I am working on a new, up-to-date edition of my textbook *Schools of Linguistics*, first published in 1980 and translated into several languages. So far I have completed a draft of the new first chapter, as seen below. (There are a scattering of rewrites and new references throughout, but the bulk of new material, relating to developments in the subject since 1980, is naturally towards the end of the chapter.) Any comments, significant references I have missed, etc. would be very gratefully received at my e-mail address <sampson@cantab.net>. The eventual published book will of course acknowledge any help received.

Geoffrey Sampson

1 Linguistics begins: the biological paradigm

Language has undoubtedly been studied and thought about for as long as human beings have had thinking time to spare from the demands of simple survival. R.H. Robins's *Short History of Linguistics* (1997) begins with the Ancient Greeks. But as an autonomous academic subject or discipline, distinct from philosophy or literary studies (while of course having connexions with them and with other subjects), it is reasonable to place the beginnings of linguistics early in the nineteenth century. Two of the leaders of the discipline later in that century, Hermann Osthoff and Karl Brugmann (1878), described it as having been half a century old in 1868.

The impetus for the new discipline stemmed from the realization, about the end of the eighteenth century, that the familiar European languages share similarities with languages of Northern India and intermediate territories, notably the ancient Indian language Sanskrit, which could not have been a consequence of cultural contacts during the historical period between the respective language communities (since there were none), and must therefore reflect descent from a common ancestor-language spoken before written records began. We now call that language Proto-Indo-European ("PIE"), and believe it to have been spoken about six thousand years ago, probably in the steppe lands north of the Black and Caspian Seas (Anthony and Ringe 2015). From PIE descended Sanskrit, Latin, Greek, "Proto-Germanic", and various other recorded or hypothesized ancient languages, and different modern languages in turn derive from these: Latin is the ancestor of the "Romance" languages, that is French, Italian, Romanian, etc.; from the hypothesized Proto-Germanic derive English, Dutch, German, and the Scandinavian languages (other than Finnish, which is not Indo-European); and so on for the other branches. The first recognition of the unity

of the Indo-European family is commonly (in Britain, at least) credited to the Welshman Sir William Jones, in a talk to a learned society in Calcutta in 1786, though others have played down Jones's significance (Swiggers 2017).

(Although the word "linguistics" was occasionally used in English in the second half of the nineteenth century, it did not come into widespread use until later – in Britain it became a familiar term only in the 1960s. Previously, the nearest equivalent word was "philology", and its practitioners were "philologers" or "philologists" – but philology referred mainly to the study of a culture through its literature, and only derivatively to the study of language itself. When the term "linguistics" became usual in Britain, at first it tended to occur in the phrase "general linguistics", to emphasize that it aims to describe the phenomenon of human language in general rather than the properties of a particular language – these days the single word "linguistics" is understood in that sense.)

It is easy for a newcomer to linguistics today to dismiss the philologers of the nineteenth century as pedants motivated more by a love of accumulating facts for their own sake than by a feeling for the excitement of scientific theory-construction. That is partly because the main focus of the discipline since the early twentieth century has been on what is called "synchronic" linguistics: the analysis of languages as they exist at a given point in time, commonly the present, ignoring (as their speakers ignore) the route by which they arrived at their present form. It is natural for students of present-day realities to see the nineteenth-century focus on "diachronic" linguistics – how languages have changed through time – as inherently old-fashioned. But, natural or not, this dry-as-dust idea of nineteenth-century linguistics is quite incorrect. It is true that the enormous effort devoted to the historical study of the Indo-European language-family was inspired partly by intellectual fashion, as opposed to considerations of rational scientific research strategy. The change of emphasis from "classical philology" to the new subject of linguistics occurred first in Germany (indeed, throughout the nineteenth century linguistics was mainly a German pursuit); and the flourishing of Indo-European (in German, "*Indogermanisch*") linguistic studies went hand in hand with the general intellectual and artistic movement of late-eighteenth- to mid-nineteenth-century Germany known as Romanticism, with its rejection of the classical tradition and its celebration of indigenous ethnic and cultural roots. (The link between linguistics and these wider intellectual and aesthetic currents is particularly clear in the work of such men as J.G. Herder (1744–1803), the leading figure in the *Sturm und Drang* movement in

literature, collector of folk songs and relics of the early culture of the Germanic people, one of whose most influential works was his *Treatise on the Origin of Language* (1772), and Jacob Grimm (1785–1863), among the founders of Germanic linguistics, and collector with his brother Wilhelm of a world-famous anthology of traditional German fairy tales.) Since race, language, and culture were taken to be intimately related, reconstruction of the prehistory of the Germanic and other language-stocks was attractive to the Romantic temperament.

But there was more to the situation than this. The history-centred outlook of nineteenth-century linguistic scientists was related to the general state of science at the time.

It is commonly the case in the history of science that at any given time there are a few outstandingly successful branches of science which are seen as models of what a science should be, so that scholars trying to investigate some new field will almost inevitably adopt the methods and ideas of the "model" sciences. The philosopher of science Thomas Kuhn (1962) coined the term "paradigm" to suggest how, at a given period, thinking about a particular subject is conditioned by some more or less coherent system of ideas which act, not so much as explicit tenets of a scientific theory, but as unspoken assumptions about the range of possible hypotheses available for a scientist to consider. For Kuhn, the most important scientific advances occur on the rare occasions when scholars manage to break out of these mental straitjackets by rejecting assumptions which their predecessors did not even feel the need to defend (as when Einstein responded to problems about the observed speed of light by arguing that space, time, and mass are observer-dependent rather than absolute quantities). And Kuhn's term "paradigm" can be used more broadly, so that the outlook of practitioners of a successful science constitutes a paradigm not only for that science itself but also for less established sciences. The nineteenth century contained two outstandingly successful scientific paradigms in this sense.

The first of these was mechanistic physics, according to which all phenomena could be described by simple, deterministic laws of force and motion – so that complete knowledge of the state of the world at the present moment would in principle allow one to predict perfectly all its future states (the view expressed by Pierre-Simon de Laplace in his 1820 work *Théorie analytique des probabilités*, and given up by the 1930s with the acceptance of quantum theory); the second was the biological theory of evolution by natural selection, which emerged from a great upsurge of interest in natural history during the eighteenth and nineteenth centuries,

and culminated in Darwin's *Origin of Species* (1859) and the storm of controversy aroused by that book.

From physics, philologists took the notion of describing the history of sound-changes occurring in a language in terms of "laws" which apply uniformly to whole ranges of examples, rather than discussing individual words in the anecdotal, case-by-case way in which a historian (in the ordinary sense) treats individual persons or events. One of the first such discoveries, for instance, was the Proto-Germanic consonant shift commonly called Grimm's Law (though in fact first stated by the Dane Rasmus Rask in 1814), whereby Proto-Indo-European consonants changed in the Germanic branch in accordance with the following rules:

PIE		Germanic
voiceless stops /p t k/	>	voiceless fricatives /f θ x/
voiced stops /b d g/	>	voiceless stops /p t k/
voiced aspirates /bh dh gh/	>	voiced stops /b d g/

Since, in other branches of Indo-European, the consonants remained unchanged, or developed differently (thus PIE voiced aspirates became voiceless aspirates /ph th kh/ in Ancient Greek, which in turn became voiceless fricatives in modern Greek), the Germanic consonant shift produces many cases of words alike in meaning but containing distinct consonants in different languages: compare, for example, the initial consonants of Greek *thyra* and English *door*, Greek *genos* and English *kin*, Greek *pous* and English *foot*.[1] "Grimm's Law" reduces hundreds of cases like these to three simple formulae.

The term *Lautgesetz*, "sound law", was first used by Franz Bopp, in a talk given in 1824 at the Berlin Academy and published in its Transactions the following year (Wechssler 1900: 51–2). Bopp's sound laws were only statements of general tendencies, and Bopp did not feel it necessary to provide explanations for cases which failed to follow the general rule. But, as the nineteenth century grew older, the concept of "sound law" took on more and more the rigorous character of genuine scientific laws like those of physics. According to the group known as *Junggrammatiker*

1 Corresponding to English *door* we find German *Tür*; that is because, after Proto-Germanic had split into the early forms of the various modern Germanic languages, a further consonant shift in the German branch (but not the English branch) altered (among other things) /d/ to /t/.

or "Neogrammarians" in the last quarter of the century, apparent counterexamples to a sound law were permissible only if they could be explained by a sub-law of their own. For instance, we saw that Grimm's Law predicts that PIE /p t k/ become voiceless /f θ x/ in Germanic, but there are exceptions where the Germanic consonants are voiced. One case is in the word for "over, above", which had /p/ in PIE (compare Greek *hypér*), yet in English (a Germanic language) the word is *over* rather than the expected **ofer*. The explanation was formulated by the Dane Karl Verner (1876), in what has been described (by Schmalstieg 1988) as "probably the most famous linguistic article of all times": voicing occurred whenever the consonant in question was preceded in PIE by an unstressed syllable. The stress in Greek *hypér* and its PIE equivalent was on the second syllable, thus the /v/ of *over* is not exceptional but regular. (After the voicing had happened, in the Germanic branch stress shifted, hence English *óver* rather than *ovér*.) Verner's discovery encouraged Hermann Osthoff and Karl Brugmann (1878) to formulate the general principle, in the introduction to a journal they founded to promote the new approach:

> Every sound change, inasmuch as it proceeds mechanically, takes place according to exceptionless laws ... all words in which the sound subject to the law occurs under the same conditions are without exception affected by the change.

