

8 Evolutionary models in other disciplines

Introductory remarks by the chapter editor, Armin Hoenen

“Stemmatology usually works with texts that change during their copying history.” If we conduct a small experiment of metaphorically zooming out and replacing the nouns in this sentence with nouns from a higher, more general category, we could say: “Genealogical science usually works with sequences that change during their transmission.” Some sciences for which this statement is applicable – though not all of them – will be the focus of this chapter.

The formulation “sequences that change during transmission” hints at evolutionary theory, although the concept of evolution more specifically entails mutation and selection as agents of change, and therefore carries strong biological connotations. Nonetheless, it has been used to convey different notions of processes of change which lead to hierarchical or temporally successive structures in various disciplines; thus, we can speak of biological evolution, text evolution, language evolution, the evolution of writing materials, and so on. The main visual metaphor for such structures, and the only figure in Darwin’s *On the Origin of Species* (1859), is the tree. The tree as a mathematical, analytical structure has been used, in turn, for a huge number of purposes, be it in one of its first attested usages, as a family tree for aristocratic families (see Lima 2014, 29); as a *stemma codicum*; or as a way of displaying folder and file structures on a computer. As Lima (2011, 43) points out, the tree has been appreciated on the one hand and attacked on the other (and not only in stemmatology). Yet it has survived criticism and continues to be widely used.

So far, in this book we have looked at many kinds of stemmatic trees. In this chapter, we will focus on fellow trees from other disciplines, which together form the forest of “trees of history”, as O’Hara (1996) proposed to call some of them. The application of the tree model in science as an analytical tool is – as already stated – very broad and has had a special role as a “tool of thought” in Europe (Klapisch-Zuber 2007, 293). The habitat of our forest is indeed vast. In fact, it is so large that we will not be able to cover all the applications of trees (for which Lima 2014, among others, could be consulted); instead, we limit ourselves to some of the disciplines most intimately related to stemmatology: linguistics, cultural evolution, musicology, and biology. What are the parallels, what are the differences, what can we learn from each other, what can we borrow or incorporate, and what are the interfaces stemmatology shares with these sciences? These are some of the leading questions to keep in mind when reading this chapter.

Phylogenetics (8.1) has functioned as a donor of many computational tools (see 5.2, 5.4) to stemmatology. Linguistics (8.2) makes complex genealogical judgements just as stemmatology does, albeit with a focus on language as a whole, not on a single work. Anthropological phylomemetics (8.3; an umbrella term proposed by C. J. Howe and Windram 2011) is, on the one hand, concerned with trees of cultural artefacts (e.g. material relevant to codicology, book binding types) and, on the

other, with the analysis of textual evolution, for instance story patterns. Finally, music (8.4) has been transmitted as musical notation, often along with texts, and musicology is a discipline which has looked to stemmatology to develop methods and models. The similarities and interdependencies between these disciplines (or some of them) were noted at an early date. Famous scientists such as Darwin, Schleicher, or Haeckel were aware of our forest, as has been outlined in publications on the similarities of these disciplines, but also on their differences. Some of the most relevant ones for stemmatology may be Platnick and Cameron (1977), Cameron (1987), and O'Hara (1996). They specifically highlight differences, such as a special focus on the survivors at the tips/leaves of the tree (the living species) in biology, and on the (presumably lost) root of the tree in stemmatology, but also similarities, such as the consensus of using shared innovations in morphological classification in both early biological cladistics and stemmatology.

Christopher Howe and Heather Windram (8.1) explain phylogeny, where trees are used to display and systematise, for instance, the relationships between species, but also between proteins or individuals. Phylogenetic trees are today usually generated from molecular data such as DNA, the universal vehicle transmitting genetic information. Its universality entails that there can be a tree of life including all living beings (and the entire fossil record); compare the site tolweb.org. In palaeontology, however, molecular data is usually not available, which is why other paradigms have to be followed here (such as selecting certain traits of the fossil – not unlike selecting certain variants as genealogically informative), and this biological subdiscipline is presumably the closest biological relative of stemmatology. Additionally, in botany, hybridisation, or the mixing of DNA of different species, is common among certain plants as well as in bacteria, which is why accommodating large amounts of “contamination” may be of great interest to botanists, bacteriologists, and stemmatologists alike. Howe and Windram walk the reader step by step through the construction of a phylogenetic tree and outline the history of the field, rooting, contamination, and other relevant phenomena.

