nature

NEWS & VIEWS

LINGUISTICS

An invisible hand

W. Tecumseh Fitch

Quantitative relationships between how frequently a word is used and how rapidly it changes over time raise intriguing questions about the way individual behaviours determine large-scale linguistic and cultural change.

In the eighteenth and nineteenth centuries, linguistics was considered a thoroughly historical science, focusing on how languages such as English or Sanskrit changed through time. By uncovering rules governing phonological change, historical linguists reconstructed dead protolanguages such as Indo-European — an ancestral dialect spoken some 10,000 years ago that diverged into a wide variety of modern languages, including Hindi, Russian, Spanish, English and Gaelic. The crowning achievement of these early linguists was a family tree of languages that became an inspiration for Charles Darwin as he pondered biological evolution (Fig. 1). Unfortunately, many historical linguists entertained quasi-mystical ideas: August Schleicher, the German author of the first great comparative grammar of Indo-European languages, believed that languages are living things, and Jacob Grimm posited a *Sprachgeist* — an internal spirit of a language driving it to change along certain lines.

Twentieth-century linguists rejected such fanciful notions, and emphasized the capacity of individuals to produce and understand utterances. Noam Chomsky famously characterized this as a conceptual shift from a historical preoccupation with 'E-language' (a set of externalized utterances) to an emphasis on 'I-language' (principles internalized by the language learner)[1]. The focus by modern linguists on the individual capacity to acquire and use language (ontogeny) led to a flowering of research, allowing the biological and neural underpinnings of language to be studied coherently, and opening the door to consideration of how the language faculty evolved biologically (phylogeny). But this approach left behind the traditional questions of the cultural evolution of individual languages (glossogeny[2]) that tantalized earlier generations of linguists.

Elsewhere in this issue, two papers[3,4] revisit these issues from a fresh perspective. Both concern language change, and come from laboratories of well-established evolutionary theorists. Both analyse historical linguistic data to show that patterns of change depend strongly on the frequency with which words are used in