This Neogrammarian principle remains the default assumption about phonological change to this day.

There have been dissenters, who did not believe that a purely mechanical principle could be capable of accounting for the behaviour of thinking human beings. For a long time, almost all of those who actively rejected the Neogrammarian principle were scholars in Romance-speaking countries, including Italy and France, such as Giuliano Bonfante (1947), or were students of Romance languages, such as Hugo Schuchardt (1885). (Some of them grouped themselves under the banner "neolinguistics" in contrast to "neogrammarian".) One example was the French-Swiss linguist Jules Gilliéron, who was writing in the years around 1920, and is often said to have put forward the anti-Neogrammarian slogan *Chaque mot a son histoire*, "Each word has its own history", implying that the words of a language do not evolve by undergoing laws which apply across the board to the whole vocabulary. It is true, of course, that many aspects of a word's history will be specific to that word, such as

developments in its meaning, or changes in its social associations, for instance becoming regarded as poetic, or as vulgar; and it is also true that a language or dialect will sometimes borrow a form from a sister dialect which has undergone a different series of sound-changes (so in modern standard English we have two words *church* and *kirk*, the latter specialized to refer to Presbyterian churches, because the sound-law which changed /k/ to /ʧ/ in certain environments in Southern England never applied in the English of Scotland and Northern England). But this does not amount to evidence against the idea that when a sound-change occurs in a language or a dialect, it applies across the board in that language or dialect. Gilliéron certainly thought of himself as an opponent of the Neogrammarians, but the truth, as William Kretzschmar (2002: 85) put it, is that Gilliéron's linguistics "does not so much attack the Neogrammarian position as it substitutes a different set of questions for those addressed by the Neogrammarians". (Kretzschmar also, incidentally, noted, *op. cit.*: 84, that he had failed to find the slogan "each word has its own history" anywhere in Gilliéron's writings.[2]) And other opponents of the Neogrammarians seemed similarly to miss their targets. The controversy has been discussed by Lia Formigari (2018), who sees the dissidents as having supposed, wrongly, that the Neogrammarians believed in laws that were more absolute even than the laws of physics. (The law of gravity says that material objects will fall towards the centre of the Earth, but we don't take it as refuted if a ping-pong ball is momentarily swept upwards by a gust of wind – other physical laws interact with the law of gravity.)

A rather different challenge to the Neogrammarian principle has come more recently from William Wang's theory of "lexical diffusion"; I defer discussion of this to p. 000 below.

While mechanistic physics provided one paradigm for linguistics, however, the influence of biology was greater. As German scholarship came to distinguish between the *Naturwissenschaften* and *Geisteswissenschaften* – between the natural and moral sciences, or in modern terms between the "sciences" and the "arts" or "humanities" – linguists were anxious to align themselves with the former: but, if linguistics is to be a natural science, then a "language" must be some kind of entity which can be described objectively along with the rest of the furniture of the natural world. It will not be adequate to interpret the term "language" as just a convenient way of referring

2 The belief that Gilliéron used this slogan has been disseminated by the widely-read book Iordan–Orr (1970: 170), which attributed the words to Gilliéron without saying where he used them.

to aspects of the purely subjective intellectual life of a nation, as one adopting the "humanities" rather than "science" approach might be inclined to do.

The solution of many nineteenth-century linguists was to regard languages as an order of natural organisms, on a par with plants and animals. Thus, Bopp (1827: 1) wrote:

> Languages must be regarded as organic bodies [*organische Naturkörper*], formed in accordance with definite laws; bearing within themselves an internal principle of life, they develop and they gradually die out, after, no longer comprehending themselves, they discard, mutilate, or misuse ... components or forms which were originally significant but which have gradually become relatively superficial appendages.

Similar views were expressed by August Pott a few years later (1833: xxvii):

> A language is in a constant state of change throughout its life: like every organic object [*organische Naturgegenstand*], it has its periods of gestation and maturation, times of accelerated and of slackened growth, its prime, decay, and gradual extinction ...

It is difficult, now, to see how Bopp's "no longer comprehending themselves" could ever have been more than a rhetorical flourish (although cf. p. 00 below). For the rest, though, these remarks are by no means unreasonable. Although languages are in some sense a product of men's minds, they do seem to have a life of their own, rather than being consciously created artefacts like a symphony or an aircraft design. It was clearly not by any process of conscious decision on the part of its speakers that the Old English of pre-Conquest days developed successively into Chaucer's English, Shakespeare's English, and now the different varieties of modern English. And groups of languages have "family trees" just as groups of biological species do. As we saw above, French, Italian, Rumanian and other Southern European languages descend from Latin, while English, German, and Norwegian, for instance, descend from "proto-Germanic", and Latin, Proto-Germanic, and various other known or postulated ancient languages descend from a still more ancient Proto-Indo-European. This cannot fail to remind us of the situation in biology where, say, Man, chimpanzee, and gorilla all descend from an extinct species of ape, while cat, lion, and tiger descend from an

extinct proto-feline, and proto-ape, proto-feline, and others themselves share a common ancestor further back in geological time. Already at the beginning of the nineteenth century scholars such as Friedrich von Schlegel (1808: 28) and Jacob Grimm (1819: xii) had suggested that the discipline most closely cognate with the new science of "comparative grammar" was comparative anatomy. The *Stammbaum* or "family tree" theory of linguistic evolution was first formally expressed by August Schleicher (1861) almost simultaneously with the appearance of Darwin's *Origin of Species* (published in England in 1859, in German translation in 1860); Schleicher's friend Ernst Häckel, an important early evolutionist, drew his attention to Darwin's book, and Schleicher responded by publishing a short treatise on *Darwin's Theory and Linguistics* (1863), in the form of an open letter to Häckel, arguing that linguistics should be regarded as one of the natural sciences to which Darwin's theory applies. (Schleicher did not say so, but it can be argued that, historically, Darwinism owed as much to linguistics as *vice versa*; see Hayek 1960: 59; Newmeyer 1975.) The linguist's language-families, languages, dialects, and idiolects[3] correspond to the biologist's genera, species, varieties, and individuals. Languages and language-families, like species, compete with one another in a "struggle for survival" (consider, in the British Isles for instance, how English has spread at the expense of the Celtic languages: Cornish and Manx are extinct, Welsh and Scottish Gaelic live on but lose ground steadily to English, Irish is kept alive in a small Gaeltacht like a protected species in a game reserve); and, on a world scale, Schleicher saw the Indo-European language-family as having reached a dominant position linguistically, as Man has become dominant zoologically.

In one respect Schleicher even argued, with justice, that the validity of the evolutionary account can be confirmed more easily for language than with respect to the plant and animal kingdoms. For the biologist it is relatively difficult to establish that the ancestor-species which he postulates in order to explain the relationships between modern species ever really existed, since they have long ago disappeared, leaving only scanty and ambiguous traces in the form of fossils. Because the time-scale of change is so much shorter in the case of language, and because writing was invented millennia ago, proto-languages can sometimes be studied directly rather than merely hypothesized. We possess plenty of documents not only in the modern Romance languages but in their proto-language, Latin, and in many of the intermediate stages; no one could claim that Latin is a figment of the linguist's

3 The term *idiolect* refers to the speech habits of an individual person.

imagination, as the notion of a common ancestor for Man and ape was pooh-poohed by opponents of the biological theory of evolution. (Indeed, Sir Charles Lyell, 1863: ch. 23, used this argument to make evolutionary theory seem more plausible in biology.)

Some readers may feel that to claim that linguistics is literally a branch of biology alongside botany and zoology is self-evidently unreasonable. Languages are not material objects: one can infer the existence and nature of languages, or even idiolects, only via the behaviour of speakers, not by direct observation as in the case of plants or animals. As Bonfante famously put it (1946: 295), "Languages are historical creations, not vegetables". This might seem to rule out the possibility of treating Darwin's theory as anything more than, at best, a suggestive metaphor for linguistics. But that would be too glib; Schleicher may have been wrong, but he meant what he wrote. What distinguishes life from non-life is still a deeply mysterious question. The biologist who has thought most deeply in our own time about what we mean by "life" is Richard Dawkins, and Dawkins argues (2006: ch. 11) for a concept of life which includes a realm of "memes". This term has been taken up by popular culture in ways that probably deviate from Dawkins's careful scientific analysis, but in Dawkins's sense the domain of memes would seem to include languages or their elements (though language is not a topic which interests Dawkins much), in which case he would regard languages as in some sense constituting a living domain parallel to the domain of gene-based life.

Until late in the nineteenth century, then, the historical approach was the natural one for the study of language, and historical linguistics looked like one of the frontiers on which exciting new scientific advances could confidently be expected. As the century neared its end, for a number of reasons this came to seem less likely.