Just as for organisms in biology, some scholars have assumed one common origin of all languages (e.g. Bengtson and Ruhlen 1994), implying one giant language tree. In section 8.2, problems similar to those in stemmatology for the genealogy of languages are discussed. They have led scholars to existential criticism: language contact on all levels of a language (Thomason 2001) calls tree models into question. Pidgins (reduced, mixed languages used as a lingua franca) and creoles (pidgins which have become mother tongues) are vivid witnesses to a problematic entity when it comes to accommodating it in a tree of languages. Another issue is the choice of the base data for tree generation. What is the DNA of a language? A list of carefully chosen words considered to withstand borrowing; or some grammatical, syntactic, morphological, phonological features; or a weighted ensemble of all of them? While in textual criticism, the basis is the usually relatively clearly delimited single work, for linguistics, it is much harder to determine the basis of tree generation. An excit-

ing interface can be found in sound shifts that lead to trees for single lexemes similar to variant stemmata. Dieter Bachmann illustrates the answers and methods linguists have developed for these challenges. He focuses mainly on the history of Indo-European, the largest and presumably best-understood language family. He recalls the development of genealogical approaches in the field and outlines important currently discussed questions, such as understanding the time-depth of certain splits or the localisation of the *Urheimat* of the Indo-Europeans. He summarises the recent influx and perception of computation in the field and the mixed reactions it has received.

Section 8.3 has a broad scope. It deals with many phenomena to which a tree model can be applied from the sphere of anthropology and of human artefacts in general. Tomasello (1999) speaks of the “ratchet effect” when humans refine tools in subsequent generations (at some point in time, the hand axe became a true axe). This implies a model of evolution for anthropology where “descent with modification” applies. The basic data on which phylomemetics operates are (manually chosen) character states of artefacts such as cross-bows or of story patterns. This choice parallels the selection of significant errors in stemmatology and that of traits in cladistics and palaeontology. Jamshid Tehrani explains why, in this field, a certain class of algorithms has superseded prior approaches. His examples from the pre-literary transmission and admixture of story patterns across cultural borders may be especially relevant to stemmatology, which also analyses literary texts.

Section 8.4 provides a view of another related field which has inherited methodology from stemmatology proper: musicology. Reading text, we might say that we literally “hear” internally the words we read, but, reading musical notation, can we really “hear” the piece inside our heads? Even if we can, must the difference between experiencing a real musical performance and reading musical notation not necessarily be much larger than that between spoken and written text? Consequently, many more factors than the mere notation may interest the musicologist in determining the transmission history of a piece. Cristina Urchueguía introduces the reader to this field, maintaining a special focus on the transmission of notation and on stemmatological questions and methods.

When reading across these disciplines, their terminologies can be tricky, as each discipline has coined its own terms more or less independently: some of the terms or elements of models in one field may exactly map onto other terms in another field (e.g. “witness”, “language”, “taxon”), while others may not be exactly congruent (“contamination” vs “loaning”, and “calquing” vs “hybridisation” and “lateral gene transfer”).

8.1 Phylogenetics

Heather Windram and Christopher Howe

Darwin and Lamarck both saw species as evolving from their ancestors by a process of descent with modification, according to which mutations in a parent organism are inherited by their offspring. The “history of [...] the evolutionary development of an organism or groups of organisms” by descent with modification is referred to as phylogenetics (Allen 1990, 897). Phylogenetic studies often attempt to infer the relationships by descent between groups of organisms, such as which were descended from common ancestors. Today, we typically use data that are directly genetically determined (DNA or protein sequences) for this (see e.g. www.ebi.ac.uk/training/online/course/introduction-phylogenetics/what-phylogenetics), but the use of genetic data does not necessarily follow from the definition, and the term “phylogenetic(s)” predates the popularisation of the term “gene” by the influential early geneticist Bateson in the early twentieth century. The use of computers to infer trees showing phylogenetic relationships from genetic data (which we would recognise today as phylogenetics, or phylogenetic inference) really began in the early 1960s with work by Edwards and Cavalli-Sforza to infer the evolutionary relationship among different human populations based on data from blood groups (reviewed by Edwards 2009), which are, of course, genetically determined.