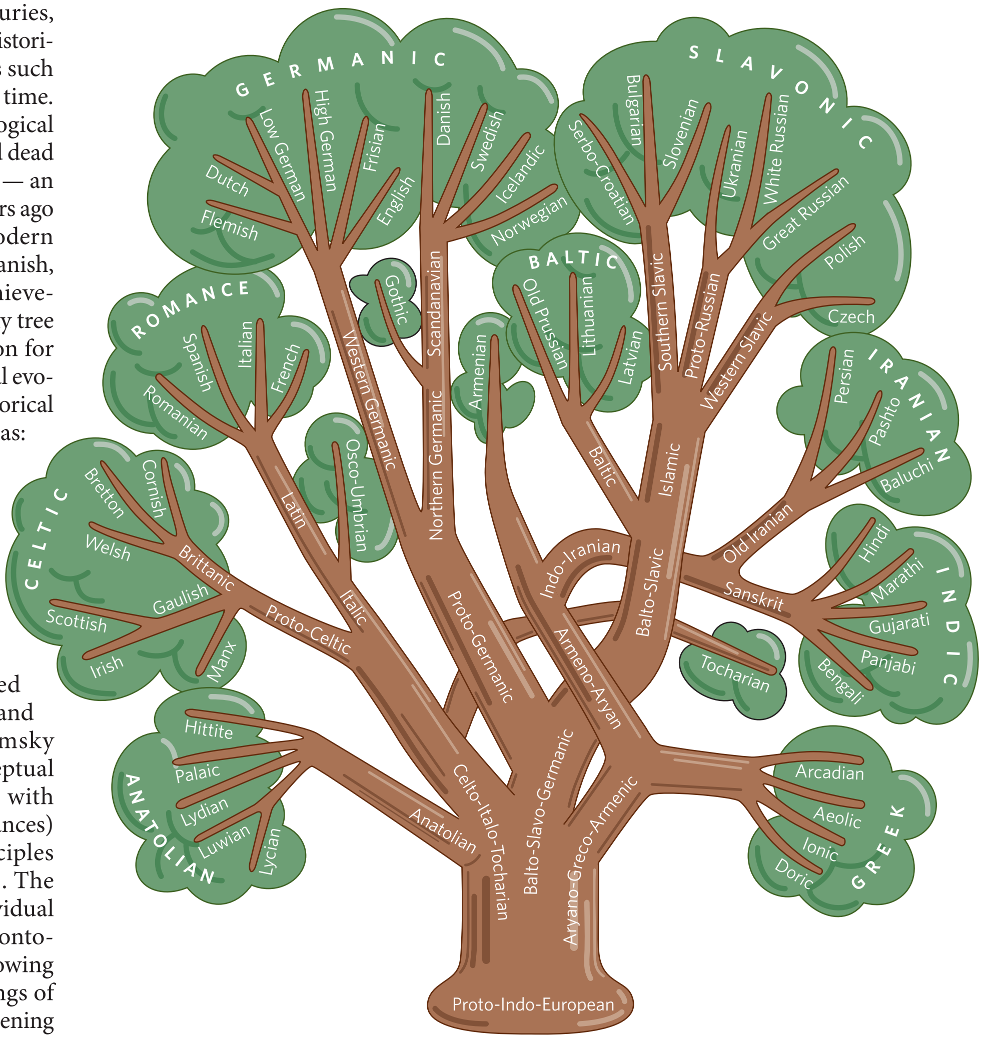

Figure 1 | The glossogenetic tree of Indo-European language. The words of language are not inherited biologically, but are passed on culturally through learning. This process of 'cultural evolution' generates a hierarchical tree of relationships among languages, here illustrated by the Indo-European family. Just as descent with modification in biological evolution (phylogeny) leads to phylogenetic trees, so the analogous process in language change (glossogeny) can lead to glossogenetic trees.

discourse, as measured from large contemporary databases. Lieberman *et al.* (page 713)[3] consider the cultural evolution of the English past-tense marker '-ed'. In Old English, this was just one of many different rules used to indicate times gone by. Today, the other once-widespread rules remain only as irregular residues, such as 'fly/flew/flown'. By tracing their disappearance, the authors derive an exact quantitative relationship between the frequency of verb use and the speed of this pruning process: a verb used 100 times more often than

Box 1 | The invisible hand in language change

Language change at the 'macroscopic' level is often influenced in counter-intuitive ways by 'microscopic' changes in how individuals use language. A nice example is found in the historical phenomenon of pejoration in words referring to women, where respectable words acquire negative connotations over the centuries. A 'hussy' was once a perfectly respectable housewife, and 'wench' just meant 'young woman', but both terms now connote a woman of loose morals. And 'lady' — once used just for a woman of noble birth — is now the standard term for any woman.

Intriguingly, words for men generally don't suffer the same fate, and sometimes even improve their connotations ('knight' originally meant just a boy or retainer). Parallel patterns have occurred in other languages (for example, as with the German *Weib*, which suffered the fate of 'wench').

The most obvious explanation for this phenomenon is that language users (or at least those who have historically been responsible for recording language — men) are consistently misogynistic. But a more convincing 'invisible hand' explanation invokes a simple individual rule: when talking to or about women, err on the side of politeness[8]. Given two options, one normal and one polite ('hussy' versus 'lady'), this rule, if applied widely and consistently, leads to 'lady' becoming the common form. 'Hussy' or 'wench', by comparison, become ever-less polite over time. The best intentions lead to pejoration as an unintended consequence. **W.T.F.**

another will regularize 10 times more slowly.

Pagel *et al.* (page 717)[4] take a broader approach, quantifying the rate at which related words (such as 'water' in English and *Wasser* in German) have been replaced by other forms (such as the French *eau*) during the cultural evolution of 87 Indo-European languages. Using frequency data from four different language corpora — sets of texts representing patterns of usage in English, Spanish, Russian and Greek — and sophisticated tree-based statistical methods over the whole glossogenetic tree, Pagel's group derives a relationship holding over millennia. The relationship explains 50% of the variation in replacement rates between different words — a level of statistical power rarely observed in the social sciences, particularly across a wide range of cultures.

Despite significant differences in their methods, both papers document the same general pattern: frequently used words are resistant to change. Relatively infrequent inflections such as 'help/holp' became regularized, whereas high-frequency English verbs retained their ancestral irregular state ('go/went' or 'be/was'). More generally, terms that occur with high frequency in Indo-European languages (such as 'one', 'night' or 'tongue') are resistant to substitution by new phonological forms. The realization that frequency of use has a significant role in language change is nothing new[5–7]. But the use of sophisticated methods developed in bioinformatics and genomics to quantify these relationships is an important step forward. We can expect similar approaches to be applied to a wide variety of languages, to determine whether the specific patterns uncovered in these papers also hold in non-Indo-European languages, such as Chinese or the Dravidian languages of southern India.

Documenting these relationships remains descriptive, not explanatory: quantifying their form does not tell us why such regularities exist. Schleicher might have characterized the situation with an E-linguistic metaphor, in terms of a struggle for survival among different word forms. But from an I-linguistic perspective, Chomsky would retort that the underlying explanation must come down to the individuals who learn and use the language. Pagel and his colleagues[4] consider two such possibilities. First, new phonological forms might arise less often for high-frequency words because errors of perception, recall or production are less common for frequently used words. Alternatively, such cultural 'mutations' might occur uniformly, but frequency of use would affect the probability of new variants being adopted by the population.