The first problem had to do with the directionality of change. It is central to the evolutionary view of biology that the replacement of old species by new is not merely a process of random changes (even if the individual variations on which evolution depends are random), but rather is a movement from less to more "fit" – variations which succeed in spreading are those which give their possessors an advantage in the struggle to survive and produce offspring, while disadvantageous traits are eliminated. Nineteenth-century linguists often took it for granted that linguistic change was similarly "directional". According to Rask (1818: 35–6), languages become steadily simpler over time:

> The language which has the most sophisticated grammar is the purest, most

> original, oldest, nearest to the source, because grammatical inflexions and endings are eroded in the development of new languages, and they require a very long time, and a certain mingling with other peoples, to evolve and organize themselves again. Thus Danish is simpler than Icelandic, English than Anglo-Saxon; and Modern Greek bears the same relation to Classical Greek, Italian to Latin, German to Gothic, and similarly in all cases known to us.

Rask's claim seems to be a purely empirical generalization about observed facts, and it is certainly correct for the cases he cites (except that German is not now held to be a direct descendant of the extinct language called Gothic). It is not clear whether Rask intended it as a strong hypothesis about all possible cases of language change – the clause about "evolving and organizing themselves again" seems to allow for some cases of languages moving in the direction of greater complexity. As the biological analogy became increasingly persuasive, though, so the directional view of language-change came to play a more central role in linguists' theorizing. One strand in the directional view was the notion that languages could be classified into a small number of types, usually three: *isolating* languages, in which each word consists of a single unchanging root (Chinese and Vietnamese are standard examples); *agglutinating* languages, in which words include affixes as well as root, but the division of a word into root and affixes is clear (e.g. Turkish, where *sevişdirilmek* means "to be made to love one another", and the word divides into *sev-* "love", *-iş-* "reciprocal", *-dir-* "causative", *-il-* "passive", and *-mek-* "infinitive"); and *inflecting* languages (e.g. Sanskrit, Ancient Greek, Latin, and the other languages cited by Rask as relatively complex), where a single word includes a number of "units of meaning" but one cannot assign these meaning-units to distinct portions of the entire word: thus Latin *sim* is the first person singular present subjunctive of the verb "to be", but one cannot divide the word up into separate pieces meaning "be", "subjunctive", "present", or the like. (This Latin example is extreme: often one can split at least the root from the inflexional ending fairly unambiguously in Latin, and even in this case one can at least say that it is the *-m* which expresses first person singular. But the three classes were intended as "ideal types" of language, and it was recognized that real languages fall between the extremes provided by the scheme.)

This three-way classification (though not the names of the categories) was first expounded by Friedrich von Schlegel (1808: ch. 4), and disseminated more widely by his brother August (A.W. von Schlegel 1818: 14–16), who treated the inflecting type as

the highest.[4] August von Schlegel further divided inflecting languages into two subclassess, *synthetic* and *analytic* languages – the former being inflecting languages in the fullest sense, the latter including some characteristics of the isolating type (prepositions in place of case-endings, subject pronouns in verb conjugations); and he treated the history of the Romance subfamily of languages as a process of decay from synthetic Latin to analytic modern languages such as French.

Not everyone who discussed typology agreed that inflecting languages were "better" or "higher" than isolating languages – Wilhelm von Humboldt (1836: sec. 24) suggested that both types have their advantages. By the mid-century, though, we find Schleicher (1848) claiming that the prehistory of languages involved a regular development from isolation through agglutination to inflexion, and that this is an evolution from less to more perfect.

There is a problem here: Rask claimed that the direction of language change was towards greater simplicity – i.e. from inflexion to isolation – while for Schleicher linguistic evolution proceeded from isolation to inflexion. But Schleicher resolved the apparent contradiction by an argument which for him was inspired by the philosopher G.W.F. Hegel, in the German world regarded as a towering intellectual figure, but which also has a close parallel in (subsequent) biological theories.

According to this argument, we must distinguish in the evolution of Man between the period of prehistory, when Man is controlled by the same laws as the rest of animate and inanimate nature, and the historical period, when Man's intellect reaches the point at which he develops free will and thus rises above the blind laws of nature. Now, Schleicher argues (following Hegel 1837: 62–3), the evolution of language presumably went hand in hand with the evolution of intellect, so that the perfection of language and of intellect would have occurred together: literature began only when Man's intellect had fully evolved, so that the earliest forms of the classical languages are highly inflexional languages – we can infer that they were preceded by agglutinating and isolating stages only by *a priori* reasoning, and by comparison with

4 The three-way classification is still commonly used today, although with no suggestion that the different types are of unequal merit. The distinction between isolating and agglutinating types, however, seems relatively superficial: the only reason why we say that Turkish has words several morphemes long while in Vietnamese words are not distinct from morphemes is because Turkish has vowel-harmony, and it is convenient to use the term "word" for the domain over which vowel-harmony operates. (A "morpheme" is a minimal meaningful unit: so the word *cat* consists of a single morpheme, whereas *undeliverable*, for instance, is made up of three morphemes, *un-*, *deliver*, and *-able*. For vowel harmony, see 000.) The distinction between these two types taken together and the inflecting type, on the other hand, does seem a basic one, though the distinction is gradient rather than sharp.

the languages of tribes who are still pre-literate today. Once the historical stage is reached, intellect becomes autonomous and ceases to depend on the superficial form of language, and language is therefore free to regress to "lower" forms: hence Rask's observation.

There are obvious and serious objections to this. If a race as intelligent as the Chinese, possessing a sophisticated civilization as old as or older than those of the Greeks, Romans, or Indians, could manage with a language which, in the Chinese classical period, was close to the isolating extreme, then how can we say that Man needed to develop inflecting languages in order to realize his intellectual potential?[5] And to what extent can we assimilate linguistics to biology, if the recorded history of languages displays exclusively decay rather than improvement?[6] But the notion that the human mind is a development which cannot be explained within the framework of natural evolution, and which frees Man from the dictates of natural laws, is very reminiscent of Alfred Russel Wallace's objections, later in the century, to Darwin's theory as applied to Man (Wallace 1870; cf. Eiseley 1958: ch. 11); and the view that language decays once the achievement of free will liberates it from evolutionary laws is parallel to the widespread, and surely very plausible, idea that products of human intelligence such as medical knowledge, by suspending the law of the survival of the fittest, must lead to lower average levels of human physical excellence.

Furthermore, when Schleicher's view of linguistic development as perfection followed by decay was attacked by Wilhelm Scherer, Scherer explicitly appealed to contemporary biology as authority for his own view of language changes. Before Lyell published his *Principles of Geology* in 1830–33, most geologists explained the existence of successive geological strata containing fossils belonging to different levels of organic complexity in terms of the "catastrophist" theory associated with Georges Cuvier, which asserted that prehistory fell into a number of distinct epochs separated by destructive upheavals, after each of which new forms of life were divinely created *ex nihilo*. Lyell replaced this view with the "uniformitarian" doctrine that the changes attested by geological evidence result from the same kinds of process that we can

5 One answer to this which is occasionally put forward is that since Chinese script is not alphabetic, the classical spoken language might have included inflexions which were not recorded in writing. George van Driem (2021: 102–4) attributes this view to the German Carl Lepsius (1860), and apparently agrees with it. I believe Karlgren (1920) may be a more relevant reference than Lepsius, but whoever was responsible for the idea, it is untenable for several reasons (see e.g. Sampson 2021).

6 This latter point seems to have been a clear and unresolved contradiction in Schleicher's thought: cf. Jespersen (1922: 72–3), or Kurt Jankowsky's disagreement (1972: 101) with Delbrück on the relation between Schleicher's Hegelianism and his view of linguistics as biology.

observe happening in our own day. Scherer (1868: x) accordingly argued, as against Schleicher, for uniformitarianism also in linguistics:

> We can hardly shut our eyes much longer to the realization that the distinction between evolution and decay, or – as it has also been put – between the nature and the history of language, rests on a fallacy. For my part, I have everywhere observed only evolution, only history.

Although discussions of linguistic evolution focused chiefly on morphology, directionality was argued also for phonological change. As late as 1893, Jan Baudouin de Courtenay (a Polish linguist of aristocratic French descent, who worked out his ideas at the university of Kazan', in Russia) argued that languages tend to replace sounds formed relatively far back in the mouth and throat with sounds formed nearer the teeth and lips: notice for instance that pharyngal and uvular consonants were common in the Semitic languages (which are among the earliest languages for which European linguists possessed records), but are rare in languages which emerged more recently, and compare the various fronting rules that have applied to velar consonants in the Slavonic languages. For Baudouin, this represented a "humanizing" tendency, by which languages are losing the beastlike sounds that characterized their primaeval origins.

