Historical Linguistics
(from the website of Jay Jasanoff)

Historical linguistics, the study of language change, is the oldest subfield of modern linguistics. The success of historical linguistics in the nineteenth century was a major force behind the growth of synchronic linguistics in the twentieth. This page gives an overview of the classical theory of linguistic change and the comparative method, followed by a discussion of modern perspectives on language change and a note on the use of language as a tool in the study of prehistory. A few reading suggestions are provided at the end.

The classical theory

The first historical linguists noticed recurrent correspondences between the sounds of cognate words in the early Indo-European (IE) languages. They explained these by positing historical sound changes or “sound laws.” One of the first sound laws to be discovered was the Germanic consonant shift (“Grimm’s Law”), which converted earlier voiceless stops to voiceless fricatives (cf. Sanskrit trayas : English three), voiced stops to voiceless stops (Skt. dvau : Eng. two), and “voiced aspirates” to plain voiced stops in Germanic (Skt. bhṛātar- : Eng. brother). As more and more sound changes were studied, an important generalization emerged: if the storable, language-specific phonetic environment for a given sound change was satisfied, the change took place; otherwise it did not. The change of voiceless stops to voiceless fricatives in Germanic, for example, always applied word-initially and after vowels and sonants, but never after stops or fricatives (Skt. star- : Eng. strew, not **sthrew). The global claim that “sound change is regular,” or that “sound laws have no exceptions,” was first made by the German “Neogrammarian” (Junggrammatiker) school in the late 1870’s and has been accepted in some form ever since. It can be considered the fundamental theorem of historical linguistics.

The regularity principle is not falsified by the phenomenon of “analogy” — the type of change in which a form is altered under the influence of a related word or pattern elsewhere in the language. The English ordinal number sixth, for example, goes back to an ancestral form containing the cluster -kst- (compare the Latin cognate sextus), with a -t- that should not, according to the regular conditioning of Grimm’s Law (see above), have shifted to -th after the fricative -s-. But the -th of the present-day English word has nothing to do with any failure of Grimm’s Law to operate correctly. In fact, the Old English form was siexta, with -t-; the -th of sixth was introduced under the influence of the other ordinal numbers, where -th was phonologically regular (fourth, seventh, etc.).

Sound change and analogy, the latter typically invoked to repair morphophonemic irregularities induced by the former, were the distinctive analytic tools of classical historical linguistics. Syntactic and semantic change were also of interest to many scholars, but the power of the regularity principle gave sound change a fascination that no other aspect of the field could equal.
The comparative method

The statement that languages are related means that they represent changed forms of a single parent language or “protolanguage,” which may or may not be directly attested. The common parent of the Romance languages (French, Spanish, Italian, etc.), which could be called “Proto-Romance,” is one of the relatively few cases of a protolanguage that is well-documented; we usually call it Latin. On the other hand, the common ancestor of the Germanic languages (English, German, Swedish, etc.) was never recorded in writing; everything we know about Proto-Germanic must be recovered by inference from the surviving daughter languages. This is also true of Proto-Slavic (the common parent of Russian, Polish, Czech, etc.), Proto-Semitic (the common parent of Arabic, Hebrew, Aramaic, etc.), and hundreds of others. The technique by which we reconstruct the words and grammar of a protolanguage by projecting backwards from its daughters is called the “comparative method.” In the domain of phonology, where sound change is constrained by the regularity principle, comparative reconstruction can be as rigorous as solving an equation. The Greek, Sanskrit, and Latin words for “five” (pέnte, pāñca, and quinque, respectively), for example, allow us to specify the Proto-Indo-European (PIE) form uniquely as *pénkʷe. The initial consonant of the PIE form could only have been *p-, which can be shown from other words to have assimilated to a following -qu- in Latin. In the second syllable, *-kʷe is the only PIE sequence that would have yielded Gk. -te, Skt. -ca, and Lat. -que; of the other imaginable choices, PIE *-kwe would have given Skt. **-kva, PIE *-ke would have given Gk. **-ke / Lat. **-ce, and PIE *-te would have given Lat. **-te / Skt. **-ta. Careful and consistent use of this procedure affords a window on three millennia of the unrecorded prehistory of Greek, Sanskrit, and Latin.

Families of related languages, including Indo-European and its main branches, were discovered long before the principle of regularity of sound change. But informal inspection is not usually a reliable way to tell whether languages are related. The longer two languages have diverged, the harder it is to distinguish inherited lexical and grammatical features from accidental resemblances, borrowing effects, and typologically-driven convergences. To prove genetic relationship we must be able to point to correspondences that could only have come about through common descent. In inflected languages, these may be shared morphological irregularities, such as the peculiar paradigm of the verb “to be” in Latin (est ‘is’ : sunt ‘are’), Gothic (ist : sind), and Sanskrit (asti : santi). More usually, relationship is proved by finding systematic phonological correspondences attributable to regular sound change. The deepest securely identifiable families are c. 6000-8000 years old. PIE and Proto-Uralic (the ancestor of Finnish, Hungarian, etc.) are usually dated to around 4000 BCE; Proto-Afro-Asiatic, the parent of Proto-Semitic, Ancient Egyptian, and various African groups, is appreciably older. Other deep families include Austronesian and Sino-Tibetan in Asia, Niger-Congo in Africa, Ritwan-Algonkian (“Algic”) in North America, and Pama-Nyungan in Australia, among many others. The enterprise known as “long-range comparison,” which seeks to link families like these in yet larger groupings of immense antiquity (e.g., “Nostratic,” “Amerind”), is regarded as methodologically unsound by most practicing historical linguists.
Modern perspectives on language change