Crucially, these two possibilities are not necessarily in conflict. An adequate explanation for glossogenetic phenomena must incorporate individual and collective levels of description, and show why they are necessarily related. Part of the challenge is the apparent circularity of explanation inherent in evolutionary systems, where the output of one generation serves as the input to the next, and causes are no longer neatly separated from effects. Such difficulties are compounded in cultural evolution: glossogeny represents an intermediate descriptive level that changes on a slower timescale than ontogeny, but faster than phylogeny. Although we, as individuals, don't generally invent words or grammatical forms, our usage (mispronunciations, or slight semantic shifts, for example) will affect their future transmission, and the population's usage en masse will determine their fate across many generations. Thus, human languages such as French or Swahili are neither natural (as envisioned by Schleicher) nor artefacts made intentionally by individual humans. Like economic, political and religious systems, they are phenomena of a third kind[8].

Although this distinction is intuitive (we apply it when distinguishing 'natural' languages such as English from 'artificial' ones such as Esperanto or C++), it remains relatively unexplored in linguistics. Theories concerning such phenomena have been developed more extensively in economics: the necessity of explaining 'macroscopic' phenomena in terms of quite different 'microscopic' behaviours was first discerned by Adam Smith, who used the evocative metaphor of the "invisible hand" to describe how individuals working to maximize profit in their own self-interest benefit society as a whole by driving up the standards of goods and services on offer.

Where should we look to gain a deeper understanding of the invisible hand in the cultural evolution of language? A promising future direction is provided by recent attempts to fuse theoretical models of cultural evolution[9] to experimental investigations of social learning in the laboratory[10,11]. Experimental investigations of 'iterated learning' — similar to the game of Chinese whispers, where one participant's output serves as input for the next — can provide empirical data to inspire, and constrain, our theories. Sophisticated new theoretical models enable language-learning 'agents' to have both innate biases (in the form of so-called bayesian priors) and powerful statistical learning systems capable of discovering and using environmental regularities[12]. Such models demonstrate the possibility of a very indirect and sometimes non-intuitive relationship between the regularities emerging at the level of a whole population and the underlying generating forces (Box 1). These forces are individual behaviour and learning (social usage) and innate constraints (in Chomsky's terms, a 'language acquisition device', often called universal grammar).

An important implication of this new synthetic approach to glossogeny is that universals of language are not identical to universal grammar — although obviously related, the two concepts should not be conflated. Another is that cultural evolution can proceed independently of either phylogenetic evolution (the interests of our genes) or our own individual goals and interests. As the papers in this issue[3,4] make clear, cultural evolution can make language easier to learn by filtering out irregular 'noise'[13] — as Lieberman *et al.* wryly point out[3], every "rule is the tombstone of a thousand exceptions". But it can also preserve the irregular cases that make learning a new language difficult: every surviving exception remains a stumbling block for a thousand new language learners.

Nonetheless, as Pagel *et al.* suggest[4], some of the most persistent 'cultural replicators' — memes[14] — evolve as slowly as some genes. By documenting and quantifying such effects, this work opens the door to a diverse range of theoretical and empirical investigations. If there is ever to be a science of memetics[15–17] to rival that of genetics, it should proceed along these lines: combining careful quantitative analysis of well-documented linguistic changes with sophisticated theoretical models capable of taking into account the multilayered complexity of cultural evolution. ■

W. Tecumseh Fitch is in the School of Psychology, University of St Andrews, Fife KY16 9JP, UK.
e-mail: wtsf@st-andrews.ac.uk

1. Chomsky, N. *Knowledge of Language: Its Nature, Origin, and Use* (Praeger, Westport, CT, 1986).
2. Hurford, J. in *Logical Issues in Language Acquisition* (ed. Roca, I. M.) 85–136 (Foris, Dordrecht, 1990).
3. Lieberman, E., Michel, J.-B., Jackson, J., Tang, T. & Nowak, M. A. *Nature* **449,** 713–716 (2007).
4. Pagel, M., Atkinson, Q. D. & Meade, A. *Nature* **449,** 717–720 (2007).
5. Bybee, J. & Hopper, P. (eds) *Frequency and the Emergence of Linguistic Structure* (Benjamins, Amsterdam, 2001).
6. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949).
7. Heine, B., Claudi, U. & Hünnemeyer, F. *Grammaticalization: A Conceptual Framework* (Univ. Chicago Press, 1991).
8. Keller, R. *On Language Change : The Invisible Hand in Language* (Routledge, New York, 1994).
9. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (Univ. Chicago Press, 1985).
10. Kalish, M. L., Griffiths, T. L. & Lewandowsky, S. *Psychonom. Bull. Rev.* **14,** 288–294 (2007).
11. McElreath, R. *et al. Evol. Hum. Behav.* **26,** 483–508 (2005).
12. Kirby, S., Dowman, M. & Griffiths, T. L. *Proc. Natl Acad. Sci. USA* **104,** 5241–5245 (2007).
13. Deacon, T. W. *The Symbolic Species: The Co-evolution of Language and the Brain* (Norton, New York, 1997).
14. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, 1976).
15. Dennett, D. C. *Darwin's Dangerous Idea* (Simon & Schuster, New York, 1995).
16. Blackmore, S. J. *The Meme Machine* (Oxford Univ. Press, 2000).
17. Mesoudi, A., Whiten, A. & Laland, K. N. *Evolution* **58,** 1–11 (2004).