So there was a widespread acceptance of the view that language change is governed by fixed developmental laws (even if there was some disagreement about *which* direction languages moved in); in this respect, the biological paradigm fitted linguistics. Towards the end of the century, though, the directional view of linguistic change became much less popular. In the same work in which he argued for directionality of phonological development, for instance, Baudouin de Courtenay (1893: 24) contradicted his predecessors by suggesting that morphological changes reveal only random "oscillations". Certainly there are counter-examples to the view that languages in the historical period uniformly become less inflexional and more isolating. For instance, modern spoken French is arguably nearer the inflecting end of the scale than was mediaeval French: consider how plurality is indicated by vowel alternation in phrases such as [lə garsõ] v. [le garsõ] "the boy"/"the boys", as against the more agglutinative situation in earlier French [le garson] v. [les garsons]. And it is easy enough to refute Baudouin's own claim about phonetic directionality: consider, for instance, the replacement of apical by uvular *r* in standard French, or the

replacement of [t n] by [k ŋ] in many environments in southern dialects of Vietnamese. Nowadays it is difficult to see the process of linguistic change at any level as more than a series of random movements in no particular direction; and, in that case, the analogy with biology falls to the ground.[7] Some scholars kept faith with the directional view well into the twentieth century: Holger Pedersen supported Baudouin's theory of "humanization" of phonology in 1924 (Pedersen 1924: 281–2), Otto Jespersen maintained his belief that natural selection makes languages steadily simpler as late as 1941. But later in the twentieth century few scholars continued to hold such views.

If one gives up the idea that language change regularly proceeds in a particular direction, it becomes difficult to follow Schleicher in applying to language Darwin's concepts of "natural selection" and "struggle for survival": what, in language, will correspond to the biological concept of aptitude for survival? And in fact the expansion of certain languages at the expense of others seems to be explainable very adequately in terms of social factors, so that there is no room for an explanation referring to the intrinsic properties of the languages themselves. Even if it were true that English is in some sense a "simpler" or "more advanced" language than Welsh, the fact that English has been expanding and Welsh contracting is undoubtedly due to the fact that England has been a centre of power and wealth and Wales has not. The work of Peter Trudgill (e.g. 2001) suggests that structural simplicity is not a *cause* but a *consequence* of a language-community becoming economically and politically influential: Trudgill argues that highly complex languages can only survive in isolated communities with few external links, so that they are rarely acquired as second languages. Where the criterion of intrinsic simplicity and the criterion of social prestige conflict in determining which of alternative languages will spread, the latter is almost invariably decisive – consider for instance the failure of Esperanto since 1887 to become a widespread second language, despite its extreme simplicity and the considerable concrete advantages that would have followed from its universal adoption.

(For much of the twentieth century, linguists had treated it as axiomatic that natural languages – leaving aside artificial languages like Esperanto – are all equally

7 William Labov (e.g. 2001: ch. 8) has shown that there are regularities in how particular sound-changes occur within a society, for instance women tend to take the lead in adopting a new speech-habit and men to follow. But Labov does not attempt to predict which particular sounds will change.

complex. However, that axiom seems to have been adopted for purely ideological rather than scientific reasons; once it was challenged early in the present century, linguists quickly gave it up. See e.g. Miestamo et al. 2008; Sampson 2017: 8–9, 143ff.)

The abandonment of the directional assumption went hand in hand with a growing emphasis on the principle that language changes originate with individual speakers. Indeed, although I have written as if it was the empirical refutation of directionality which undermined the view of linguistics as a branch of biology, this is a *post hoc* rationalization rather than an accurate account of the theoretical developments of the late nineteenth century. It would probably be truer to say that the linguists of the time first adopted the general methodological approach that language must be treated in terms of the psychology of individual speakers, rather than in terms of a *Sprachgeist* having some kind of existence above and beyond individuals, and only subsequently noticed empirical evidence which tended to refute the view they were giving up. (Philosophers of science are familiar with the idea that relevant data are often noticed only *after* adoption of the theory for which the data are decisive; see e.g. Lakatos 1970: 158–9.) Furthermore, although those who stressed individual psychology certainly believed that their approach was incompatible with the view of linguistics as biology, they seem to have been wrong in this – as I shall show shortly.

The point of stressing individual psychology was as a reaction to the views of earlier, Romantically-inspired linguists such as Grimm, who held that the nature of the language of a nation was determined by its *Sprachgeist* or *Volksseele* ("genius of the language", "race-soul" – these and similar terms were used more or less interchangeably to denote some kind of spiritual entity embodying the aesthetic, moral, and intellectual values of a nation).[8] It was his belief in a conscious *Sprachgeist* that allowed Bopp to write of languages "ceasing to comprehend themselves" (p. 00 above). This mystical but popular idea was attacked already in 1858 by Rudolf von Raumer (1858: 374):

> Whenever linguistic change, particularly sound change, is discussed, people are apt to appeal straight away to the "*Sprachgeist*" and its marvels. ... But ... the "*Sprachgeist*" does nothing of itself, separately from men, rather all changes in a language are brought about by men themselves.

8 The nineteenth-century concept of *Sprachgeist* has been analysed by Rosemarie Lühr (2016).

The same point was hammered home repeatedly and forcefully by the Neogrammarians. Osthoff and Brugman (1878: xii) held

> that language is not a thing, standing outside and above men and leading its own life, but has its true existence only in the individual, and that therefore all changes in the life of a language can originate only with individual speakers.

and Hermann Paul, in his standard textbook *Prinzipien der Sprachgeschichte*, wrote (1880: 11):

> All psychic processes are executed in individual minds and nowhere else. Neither race-mind [*Volksgeist*] nor elements of the race-mind such as art, religion, etc. have a concrete existence, and consequently nothing can occur in them or between them. So away with these abstractions.

This last quotation makes it particularly clear that we are dealing here not with a modification of linguistic theory necessitated by the observation of awkward data, but rather with a very general shift in conceptions of the nature of social phenomena. However, from views such as those quoted it might well seem to follow that one cannot assimilate linguistics to biology as a science treating a class of natural objects. Paul (1891: 118) accordingly attacks Schleicher, "who, being wedded to the view that linguistics is a natural science, was unable to succeed in forming any correct views about the nature of language development". According to Kurt Jankowsky (1972: 147), "For Hermann Paul, linguistics was a historical discipline, not a natural science".[9]

However, it is surely wrong to assume, because the *Sprachgeist* notion is admittedly nonsensical, that Schleicher's equation of linguistics with biology must necessarily be given up too. For Schleicher, a language corresponded to a biological species, and an idiolect in linguistics to an individual member of a species in biology. We do not accuse the biologist of mysticism because he recognizes as a theoretical construct the species "carrot", even though all he tangibly observes are individual carrots. The analogue of the principle that linguistic changes originate in individual psychology is the claim that, in biology, it is innovations in the DNA of individuals

9 One problem in interpreting comments like this last one is that history was seen as the clearest case of a *Geisteswissenschaft*, so that calling a subject "historical" involves some ambiguity between the diachronic/synchronic and the arts/science distinctions.

which lead to the evolution of new species (rather than variations in individual genomes being caused by the striving of the species as a whole towards some goal) – and this is a cardinal tenet of Darwinist theory.[10]

Apart from the lack of consistent direction in language change, another real problem for the evolutionist view of language had to do with the causation of changes. The difficulty here lay not so much in accounting for innovations in morphology, which might be explained with some plausibility as developments towards a simpler system, or as restoring intelligibility where unstressed case endings or the like had been eroded in rapid speech; the problem concerned rather the sound-shifts (such as Grimm's Law), which, by causing the pronunciations of words to diverge among different groups of speakers, seem to be quite arbitrary, unmotivated hindrances to communication. Sound-changes were law-like in the sense that they applied to all words containing the relevant sounds, *in a given language at a given point in time*; seen from a wider perspective, however, they were merely isolated, idiosyncratic events: Grimm's Law applied just to the Germanic dialect of Indo-European in a particular century, not to all languages at all times. We would not be impressed by a physicist who invented one law of gravity for seventeenth-century Italy, another for present-day England, and so on.

Quite a number of theories were advanced as to the causation of phonological change (see summaries in Oertel 1902: 189ff.; Jespersen 1922: chs 14–15). One view was what would nowadays be called the "substratum" theory: when a group of people adopt a new language (that of their conquerors, for instance), they are likely to carry habits of pronunciation over from the old language to the new. This theory is certainly correct in many cases: the Welshman's English accent is heavily influenced by the phonology of Welsh, even though most Welshmen today do not speak that language. It is emerging from genome studies, a type of data which linguists are only just beginning to take on board,[11] that this kind of effect may have been more significant in human prehistory than one might have supposed. Comparison of

10 Paul would not have accepted this point as robbing his objection to Schleicher of its force. Paul's disagreement with his predecessors was based not merely on a different view of social phenomena, but on novel presuppositions about the nature of science in general. He was influenced by the "descriptivist" view of science, advocated in his time by the physicist and philosopher Ernst Mach, according to which only observables really exist, while theoretical entities – atoms, for instance – are convenient fictions introduced to abbreviate statements about observables. Thus, for Paul, the biologist who recognizes a species "carrot" might well be accused of mysticism (see e.g. Paul 1880: 37). However, descriptivism (on which see e.g. Nagel 1961: ch. 6) does not seem to modern philosophers to provide an adequate account of the nature of scientific theories.

11 A useful but rather idosyncratic survey is in van Driem (2021).

mitochondrial DNA (inherited only through the female line) with Y-chromosome DNA (present only in males) suggests that, after various territories were first settled by humans, a usual pattern was for female descendants of the original populations to remain in the same place while groups of men often invaded other territories, mating with the local women and sometimes eliminating the local males (see e.g. Metsapalu et al. 2004, Balaresque et al. 2010, Lippold et al. 2014). Examining correlations between language and genome for a sample of speakers of Indo-European languages, Zhang et al. (2019) claim to find evidence of a tendency for individuals to have adopted their father's language but with their mother's pronunciation: they grew up speaking Fatherese with a Motherese accent, which would make sense in a scenario where men were aggressors and women provided child-care.