8.1.1 History

8.1.1.1 Development of phylogenetic and related methods

Edwards and Cavalli-Sforza developed three methods for phylogenetic inference (Edwards 2009), which have remained fundamental to the field. These were (i) the method of least-squares estimation on an additive tree (in effect, the distance matrix method); (ii) the method of minimum evolution (in effect, maximum parsimony); and (iii) the method of maximum likelihood. The methods are discussed in more detail below. A number of other scientists developed approaches to classification and/or phylogenetic inference around the same time, with varying degrees of similarity to those of Edwards and Cavalli-Sforza. The German entomologist Willi Hennig developed the approach he referred to as “phylogenetic systematics” (the title of his classic book in its translation into English, 1966), which later became known as “cladistics”. This grouped organisms based on their sharing derived characters (synapomorphies) that others did not have, as a result of their sharing a common ancestor. Groups would therefore reflect evolutionary relationships. The use of characters shared by one or more taxa to the exclusion of others is, in effect, a maximum parsimony criterion (see below). This use of synapomorphies in preference to other characters differentiated cladistics (and maximum parsimony) from the “numerical taxonomy” advanced by Sneath and Sokal (1962). The latter method

classified organisms based on a number of characters that were all weighted equally rather than used selectively, with the organisms grouped in a cluster analysis. It was described as a “phenetic” classification, and Sneath and Sokal (1962, 856) noted that it “practises the strict separation of phylogenetic speculation from taxonomic procedure”. Nevertheless, Camin and Sokal (1965) employed the data matrices used for numerical taxonomy for a parsimony-based process of phylogenetic inference.

8.1.1.2 Application of molecular sequence data

Probably the biggest driver for the application of methods for phylogenetic inference in biology was the development of techniques for the rapid determination of DNA sequence data in the 1970s. DNA is composed of chains of units called nucleotides. There are four forms of these, referred to as adenosine (A), cytidine (C), guanosine (G), and thymidine (T) nucleotides. The order in which particular nucleotides come in a chain comprises the information carried by DNA. This information is used by the cell to build proteins, the molecules responsible for many of the functions of cells. Just as DNA molecules are chains of nucleotides, proteins are chains of units called amino acids. There are twenty different amino acids that occur in proteins (although in some instances individual amino acids may be chemically modified). The order of individual amino acids is determined by the order of nucleotides in the DNA. The region of DNA containing the information for the order of amino acids in a given protein is termed the “gene” for that protein. (The order of amino acids in a protein determines the shape of the protein and its function in an organism.) As there are twenty amino acids to be specified, but only four varieties of nucleotide to specify them, organisms use groups of three nucleotides to specify each amino acid. These groups are referred to as codons. Some amino acids can be specified by many different codons; others are specified by a single codon.

In 1977, two landmark papers were published that described simple methods for determining the sequence of nucleotides in defined pieces of DNA – referred to as DNA sequencing. These methods were developed by Maxam and Gilbert (1977) at Harvard, and Sanger, Nicklen, and Coulson (1977) at Cambridge, UK. The Sanger method became the more widely adopted, although it in turn has been largely superseded by so-called “high-throughput” methods capable of generating even larger amounts of data. These methods, and especially the Sanger method, led to an explosion in DNA sequencing (and protein sequencing by inference from DNA sequences), and the resulting data became widely used for phylogenetic analysis. Phylogenetic inference is used throughout biology. It can be used to study evolutionary events ranging from the very ancient, such as the origin of the main groups of organisms (Bacteria, Archaea, and Eukaryota) billions of years ago, to the very recent, such as the seasonal origin of different strains of influenza (T. A. Williams et al. 2013; Nelson et al. 2007).