The advent of synchronic linguistics made it possible to understand sound change, analogy, and other kinds of linguistic change in a more general context. All observable linguistic change consists of an inception phase — the change proper — and a period of diffusion. In the commonest case, the initiating event is a juvenile learning error: a child pluralizes foot as foots, for example, or misparses an acoustic signal and wrongly fronts a vowel before *i. Such innovations are normally corrected before they can spread. Occasionally, however, they escape correction and become acceptable variants, potentially acquiring prestige value and being taken up by other speakers. Sociolinguistic studies have greatly clarified this phase of the change process. Claimed violations of the regularity principle, such as “lexical diffusion” — the supposed word-by-word progress of sound change through the lexicon over a period of decades or generations — have nothing to do with the essence of sound change itself, but reflect one aspect of the social mechanism by which all change is propagated.

For early historical linguists, who lacked a developed theory of underlying structure, all change was surface change, conditioned by surface facts. With the rise of generative grammar, language change came to be seen as grammar change, thus focusing attention on the possibility that some change might be controlled by non-surface linguistic factors — inefficiently exploited phonological contrasts, marked rule orderings (or constraint rankings), typologically inconsistent word order choices, etc. The sometimes overused tool of analogy invited particular scrutiny in this context. Critical discussion of analogy centered over whether the phenomena traditionally labeled “analogical” might be better explained in terms of rule loss, rule reordering, or other grammar-internal operations. Interestingly and perhaps somewhat surprisingly, the answer was (mostly) no; the status of surface-driven analogy as a primary mechanism of change was robustly reaffirmed. But the effort to put limits on analogy — to discover the conditions under which this type of change is likely to operate and in what direction — remains an important goal.

The tension between surface-driven and structure-driven explanations recurs in the most rapidly growing area of historical linguistics, historical syntax. Traditional accounts recognize three principal mechanisms of syntactic innovation: 1) analogy-like processes of structural reanalysis and extension; 2) functional “bleaching” of marked or expressive configurations into unmarked ones; and 3) as in all areas of language change, borrowing of foreign usages from non-native speakers. A recently discussed possibility, suggested by the “principles and parameters” framework in syntactic theory, is that a single comparatively minor diachronic event can lead to the resetting of a major syntactic parameter, producing a cascade of seemingly unrelated changes in its wake. Given the difficulties of using dialectally and stylistically diverse written records, however, the evidence for such chains of causation is hard to evaluate.

Language and prehistory

Linguistic evidence, properly interpreted, can be an important source of information about the past. Even if we knew nothing about European history, the number and range of French loanwords in English would suggest an event like the Norman conquest of
1066, which exposed England to Francophone domination for more than two centuries. Prolonged cultural contact can similarly be inferred from the borrowed Chinese words in Japanese, the Middle Iranian words in Armenian, and the Arabic words in Persian. Language is sometimes our only source of information about prehistoric migration and settlement patterns. Ancient and early medieval sources tell us very little about the early history of the Finns; yet the Finnish lexicon, with nested layers of Norse, Germanic, Baltic, Indo-Iranian, and Indo-European loanwords, is almost as explicit as an actual historical narrative. Genetic affinities, of course, are revealing as well. The Polynesian languages (Hawaiian, Tongan, Samoan, etc.) are a branch of the Southeast Asia-based Austronesian family; this rules out the possibility, once seriously entertained, that Polynesia was settled from South America. The Dravidian languages (Tamil, Telugu, etc.) are today largely confined to south and central India, but the presence of a Dravidian language (Brahui) in western Pakistan points to a wider original territory, lending support to the possibility that the early Indus Valley civilization may have been established by Dravidian speakers.

No less important than what linguistic evidence can do is what it *cannot* do. It cannot provide us with fixed dates or absolute chronologies. Language change does not unfold at a constant rate; this is why the quantitative technique known as glottochronology, which purports to compute the chronological distance between two related languages from the percentage of shared vocabulary they retain from their period of unity, is fatally flawed. And although linguistic evidence can lead us to set up temporally remote protolanguages, the translation of linguistic relationship into real-time history is a hazardous enterprise. The nineteenth and early twentieth-century scholars who created the myth of the “Aryans” committed every possible methodological error in leaping from Proto-Indo-European to the Proto-Indo-Europeans — the error of confusing language with “race”; of uncritically ascribing language spread to violent conquest; of attributing conquest to racial superiority; and of selectively interpreting the material evidence to locate the IE homeland where their prejudices led them to expect it. Current-day reimaginings of the past are usually more subtle. But the use of linguistic data to support prehistoric scenarios of conquest or ownership, often with an ethnic or national bias, remains surprisingly common. Linguistically literate readers should be prepared to correct for this practice when they encounter it.

*Further reading*

The basic concepts and controversies of historical linguistics are well surveyed in Brian D. Joseph and Richard D. Janda, eds., *The Handbook of Historical Linguistics* (Blackwell, 2003), with compendious bibliographical references. Two important books that have appeared since then are Lyle Campbell and William J. Poser, *Language Classification: History and Method* (Cambridge University Press, 2008), and Mark Hale, *Historical Linguistics: Theory and Method* (Blackwell, 2007). Campbell and Poser give a detailed glimpse into the content of the disputes over long-range comparison and other attempts to go beyond the comparative method. Hale presents a “neo-Neogrammarian” overview of historical linguistics that stresses the continuity between classical and more recent approaches.