But much of the world has happily become more peaceable in historical times, and many sound-changes have clearly happened within one language, independently of other languages. The Great Vowel Shift which occurred in English between the fifteenth and eighteenth centuries, for instance (the series of sound-changes which are responsible for the fact that the modern English pronunciation of the vowel-letters contrasts with their value in Continental languages) cannot be explained in terms of substratum theory.

Another idea was to extend to phonology the theory that languages tend to become simpler: sound-changes might be caused by a tendency to greater ease of articulation. Again this explanation works well for some cases (e.g. elision of unstressed vowels or of consonants in consonant-clusters); but there are counter-examples. It is generally agreed that front rounded vowels are less natural than back rounded vowels, yet French regularly developed front rounded [y ø] from Latin back rounded [u: o:] (e.g. *lūnam* > *lune*, *nōdum* > *nœud*). The language-families that have been studied in depth seem to have undergone very many sound-changes, so the ease theory seems to imply that the earliest languages must have been unusually full of difficult sounds and sound-combinations, which feels implausible.[12] And the theory says nothing as to why particular ease-increasing changes happen when and where they do. Words spelled in English with initial *kn-* and *gn-*, like *knee*, *gnaw*, originally began with *k* and *g* sounds which later dropped; in German, *Knie* "knee" is still

12 On the other hand, those who argue that *grammatical* change is a process of simplification are not forced to postulate perversely complex proto-languages. E.H. Sturtevant (1947: 107–9) pointed out that it is in the nature of sound-change to create grammatical irregularity; so, if one accepts the axiom that sound-changes occur, *other* linguistic changes might be motivated by a tendency towards simplicity while the overall complexity of a language remains more or less constant.

pronounced [kni:]. To quote Leonard Bloomfield (1933: 385),

> The English change of [kn-, gn-] to [n-] seems natural, after it has occurred, but why did it not occur before the eighteenth century, and why has it not occurred in the other Germanic languages?

Grimm himself explained the law that bears his name in terms of the psychology of the Germanic race:

> ... from one point of view the sound-shift strikes me as a barbarity and a rejection of civilization, which other, more peacable peoples avoided, but which is connected with the Germans' mighty progress and struggle for freedom which inaugurated the Middle Ages and was to lead to the transformation of Europe (1848: 417)

> The Roman Empire had decisively lost its strength after the end of the first century, ... and the invincible Germanic race was becoming ever more vividly aware of the unstoppability of its advance into all parts of Europe. ... How could such a forceful mobilization of the race have failed to stir up its language at the same time, jolting it out of its traditional rut and exalting it? Does there not lie a certain courage and pride in the strengthening of voiced stop into voiceless stop and voiceless stop into fricative? (1848: 437)

Many of Grimm's contemporaries accepted this kind of explanation, but majority scholarly opinion has long rejected them. Some of the same changes which Grimm took as symptomatic of courage and vigour were treated by Karl Müllenhoff (1892: 197) as indicating laziness or enervation, and later research has not established any empirical correlations between particular sound-changes and particular psychological characteristics.

Others explained sound-shifts in terms of anatomy or physical geography. Quite late in the century, Hermann Osthoff (1879: 16) claimed that "modification of the vocal organs is in general the real cause of historical sound-changes in languages". And there have been sporadic explanations put forward since, some better-argued and more plausible than others, for particular phonological facts. Heinrich Meyer (1901) suggested that phonetic changes of the kind described by Grimm's Law might

correlate with relatively energetic breathing, which could in turn be caused by living in a hilly region; Hermann Collitz (1918) took this idea up, and quoted several other cases of sound-shifts in different parts of the world which tended to confirm it. Leonard Brosnahan (1961) published impressive data on correlations between existence of the interdental /θ ð/ phonemes and prevalence of particular blood-groups in the languages and populations of different areas of Europe. Gerry Knowles (1973: 116ff.) was inclined to take seriously the idea that the "adenoidal" quality of the Scouse (Liverpool English) accent was caused by bad sanitary conditions in nineteenth-century Liverpool housing. A group of Swiss linguists (Blasi et al. 2019) have published evidence that the labiodental [f v] sounds, rare in ancient languages, have become commoner because of anatomical changes caused by modern dietary habits. But despite scattered hints like these, no-one has produced an elaborated general theory of physical influences on phonology.

A number of twentieth-century linguists writing in French, notably André Martinet (1955), felt that at least some sound-changes will be predictable from systematic properties of the phonological structure of a language. Phonemes will shift in ways that increase their contrast with other phonemes, and in particular mergers between phonemes will be avoided unless the language has few word-pairs that are kept apart by that phoneme contrast: in English, /θ/ and /ð/ could merge, since there are few word-pairs like *thigh* and *thy*, but /t/ and /d/ will not merge because there are many pairs like *tip* and *dip*. This is a plausible-looking idea; but it is a testable hypothesis, and the evidence is against it (King 1967, Sampson 2019).

It would not be fair to say that physical theories of sound-change, or for that matter theories like Grimm's in terms of national solidarity, have been decisively refuted. Scholars have simply given up working on such theories. The unpopularity of explanations which appeal to the concept of national psychology may have more to do with rejection of nationalism among educated circles in recent decades than with considerations of rational research strategy. On the other hand, one may well feel that "the truth will out": scholars discussed possible correlations of phonological changes with extraneous factors over a period of many decades, and if they did not end by producing a convincing theory of such correlations, then perhaps there are none to be found, and sound change really is largely random. Leonard Bloomfield (1933: 385–6) felt justified in drawing this conclusion, and later scholars have not dissented. The Neogrammarians of the late nineteenth century felt that sound "laws", to be worthy of the name, must in principle be independent of particular times and places

(Jankowsky 1972: 155–6), so that if one group of speakers applied Grimm's Law while others did not, there must have been some independently verifiable special circumstance applying to that group which regularly causes a sound-change of a similar kind whenever it occurs. Linguists since the mid-twentieth century, on the other hand, have not even felt a need for a theory of the causes of sound-shifts. Paul Postal (1968: 283) found it clear that

> ... there is no more reason for languages to change than there is for automobiles to add fins one year and remove them the next, for jackets to have three buttons one year and two the next ...

There are actually two separate issues here: why sound-changes occur at all, and why the particular changes that occur are what they are. The former point might be adequately answered by considering how teenagers in our own time adopt novel speech styles, including new phonological features, as a badge of solidarity with their generation. William Labov's work (1972: ch. 1) on Martha's Vineyard, a holiday island off the Massachusetts coast, has shown that people in general, not just young people, unconsciously use their speech patterns to reflect their social allegiances. (The indigenous dialect of this island includes unusual versions of the /ai au/ diphthongs; Labov found that the extent to which native Vineyarders use these pronunciations correlates with the individuals' self-identification with the island community or with the wider American society represented by the summer visitors.)

The urge of young people to differentiate themselves from their elders must surely be more or less universal, and it would be enough on its own to explain why plenty of sound-changes occur, whether or not some changes have other causes; once the older generation has died off, what was teenage-talk becomes the standard language. But *any* innovation will create distinctiveness, so if this is the reason for sound-changes there is no need to look for factors determining the particular changes that occur. Perhaps one admired youngster happens by chance to speak in some slightly unusual way, and this is seized on and exaggerated by others who look up to him. Teenage-talk might feel like an undignified explanation for a development as ancient and momentous as the creation of the Germanic languages, but if teenage-talk is a widespread phenomenon in our own time, uniformitarianism requires us to suppose that it was prevalent also among our Germanic-speaking linguistic ancestors millennia ago.

Even if nineteenth- and twentieth-century linguists failed to develop a theory of the causes of sound-change, in a sense that did not disturb the analogy with biological evolution. Darwin also had to treat the occurrence of modifications in the offspring of given parents as an unexplained axiom; it was not until much later that people began to understan either the biochemical mechanisms by which the "genetic blueprint" is transmitted from generation to generation, or the phenomena (meiosis, and radioactivity) which could lead to random modifications of that blueprint. However, Darwin's theory provided a satisfying explanation for so many other biological truths that people might be willing to take this gap in the argument on trust. In linguistics, with no clear directionality of change and no clear analogue of "fitness for survival", the failure to find causes for change was yet another factor making an evolutionary theory of language unattractive.

It is also true that, between the 1860s and the end of the nineteenth century, various counter-arguments (largely founded on what eventually proved to be false assumptions about the unknown mechanisms of genetic inheritance) made Darwin's theory seem steadily less convincing to biologists, including its author (Eiseley 1958: 209ff., 233ff.). This is no doubt another reason why the equation of linguistics with biology was abandoned – by the end of the nineteenth century biology no longer seemed to offer such an attractive paradigm as it had forty years earlier.

In 1880, Hermann Paul could still insist (1880: 20) that the historical approach to language was the only scholarly method available for linguistic study. But then, despite his disagreement with Schleicher, Paul did still believe in the applicability of the concept of natural selection to language.[13] By the end of the century, though, the data for historical linguistics came to seem a mere assembly of sound-shifts which had occurred for no good reason and which tended in no particular direction; and the science to which linguists had looked as a model for their attempts to reduce this chaos to order had itself fallen on hard times.