8.1.1.3 Sequence-based phylogenetic methods

The principle underlying phylogenetic analysis based on sequence data is that, as cells in an organism divide, their DNA is first copied (replicated) so that each of the resulting cells inherits a full set of DNA from the parental cell. Errors may occur during the copying process. These are referred to as mutations, and each time the DNA carrying the mutation is copied the mutation is propagated. If a mutation occurs in cells that will ultimately give rise to egg or sperm cells, the mutation will be passed to the next generation. When species or other taxonomic groups of organisms (which we refer to here as taxa, but are sometimes also referred to as operational taxonomic units, or OTUs) become separated during evolution, different mutations occur in different taxa. In general, the more closely related two taxa are (i.e. the more recently they shared a common ancestor), the more similar their DNA sequences will be. Methods for phylogenetic tree building exploit these differences. Some methods assume that there is a constant rate of mutation over time, which is referred to as a molecular clock. However, not all methods require this. For example, cladistic methods, which assume that organisms sharing a particular mutation to the exclusion of others share common ancestry, do not need to presuppose a molecular clock.

8.1.1.4 Application of phylogenetic methods to textual scholarship

The process of copying of DNA with the incorporation of changes clearly shows many attractive similarities to the copying of manuscripts by scribes. Platnick and Cameron (1977) commented on the similarities between stemmatology and phylogenetic reconstruction – although Griffith (1969) had previously applied principles from numerical taxonomy to a range of classical and biblical texts. Platnick and Cameron noted that textual criticism (and linguistics) resembled “phylogenetic systematics in being primarily concerned with constructing and testing hypotheses about the interrelationships of taxa connected by ancestor-descendant sequences” (1977, 380). They pointed out that both fields of study used similar data (“The cladist need only substitute ‘taxa’ for ‘witnesses,’ ‘derived character’ for ‘error’”; 381) – although they argued that chronological data might be available and applicable in the field of textual criticism, but not in cladistics. Lee (1989) applied computer programs from phylogenetic analysis to St Augustine’s *Quaestiones in Heptateuchum*, and Peter Robinson and Robert O’Hara (1996) described the application of phylogenetic programs to a Norse narrative, discussing in detail the similarities between textual analysis and cladistics (see 5.1.2.1). The phylogenetic approach received much attention with the publication of an analysis of the prologue to Chaucer’s *The Wife of Bath’s Tale* (Barbrook et al. 1998), and a number of authors have commented on the similarities of the fields and the applicability of different methods (e.g. C. J. Howe et al. 2001; Macé and Baret 2006). Since these initial studies, phylogenetic methods have been applied to a range of textual traditions (reviewed by C. J. Howe and Windram 2011), including a set of music manuscripts (Windram, Charlston, and

Howe 2014). Adopting the term “meme” used by Dawkins as a “unit of cultural transmission”, the term “phylo memetics” has been proposed to refer to the phylogenetic analysis of non-biological data (Dawkins 1976, 206; see C. J. Howe and Windram 2011). This includes not only texts but also data from other disciplines, such as languages and folk tales (see 8.3).

8.1.2 Methods

It is important to recognise that phylogenetic methods depend on a model of how the data is evolving. If the data has not really been generated in accordance with that model (model misspecification), the phylogenetic tree generated may be erroneous. The phylogenetic methods commonly applied to textual traditions represent a subset of those usually used with sequence data. We focus here on the methods most widely used with texts. We give a brief summary here, although more detailed descriptions are available elsewhere (see e.g. Lemey, Salemi, and Vandamme 2009; 5.3 above). Many of these have been tested for their accuracy in stemmatology, alongside other programs developed specifically for textual analysis (see e.g. Roos and Heikkilä 2009; 5.3.7 above).

The maximum parsimony method aims to find a tree structure (topology) that requires the smallest number of character changes to give rise to the DNA or protein sequences seen at the ends of the branches of the tree (referred to as terminal nodes). The total number of possible trees increases more than exponentially as the number of taxa increases (Felsenstein 1978b), so computer algorithms employ a variety of methods to reduce the number of trees that have to be considered, for example excluding further consideration of partially constructed trees that already require more character changes than the best trees (most parsimonious) recovered so far. Distance matrix methods aim to identify the tree that gives the best fit (in terms of numbers of changes along the branches of the tree) to a matrix of distances between the DNA or protein sequences of the taxa. Neighbour-joining builds up the tree in stages, and is therefore referred to as an agglomerative clustering method. NeighborNet and the related split decomposition are also distance matrix methods (more details in 5.3.3).

A family of probability-based methods are often used with sequence data. Maximum likelihood methods calculate the probability of each tree giving the observed data under a specific evolutionary model. The tree with the highest probability is the preferred solution. These methods are computationally demanding. Bayesian methods in effect start with a standard set of prior assumptions, and then consider the effects of randomly changing those parameters. If the new set of parameters is worse than before (in terms of the probability of generating the observed data), they are rejected.

