogenetic evolution to be estimated.

Frequency and stability

For many years, it has been suspected that frequently used words (such as 'you', 'hand', 'eye' or 'two') are relatively resistant to change, whereas rarer words such as 'holster' or 'knuckle' tend to change faster. This observation can be put on a firmer empirical footing by directly correlating the rate of lexical replacement (analogous to mutation rate in a genetic system) and word frequency. Pagel *et al.* [11] found, as predicted, that lexical 'evolution' occurs faster in rare words.

This effect is not limited to word forms but applies to linguistic rules as well: Erez Lieberman, working with evolutionary theorist Martin Nowak and his colleagues at Harvard [12], found that the rate of 'regularization' of past-tense rules in English also correlates with frequency. The regular past tense in English is formed by adding '-ed' to the end of the verb (e.g. walk → walked), but many 'irregular' verbs use different rules to form the past tense (e.g. run → ran; be → was).

Tracing the historical records for 177 Old English verbs, these researchers found that rarely used forms were more likely to 'regularize' by adopting the '-ed' rule than frequently used verbs. Thus, although the frequent 'run/ran' pair has survived, the less-frequent Old English 'help/holp' has become regularized: we now use 'helped' as the past tense of 'help'. The relation between frequency and regularization turned out to allow a surprisingly succinct analysis: 'The half-life of an irregular verb scales as the square root of its usage frequency: a verb that is 100 times less frequent regularizes 10 times as fast.' [12]. The fact that high-frequency words change slowly has an interesting corollary: the slowest-changing words are replaced at rates comparable to the fastest-changing genes. Thus, human cultural evolution is capable of generating replicators with a fidelity overlapping that of DNA [11].

The future of glossogenetics

Finally, most recently, Pagel and colleagues [13,14] have broadened their linguistic database to include further language families (Bantu, Indo-European and Austronesian). They applied techniques for estimating the rate of change after branching events, originally developed (again) for molecular data. These analyses strongly suggest that glossogeny does not occur at a uniform rate but instead is 'punctuational' in the sense that change occurs more rapidly during and just after divergence events than after the new dialect or language is mature and independent. These results provide another suggestive analogy to biological evolution and rapid divergence after speciation, and again demonstrate the virtue of a quantitative data-driven approach over previous qualitative analyses. Together, these studies illustrate the potential for applying both theoretical models and empirical tools

developed in a genetic context to important aspects of human language, but they only begin to address the many fascinating questions one can ask about the relationship(s) between genetic and cultural 'evolution'. How far can one press the analogy without strain? How adequate is the parallel between genes as units of selection in biological, darwinian processes with words, or syntactic rules, in cultural evolution? How does the existence of borrowing between languages complicate the essentially lineal analyses attempted to date? In particular, is cultural evolution more like bacterial evolution, with abundant horizontal transfer, than eukaryotic evolution? Although the factual and logical basis for evolutionary models of culture are as strong, today, as the evidence Darwin brought to bear on biological evolution in 1859 [15], much has changed since that time. Thus, despite a blossoming of interest in theoretical modelling (e.g. Refs. [10,16]) and empirical tools (e.g. Ref. [11]), the study of cultural evolution from a biological perspective remains in its infancy, and we can expect substantial progress in the coming decade.

References

1 Dawkins, R. (1976) *The Selfish Gene,* Oxford University Press
2 Hurford, J. (1990) Nativist and functional explanations in language acquisition. In *Logical Issues in Language Acquistion* (Roca, I.M., ed.), Foris Publications, pp. 8, 5–136
3 Darwin, C. (1859) *On the Origin of Species.* (1st edn), John Murray
4 Cavalli-Sforza, L.L. *et al.* (1992) Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5620–5624
5 Cavalli-Sforza, L.L. *et al.* (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. U. S. A.* 85, 6002–6006
6 Cavalli-Sforza, L.L. (1997) Genes, peoples, and languages. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7719–7724
7 Feldman, M.W. and Laland, K.N. (1996) Gene-culture coevolutionary theory. *Trends Ecol. Evol.* 11, 453–457
8 Cavalli-Sforza, L.L. and Feldman, M.W. (1983) Cultural versus genetic adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 80, 4993–4996
9 Rogers, D.S. and Ehrlich, P.R. (2008) Natural selection and cultural rates of change. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3416–3420
10 Richerson, P.J. and Boyd, R. (2005) *Not by Genes Alone: How Culture Transformed Human Evolution,* University of Chicago Press
11 Pagel, M. *et al.* (2007) Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449, 717–721
12 Lieberman, E. *et al.* (2007) Quantifying the evolutionary dynamics of language. *Nature* 449, 713–716
13 Pagel, M. *et al.* (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* 314, 119–121
14 Atkinson, Q.D. *et al.* (2008) Languages evolve in punctuational bursts. *Science* 312, 588
15 Mesoudi, A. *et al.* (2004) Is human cultural evolution Darwinian? Evidence reviewed from the perspective of 'The Origin of Species'. *Evolution Int. J. Org. Evolution* 58, 1–11
16 Odling-Smee, J. *et al.* (2003) *Niche Construction: The Neglected Process in Evolution,* Princeton University Press