Consequently, for much of the twentieth century, the historical approach that so many had found exciting in the nineteenth century rather lost its oomph. Thomas Kuhn saw the history of any science as consisting of occasional "revolutions" whereby the reigning paradigm is replaced by a new paradigm, separated by periods of "normal science" during which researchers fill in details without calling the current paradigm

13 "... utility [*Zweck*] plays the same role in the evolution of linguistic usage as Darwin attributed to it in the evolution of organic nature: the greater or lesser fitness [*Zweckmässigkeit*] of newly arising forms is decisive for their retention or extinction" (Paul 1880: 32).

into question. In the twentieth century, historical linguists continued to work on the detailed histories of words and structures in various languages, but few new general ideas emerged within the historical approach to change the overall complexion of the discipline.

Towards the end of the twentieth century this changed.

One development in historical linguistics lay in far greater awareness of the realities of language-families other than our own. Throughout the nineteenth century, the "primary goal" of linguistics, as Patrick Honeybone (2005) says, "was the reconstruction of Proto-Indo-European" in particular, and long after the nineteenth century many linguists supposed that Indo-European is a typical example of the world's language-families. But in the twentieth century linguistics turned more of its attention to non-European languages, such as the indigenous languages of American Indians; and although these linguists were working in the synchronic rather than historical framework, their findings have led recent historical linguists to appreciate that Indo-European is a very untypical family.

For one thing, it split at an early date into a large number of major branches – subfamilies at the level of Germanic, Italic, Slavonic, and so on;[14] Johanna Nichols (1990) counts about fifteen (including now-extinct subfamilies like the Tocharian spoken in Sinkiang in the fifth to eighth centuries AD), of which eight to ten survive today. On a world scale, Nichols finds, it is quite unusual for a family to split into more than two surviving subfamilies.[15] And it seems that after PIE broke up, speakers of the various branches tended to move far enough in the so-called *Völkerwanderungen* (migrations of peoples) to lose contact with one another. This created a favourable situation for the present-day historical linguist, since it makes it easy to distinguish cases where words are borrowed between languages in modern times from words which show correspondences because they have descended separately from the same PIE ancestor root. (Latin /p/ in *pater* corresponds to English /f/ in *father*, but to English /p/ in *paternal*; we can set the latter correspondence aside because we know that English-speakers borrowed *paternal* from Latin after they had become literate and educated, thousands of years after the Germanic and Italic subfamilies had separated. In some language-families, after the proto-language split the resulting speech-

14 "Italic" covers Latin and its Romance descendants, together with some extinct sister-languages of Latin, such as Umbrian.

15 Johanna Nichols uses the term "stock" for what I am calling a language-family, that is a maximally inclusive set of languages which can all be shown to descend from a shared ancestor, and she uses "family" for what I am calling a "subfamily".

communities remained in contact with one another so that borrowing of words might happen at any period, blurring the evidence for reconstruction of the ancestor-language.) Also, it was possible to discover Grimm's Law because there are many sets of cognate words akin to *pater ~ father* in the various branches of Indo-European. That is not always true in other families. Rainer Vossen (2017) tells us that "in African languages it is often not so easy to find even two examples of a given regular correspondence" – not just because data on some of these languages are meagre, but because some languages contain very rare phonemes: "In Central Khoisan, for instance, consonants such as *tx* or *dz* occur in just a handful of words; in inter-language comparison, these words may not be cognates throughout so that regularity cannot be proven." And our limited knowledge of cultures distant from us in time and space can make it difficult to see that words are cognates. According to Paul Black (2017), Geoffrey O'Grady has proposed five separate reconstructed roots in the Proto-Pama-Nyungan ancestor of many Australian languages, each of which means "raw" in some descendant languages and "one" in others – a European might wonder what connexion Australian aborigines saw between these concepts.[16]

There have even been controversies about whether the Neogrammarian principle which has become axiomatic for Indo-Europeanists, that sound-laws apply across the board to all relevant words in a language, is equally valid for other language-families. William Wang and his associates have argued (e.g. Wang 1969, 1977, Wang and Lien 1993) that the history of the differentiation of Chinese into its modern dialects shows that sound-changes spread gradually from one word to another across a vocabulary, and sometimes cease to operate before all the words to which they could apply have been affected. Edwin Pulleyblank (1982, esp. p. 405) replied to Wang, arguing that Wang's data can be explained differently, but it is not clear why a basic principle that seems cogent with respect to one language-family should appear so debatable with respect to another.

Finally, it is obvious that the Indo-European family is very unusual on a world scale in containing several proto-languages such as Latin which were written and hence can be studied directly. For most language-families only modern descendant languages are available for study. Indo-Europeanists do not always appreciate how favourable their research situation is.

Greater awareness of non-Indo-European languages has been one factor helping linguists to appreciate that there is more to the history of languages than

16 Unfortunately there seems to be some error in Black's citation of O'Grady (2004: 83–4).

their descent from their ancestor-languages. Consider the concept *Sprachbund*, "language federation", a term which was coined by Nikolai Trubetzkoy (1930), a Russian prince who emigrated after the revolution and became a professor of philology in Vienna. Trubetzkoy used *Sprachbund* for a group of languages with structural similarities which are striking enough to suggest divergence from a common ancestor, but which are known not to have come about that way – the alternative being that the languages *con*verged through mutual contact. For instance, many Balkan languages are regarded as forming a *Sprachbund*, including among others Albanian, Romanian, Bulgarian, Greek, and Turkish – the first four represent four main branches of Indo-European which separated at an early date, and Turkish is not Indo-European at all. One striking common feature is loss of the infinitive. Languages in other parts of Europe express an idea like "I want to go" with a first-person verb followed by an infinitive form (*je veux aller*, *ich will gehen*), and Ancient Greek did the same, but modern Greek has no infinitive and has to say *thelo na pao*, "I want that I go"; other Balkan languages are similar. Even more striking are the similarities between the languages of mainland south-east Asia. For a Chinese-speaker who begins to study Vietnamese, for instance, it would be very easy to take the latter as just one more regional dialect of Chinese, alongside Cantonese, Hakka, and others – and early linguists supposed that these languages did belong to one family. But we now know that they do not: Vietnamese is an Austroasiatic language, sharing a common ancestor with, for instance, the Munda group of languages in north-east India. Vietnamese has a number of high-frequency words that lack Chinese cognates, because they derive from the Austroasiatic ancestor-language, but the vast bulk of Vietnamese vocabulary is borrowed from Chinese; and structurally Vietnamese is much closer to Chinese than to its Austroasiatic relatives (Vietnamese and Chinese have very similar tone systems, the Munda languages are not tone languages; Vietnamese and Chinese are isolating languages, the Munda languages have complex systems of verb and noun inflexion; etc.). Nicole Creanza et al. (2015) have shown that it is common for unrelated but geographically adjacent languages to resemble one another in terms of the sounds they use. Phenomena like these led to the growth of a flourishing branch of the discipline called "areal linguistics", whose founding charter was a classic paper by Murray Emeneau (1956) – more recently, see e.g. Thomason (2000), Hickey (2017).

Some linguists find the earlier exclusive emphasis on ancestral relationships between languages not just one-sided but objectionable, smacking of ideas about

racial purity in societies. Edward Vajda (2020: 733–4) writes:

> ... inherited from the early days of diachronic linguistics is the still-prevalent attitude that proving a genetic relationship between languages represents some sort of pinnacle of achievement in historical linguistics, with evidence of borrowing or other forms of language contact often treated as chaff to be winnowed away ... In reality, genetic relationship is but one of many important facts to be investigated in tracing the history of languages. Modern historical linguists should be interested in all aspects of language history.

Be that as it may, other new developments in historical linguistics have tended to reinforce the parallels between linguistics and biology.

One notable innovation has been "linguistic cladistics", pioneered by Don Ringe of the University of Pennsylvania. In biology, where all life-forms ultimately share a common ancestry, a "clade" is any intermediate shared-ancestry grouping between the individual animal or plant, and the entire realm of life. A species, a genus, or the whole kingdom of animals, are clades of different levels, but there will be "anonymous" clades between, say, a genus and its various species – the single extinct species from which the genus descends may have diversified into two or three species over evolutionary time and each of those species diversified again, before we reach the species that exist today. A clade is a node of the evolutionary tree, in which present-day species are the leaves at the tips of the branches. Before the discovery of how a genome is encoded in the sequence of base-pairs in the "double helix", it was not possible to plot the precise shape of an area of the evolutionary tree in detail: species were grouped into genera by reference to bodily similarities, but those groupings were tentative and debatable. Now, with precise information on the genomes of representative members of a set of related species, biologists can use mathematical techniques to search for a tree which minimizes the number of changes needed to arrive at those genomes from a single ancestral genome, giving an objective reason to believe that this tree accurately reflects the evolutionary history of the life-forms in question.