When phylogenetic methods are used to determine the copying history of a textual tradition, the DNA or protein sequences are replaced by the words in texts.

This requires the conversion of the texts into datasets resembling DNA or protein sequences, and in a format that the phylogenetic programs can use. This conversion has been described elsewhere (e.g. C. J. Howe and Windram 2011). Essentially, each text is converted into a string of characters. At any one position in the string, a character may show different states in different texts (e.g. changes of a word, or mutations in the case of DNA sequences; see 3.3.1), with any one position in the string corresponding to the same feature of the text in different witnesses. Once the texts have been converted into a dataset analogous to biological sequence datasets, the phylogenetic programs can be used essentially unchanged to analyse them.

The methods most widely used for analysis of textual traditions are arguably maximum parsimony, and distance matrix methods such as neighbour-joining and NeighborNet. Maximum likelihood methods are not widely used with texts because of the difficulty of formulating a plausible underlying evolutionary model. Bayesian methods have been used with non-biological data (e.g. the analysis of the *Little Red Riding Hood* folk tale in Tehrani 2013; see 8.3 below). It will be interesting to look at the application of these methods to texts in more detail (for more on Bayesian methods, see 5.3.6).

8.1.3 Tree structure

8.1.3.1 Topology

In their simplest format, the programs generate a tree with each of the texts at the end of a branch in the tree. Trees can be considered equivalent if they can be interconverted without the need to break and rejoin any of the branches. They are said to have the same *topology*. The distance along the branches between any two texts is a measure of the amount of difference between them. However, it is not always easy to interpret these distances. A large distance might indicate that multiple rounds of copying separate two witnesses, or that there were a few rounds of inaccurate copying.

8.1.3.2 Multifurcation

The trees generated by software are usually bifurcating; that is to say, internal branches (often referred to as edges) divide into two. That is not surprising, as it is typically an inevitable consequence of the algorithm used, making the tree reconstruction more tractable. Some methods allow for trees to be represented as multifurcating, that is, with a particular branch giving rise to several others. A multifurcation, or *polytomy*, might be an appropriate representation when a taxon radiates simultaneously into several others or when the method is unable to resolve the branching order (fig. 8.1-1).

A polytomy that arises from radiation of a taxon into several others is often referred to as a *hard polytomy*. One that arises when the branching order cannot

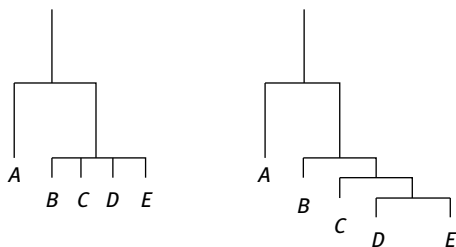


Fig. 8.1-1: A polytomy (left) can arise when a taxon simultaneously diverges into several others, or when a method is unable to resolve the branching order of a number of closely related taxa that are in principle resolvable (right).

be resolved is similarly a soft polytomy. The biological relevance of hard polytomies has been questioned (Lemey, Saelmi, and Vandamme 2009), but a hard polytomy could exist in the case of divergence of viral strains, for example where infection of an individual cell or organism could generate multiple progeny viruses by DNA replication using the same template. In the case of texts, it is clearly possible for the same text to be copied multiple times, in which case a hard polytomy would be appropriate. The question of how to allow for polytomies in textual traditions has been discussed by Phillips-Rodriguez, Howe, and Windram (2010).

8.1.3.3 Networks

Where evolutionary relationships among taxa are conflicting – e.g. if some characters in taxon *A* indicate a most recent common ancestry with *B*, whereas others indicate a most recent common ancestry with *C* – it is not possible to depict them all in a two-dimensional branching tree. In a biological context, this might arise if a taxon was derived by hybridisation between two others, or by transfer of a limited number of genes between separate taxa. In the context of textual analysis, this would occur as a result of contamination. In these cases, depicting relationships as a network may be appropriate. This is discussed in more detail below.