ORGANIC CHEMISTRY

Zippier synthesis in water

Masayuki Inoue

Complex toxin molecules are the ultimate challenge for organic chemists — even successful syntheses often involve an impractical number of steps. A biologically inspired reaction might simplify things.

Nature creates many complex organic molecules, but some of the most spectacular are the brevetoxins and ciguatoxins produced by marine microorganisms[1]. Brevetoxins are the toxic component of dramatic natural phenomena known as red tides (Fig. 1) — toxic algal blooms that kill huge quantities of fish — whereas ciguatoxins are found in the flesh of fish, and can cause widespread poisoning in humans if they enter the food-chain. Many researchers have invested enormous effort into making these intricate molecules, and have come up with heroic syntheses that stretch the limits of organic chemistry. Reporting in *Science*, Vilotijevic and Jamison[2] now point us towards a potential synthetic short cut. They have found that pure water 'zips up' simple starting materials to make part of the core structures of the molecules, without the need for any other additives.

Brevetoxins and ciguatoxins belong to the family of ladder-like molecules known as polycyclic ethers, which contain fused rings of carbon and oxygen atoms (Fig. 2a, overleaf). A remarkable feature of these molecules is the striking regularity with which the oxygen atoms bridge the carbon framework. In the 1980s it was proposed[3,4] that the key step in the biosynthesis of these compounds is a single 'zip reaction' of a linear precursor molecule into the polycyclic product, catalysed by an unknown enzyme (Fig. 2a). Since then, chemists have tried to emulate such cyclization cascades in the laboratory using a variety of organic solvents, artificial catalysts and specially designed precursor molecules. But these efforts have met with only partial success[5,6].

Such cyclization reactions pose two big problems. First, even for simple cases in which a single ring of atoms is formed, two potential products can be made: a smaller ring of five atoms, or a larger ring of six atoms (Fig. 2b). The cyclization of ten consecutive rings, such as occurs in brevetoxin-A, could therefore theoretically generate 1,024 (2^{10}) products. Designing a reaction that yields only the desired product out of all of these options is challenging, to say the least.

The second, more serious problem stems from the intrinsic preference of cyclization reactions to make one product rather than the other. Organic chemists have known for many years that smaller rings form more readily than their larger counterparts. But assuming that brevetoxin-A is formed from a linear precursor, consecutive 'larger' ring sizes appear in the structure where smaller sizes would be expected. The ring structure of brevetoxin-A could thus theoretically be the least favourable option of all 1,024 possible products. The synthesis of brevetoxin-A in a polycyclization reaction without the aid of an enzyme starts to look unrealistic.

Nevertheless, synthetic chemists have succeeded in reversing the intrinsic selectivity of these cyclization reactions by carefully designing their starting materials. An early example of this involved manipulating the epoxide reacting group in the starting material[7]. Epoxides contain two carbon–oxygen (C–O) bonds; during cyclizations, the product of the reaction — either a five-membered or a six-membered ring — is determined by which of the two bonds breaks. But if an olefin, which contains a carbon–carbon double bond (C=C), is attached to the epoxide, the C–O bond adjacent to the olefin breaks preferentially (Fig. 2b). This is because the olefin partially donates some of its electrons to the adjacent C–O bond, destabilizing the bond so that it ruptures if treated with acid in an organic solvent. Strategic attachment of an olefin to the epoxide can thus direct cyclization reactions to form six-membered rings. This was a crucial synthetic discovery. But it is obviously not how polycyclic ethers are assembled in nature, because no possible brevetoxin precursor uses directing groups such as an olefin.

Vilotijevic and Jamison's method[2] provides a considerable advance over earlier approaches to polycyclic ether formation. They show that a series of three epoxides in a molecule can be reacted in one step to zip together ladder-like structures of six-membered rings, simply by heating them in water (Fig. 2c). The reaction occurs with a good chemical yield at neutral pH, and in the absence of olefin directing groups or other epoxide activating agents.

The reaction mechanism for this unusual transformation[2] has not yet been established, but it is known

B. BACHMAN/SPL

Figure 1 | Crimson tide. Toxic algal blooms wash ashore in Queensland, Australia.