Ringe and his collaborators apply these techniques to the corresponding issues in historical linguistics. Within the Indo-European language family, for instance, the obvious clades would be language groupings such as the Germanic languages, the Slavonic languages, or the Italic languages. These clades are uncontroversial. But

there has been plenty of debate about the shape of the Indo-European *Stammbaum* at higher levels. For instance, August Schleicher believed that the Italic clade was sister to the Celtic clade (including Welsh, Irish, etc.), so that these languages had a common "proto-Italo-Celtic" ancestor, making them closer relatives to one another than to other Indo-European languages. Geographically this is not as implausible as it might seem: those Celtic languages which survive today are all spoken on the north-west fringes of Europe, but it is known that at an early date the Celts lived in and around Switzerland. Nevertheless, the two subfamilies are not outstandingly similar, and linguists have continued to debate whether Schleicher was right about proto-Italo-Celtic.

Ringe and his team apply the techniques of cladistics to language by encoding an individual language as a set of "characters" taking different values, analogous to the A, T, C, G of the biological genome. A character can be the root used to express a given meaning, a grammatical element, or a sound-change; for instance, the word for "make" derives from an Indo-European root *werǵ-* in Armenian and the extinct Germanic language Gothic, but from a different root in another Germanic language, Old Norse; the superlative ending derives from *-isto-* in Greek, the ancient Avestan language of Persia, and in Germanic languages including English, from *-isṃo-* in Latin and in Celtic languages, and from various other roots in other Indo-European languages that have superlative forms. Clearly, if linguists are right to group "Germanic" languages including Gothic together as a low-level clade, then the fact that Gothic agreed with Armenian but not with other Germanic languages about the word for "make" will be an anomaly in the tree structure (the same vocabulary innovation would have had to occur in separate speech communities) – but, if Gothic were moved out of the Germanic clade, other anomalies would be created. The aim is to find a tree structure which minimizes the number of anomalies.

It is easy to count the number of anomalies in any particular tree structure, but the difficulty lies in locating the "best" tree in an ocean of alternative structures. Ringe et al. (2002) looked at a representative sample of 24 Indo-European languages, encoding each as its values for 370 characters. Mathematically it can be shown that the number of distinct tree structures which can be drawn over 24 leaf nodes is truly colossal – it is roughly equal to the number written as "25" followed by 27 zeros (assuming that all splits are binary – if nodes were allowed to have three or more "daughter" nodes, the number would be higher still). Even a computer cannot work through such a vast space of possibilities one by one, so much of the difficulty of

cladistics lies in developing computational techniques for approximating the "best" tree without examining each individual possibility. But geneticists have devised a number of techniques to address this problem, which Ringe's group adapted for analysing language data (Barbançon et al. 2007).

In the writings which introduced cladistics to the linguistic community (Ringe et al. 2002, Nakhleh et al. 2005), the group found that an almost perfectly anomaly-free tree can be produced for the Indo-European family, and its shape is fairly much as expected; it confirms the hypothesis that Italic and Celtic constituted a clade. But there are anomalies, however almost all of these relate to the same single subfamily, Germanic. Their tentative explanation is that while most of the speciation events in the history of Indo-European were clean splits, after which there was little contact between the separate resulting speech-communities, "part of the IE family, but only a part, must have evolved otherwise than through clean speciation" (Nakhleh et al. 2005: 388): Germanic "was originally a near sister of Balto-Slavic and Indo-Iranian ... [but] at a very early date it lost contact with its more easterly sisters and came into close contact with the languages to the west" (Ringe et al. 2002: 111).

This work also sheds new light on a continuing debate about the *Stammbaum* model in general. Johannes Schmidt (1872) had argued that the family-tree model failed to fit the facts of Indo-European for which Schleicher designed it. There were many cases where some trait was common to two language-groups, say A and B, lying relatively far apart on Schleicher's tree diagram, while being absent from other groups descending from the postulated common ancestor of A and B; but this situation could not be rectified simply by redesigning the tree diagram so as to make A and B adjacent, because it was also the case that B shared some trait missing in A with group C, say. According to Schmidt's "wave theory", such findings could be explained only by abandoning the family-tree theory and seeing the process of linguistic change instead in terms of innovations originating at different geographical points and spreading outwards over arbitrary areas of territory, so that the resulting languages show a pattern of overlapping rather than hierarchically-organized relationships. Ringe et al. (2002: 106–8) identify "a widespread and long-standing conviction among historical linguists that an evolutionary tree (or *Stammbaum*) is hardly ever an appropriate model for the diversification of a language family", but they see their own findings as refuting this radical rejection of the tree model. "Our analysis shows dramatic support for the claim that the divergence of IE was largely treelike" (Nakhleh et al. 2005: 391).

(There is an obvious tension between Ringe's reassertion of the tree model of language evolution, and the *Sprachbund* concept discussed earlier. Ringe et al. (2002: sec. 7.6) discuss this to some extent, but there is clearly more to be said on this point.)

Other scholars have been developing the techniques pioneered by Don Ringe in new directions. Ringe's work on Indo-European takes descent of the modern languages from a single shared ancestor for granted and is concerned only with the history of branching since the original break-up of PIE. But, for instance, Nataliia Hübler (2020) applies related techniques to the so-called "Altaic" languages, whose genetic unity is highly debatable. Until the 1960s it had long been accepted that numerous languages from Turkish across Siberia to Mongolian and possibly Korean and Japanese formed a single "Altaic" family, yet it has recently been thought possible (see e.g. Voivin 2009, Robbeets 2020) that these languages belong to four or five separate families, with their considerable similarities arising purely from contacts between speech-communities. Hübler seems to suggest (*op. cit.*, p. 159) that her findings not only suggest a tree-shape for the Altaic family, assuming it is one, but support the idea that it is indeed a single family descending from a common ancestor.

Even in a case as favourable as Indo-European, comparing languages and reconstructing their proto-language cannot get us back earlier than about 6000 BP (before the present), the probable age of the PIE language. (Human beings have certainly been speaking languages for far longer than six thousand years, and it might be that PIE and some other language-family or families shared a common ancestor earlier than that, but if so the surviving evidence does not establish this.) In another new biologically-inspired approach, Johanna Nichols sets out to look at the languages of the world on a much longer timescale (developing ideas pioneered by G.A. Klima, whose work is little known in the West because he wrote in Russian). Rather than trying to find deeper genetic relationships than those already known, Nichols borrows techniques from another area of biology, namely biogeography. Just as botanists, for instance, divide the land surface of the world into separate floristic regions with respect to the distribution and types of plant species (the standard division recognizes six top-level floristic "kingdoms", varying in size from the Holarctic kingdom, covering roughly everywhere north of the tropics, down to the small "South African" kingdom comprising just the southern tip of Africa), so Nichols argues that the world is divided into zones with respect to the distribution of languages; her approach pays little attention to questions about the languages' ancestry. There is a contrast between "spread zones" and "residual zones" (Nichols 1992: 13ff.): residual zones

(such as the Caucasus) contain high densities of languages and language-families (large numbers per million square miles), which tend to have long histories in the same area, while spread zones (such as the Eurasian steppe) have low language densities, and their relatively few languages have shallow time-depths, since over the millennia new languages often spread in and replaced those which were there previously. The contrast correlates with physical geography: areas of high language density tend to be relatively close to the Equator, and to be near coasts or to be mountainous (*op. cit.*: 232–4).

Nichols measures the linguistic diversity of different areas, in terms of how far their languages agree or differ with respect to various fundamental grammatical properties, for instance whether they mark the head/dependant relationship on the head or on the dependant. (In most grammatical constructions, one can identify one member as its head and the other(s) as dependent on it: in *white house* or *the woman's house*, *house* is head, because a white house is a kind of house, and *the woman's house* is a house rather than a woman.) In the "white house" construction, French marks the dependant: *blanche* in *maison blanche* has the feminine ending to agree with *maison*. In "the woman's house" construction, English marks the dependant by adding apostrophe-s, but Biblical Hebrew on the other hand marks the head: "house" is *bajit* in isolation but *bet* in *bet ha'išša*, "the woman's house". Many languages always mark dependants in various constructions, many others always mark heads, and some languages mark both, or are mixed (Hebrew marks the head of a genitive construction, but it inflects a dependent adjective to agree with its head noun, like French). Johanna Nichols is less interested in whether the languages of an area are head-markers or dependant-markers than in how diverse the languages of an area are, with respect to this and many other basic grammatical properties. So, looking at a sample of 174 of the world's languages, she finds for instance (Nichols 1992: 242) that the indigenous languages of eastern North America are rather consistently head-markers (only eight per cent are anything else), whereas among languages of New Guinea, while head-marking is the commonest single type it accounts for only 37 per cent of the sample languages.

Nichols (e.g. 1992: 258–9) claims that findings of this sort correlate with the prehistory of the peopling of the world, with grammatical clines still visible from west to east as Mankind spread out from its origin in Africa to occupy eventually all habitable territories. According to her the findings provide new answers to open questions, such as the date when humans first arrived in the Americas. It is

uncontroversial that the New World was first colonized from Siberia, there having at some periods been a land bridge across the Bering Strait and at other periods sea passages short enough for people to cross. The archaeological consensus had been that this first happened about 12,000 BP, though a later view (Adovasio et al. 1990) pushed this back to ca 14,000 BP. As Nichols (1990) interprets her linguistic data, both these dates are too low, with a more realistic figure being 20,000 BP as a bare minimum and 50,000 BP a possibility. "The linguistic evidence is not fully precise as to dates, but it is absolutely unambiguous in regard to ballparks: the New World has been inhabited for tens of millennia."

It is too soon to know whether Nichols's theories will prevail. The theories are complex and not all their elements are yet precise enough to evaluate; and the confidence she expressed in the passage just quoted was called into question by subsequent findings from genome data of the kind discussed on p. 000, which suggest a date for the earliest Americans that agrees better with the archaeologists' figures than with Nichols's (e.g. Lippold et al. 2014). However, archaeological findings emerging as I write (Bennett et al. 2021) appear to vindicate Nichols.

Certainly, with work of the kinds pioneered by Don Ringe and by Johanna Nichols continuing to appear, we can certainly say that the historical approach to linguistics is again a live enterprise, after decades of relative somnolence. The subject is unquestionably generating new ideas of interest to people other than narrow specialists.

REFERENCES (other than those listed in first edition)

Adovasio, J.M. et al. 1990. "The Meadowcroft rockshelter radiocarbon chronology 1975–1990". *American Antiquity* 55.348–54.

Anthony, D.W. and D. Ringe. 2015. "The Indo-European homeland from linguistic and archaeological perspectives". *Annual Review of Linguistics* 1.199–219.

Balaresque, P., et al. 2010. "A predominantly Neolithic origin for European paternal lineages". *PLoS Biology* 8(1), <www.plosbiology.org>.

Barbançon, F., et al. 2007. "An experimental study comparing linguistic phylogenetic reconstruction methods". University of California at Berkeley technical report, <statistics.berkeley.edu/tech-reports/732>.

Bennett, M.R. 2021. "Evidence of humans in North America during the Last Glacial

Maximum". *Science* vol. 373, issue 6562, 23 Sep 2021.

Black, P. 2017. "The comparative method in Australian linguistics". In Klein et al. (2017), pp. 129–37.

Blasi, D.E., et al. 2019. "Human sound systems are shaped by post-Neolithic changes in bite configuration". *Science* vol. 363, issue 6432.

Brosnahan, L.F. 1961. *The Sounds of Language*. Heffer (Cambridge).

Creanza, N., et al. 2015. "A comparison of worldwide phonemic and genetic variation in human populations". *Proceedings of the National Academy of Sciences of the USA* 112.1265–72, 3 Feb. 2015.

Dawkins, R. 2006. *The Selfish Gene*, 40th anniversary edn (first edn 1989). O.U.P.

van Driem, G. 2021. *Ethnolinguistic Prehistory: the peopling of the world from the perspective of language, genes and material culture.* Brill (Leiden).

Emeneau, M. 1956. "India as a linguistic area". *Language* 32.3–16.

Formigari, L. 2018. "Wilhelm Wundt and the Lautgesetze controversy". *History and Philosophy of the Language Sciences* 17 Jan 2018, <hiphilangsci.net/2018/01/17/wundt-lautgesetze/>.

Hickey, R., ed. 2017. *The Cambridge Handbook of Areal Linguistics*. C.U.P.

Honeybone, P. 2005. "Brugmann, Karl". In S. Chapman and P. Routledge, eds, *Key Thinkers in Linguistics and the Philosophy of Language*, Edinburgh University Press, pp. 45–8.

Hübler, N. 2020. "Typological profile of the Transeurasian languages from a quantitative perspective". In Robbeets and Savelyev (2020), pp. 145–59.

Iordan–Orr. 1970. Iorgu Iordan, *An Introduction to Romance Linguistics: its schools and scholars* (1932), revised and translated by J. Orr, with a supplement by R. Posner. Blackwell (Oxford), 1970.

Jespersen, O. 1941. *Efficiency in Language Change*. Munksgaard (Copenhagen).

Karlgren, B. 1920. "Le proto-chinois, langue flexionelle". *Journal Asiatique*, 11th series, 15.205–32.

King, R.D. 1967. "Functional load and sound change". *Language* 43.831–52.

Klein, J. et al., eds. 2017. *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1. De Gruyter Mouton (Berlin).

Knowles, G.O. 1973. "Scouse: the urban dialect of Liverpool". PhD thesis, Leeds University; <etheses.whiterose.ac.uk/2080/1/Knowles_GO_English_PhD_1973.pdf>.

Kretzschmar, W.A. 2002. "Dialectology and the history of the English language". In D. Minkova and R. Stockwell, eds, *Studies in the History of the English Language: a*

millennial perspective. Mouton de Gruyter (Berlin), pp. 79–108.

Labov, W. 2001. *Principles of Linguistic Change*, vol. 2: *Social Factors*. Wiley–Blackwell (Oxford).

Lepsius, C. 1860. "Über die Umschrift und Lautverhältnisse einiger hinterasiatischer Sprachen, namentlich der Chinesischen und der Tibetischen". *Physikalische Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin*, 1860, pp. 449–96.

Lippold, S., et al. 2014. "Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences". *Investigative Genetics* 2014, 5:13; <www.investigativegenetics.com/content/5/1/13>.

Lühr, R. 2016. "Indogermanistik und Germanistik". *Historische Sprachforschung* 129.233–63.

Metsupalu, M., et al. 2004. "Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans". *BMC Genetics* 5:26: <www.biomedcentral.com/1471-2156/5/26>.

Miestamo, M. et al., eds. 2008. *Language Complexity: typology, contact, change.* John Benjamins (Amsterdam).

Nakhleh, L., et al. 2005. "Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages". *Language* 41.382–420.

Nichols, J. 1990. "Linguistic diversity and the first settlement of the New World". *Language* 66.475–521.

O'Grady, G.N. 2004. "Pama-Nyungan again under unjustified attack!!". *Mother Tongue* 9.7–109, <www.aslip.org/MT/9 Mother Tongue IX (2004)_text.pdf>. @CHK with Paul Black@

Osthoff, H. and K. Brugmann. 1878. Preface to vol. 1 of the journal *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*.

Pulleyblank, E.G. 1982. Review of Wang 1977. *Journal of Chinese Linguistics* 10.392–416.

Ringe, D., et al. 2002. "Indo-European and computational cladistics". *Transactions of the Philological Society* 100.59–129.

Robbeets, M. 2020. "The classification of the Transeurasian languages". In Robbeets and Savelyev 2020, pp. 31–9.

Robbeets, M. and A. Savelyev, eds. 2020. *The Oxford Guide to the Transeurasian Languages*. O.U.P.

Robins, R.H. 1997. *A Short History of Linguistics*, 4th edn. Routledge. (First edn 1967.)
Sampson, G.R. 2017. *The Linguistics Delusion*. Equinox (Sheffield).
Sampson, G.R. 2019. "An unaddressed phonological contradiction". *International Journal of Chinese Linguistics* 6.221–37.
Sampson, G.R. 2021. Review of van Driem (2021). *Linguist List* post 32.3679, <linguistlist.org/issues/32/32-3679/>.
de Schlegel, A.W. 1818. *Observations sur la langue et la littérature provençales*. Librairie Grecque-Latine-Allemande (Paris).
Schlegel, F. 1808. *Über die Sprache und Weisheit der Indier: ein Beitrag zur Begründung der Alterthumskunde*. Mohr und Zimmer (Heidelberg).
Schmalstieg, W.R. 1988. "Congratulations to an important friend of Baltic studies: the one-hundredth volume of *Kuhns Zeitschrift*". *Lituanus* vol. 34 no. 4, <www.lituanus.org/1988/88_4_01.htm>.
Swiggers, P. 2017. "Intuition, exploration, and assertion of the Indo-European relationship". In Klein et al. (2017), pp. 138–70.
Thomason, S.G. 2000. "Linguistic areas and language history". In D. Gilbers et al., eds, *Languages in Contact*, Rodopi (Amsterdam), pp. 311–27.
Trubetzkoy, N.S. 1930. "Proposition 16. Über den Sprachbund". *Actes du premier congrès international des linguistes à la Haye, du 10–15 avril 1928*, pp. 17–18. A.W. Sijthoff (Leiden).@CHK@
Trudgill, P. 2011. *Sociolinguistic Typology: social determinants of linguistic complexity*. O.U.P.
Vajda, E. 2020. "Transeurasian as a continuum of diffusion". In Robbeets and Savelyev (2020), pp. 726–34.
Verner, K. 1876. "Eine Ausnahme der I. Lautverschiebung". *Kuhns Zeitschrift* 23.97–130.
Voivin, A. 2009. "Japanese, Korean, and other 'non-Altaic' languages". *Central Asiatic Journal* 53.71–132.
Vossen, R.. 2017. "The comparative method in African linguistics". In Klein et al. (2017), pp. 114–20.
Wang, W.S.-Y. 1969. "Competing changes as a cause of residue". *Language* 45.9–25.
Wang, W.S.-Y. and Chinfa Lien. 1993. "Bidirectional diffusion in sound change". In C. Jones, ed., *Historical Linguistics: problems and perspectives.* Longman.
Wang, W.S.-Y., ed. 1977. *The Lexicon in Phonological Change*. Mouton (the Hague).
Wechssler, E. 1900. *Giebt es Lautgesetze?* Max Niemeyer (Halle).

Zhang Menghan et al. 2019. "Reconciling the father tongue and mother tongue hypotheses in Indo-European populations". *National Science Review* 6.293–300.