8.1.3.4 The root and outgroups

Many tree-building programs do not initially indicate the earliest part of the tree, referred to as the root. Care needs to be taken to avoid making an unjustified assumption about where the root lies, as shown in figure 8.1-2.

Outgroups (see 5.2.1) are used in order to allow the rooting of a phylogenetic tree in biological analysis. The outgroup is a taxon, or group of taxa, that is evolutionarily distinct from the group of species (or other taxa) that are being studied. The latter are termed the ingroup and have a close evolutionary relatedness. The outgroup must be evolutionarily related to the ingroup, but must be known to have diverged from the ingroup before the members of the ingroup diverged from one

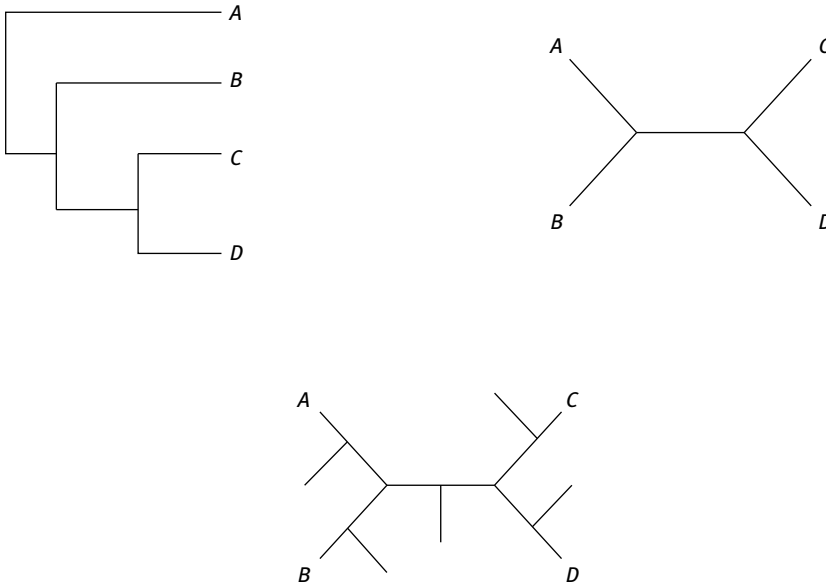


Fig. 8.1-2: Rooting a tree. Although the tree on the top left might be taken to imply that A was the earliest-diverged taxon, the tree is topologically identical to that on the top right, which makes no implication as to which taxon diverged earliest. This could be resolved by locating the root of the tree, which could go in any of five places, marked in the lower panel.

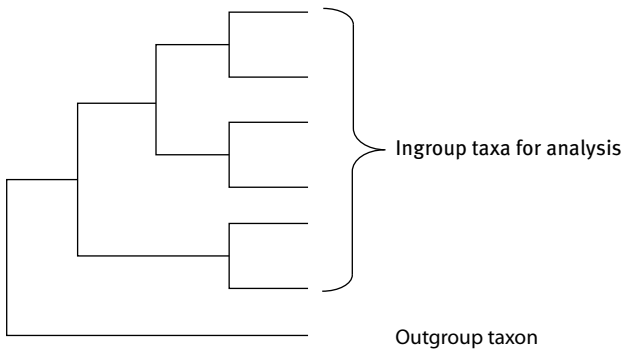


Fig. 8.1-3: Relationship of an outgroup to the ingroup.

another. Nevertheless, it must be sufficiently related that its molecular sequences (either DNA or amino acid) can be aligned unambiguously with those of the ingroup. The resultant phylogenetic tree can be rooted by the outgroup, as this is already known to represent the earliest branch on the tree (fig. 8.1-3).

It is generally not possible to use outgroups for the study of textual traditions. All witnesses to a particular text would be included in the ingroup, and there is no suitably related outgroup to include in the analysis. Other texts by the same author,

or even other sections of the text under analysis, would clearly not align with the text being studied. However, in certain circumstances – e.g. if the identity of the earliest source is known from external (non-textual) evidence – it may be possible to root a textual phylogenetic tree (for an example, see 4.5.3). Rooting a tree is, in effect, indicating the direction of change along the edges within the tree. It may also be possible to do this based on an understanding of the changes involved (i.e. if change X to Y is much more likely than Y to X).

Some methods root the tree by finding the longest path separating two taxa on the tree, and placing the root half-way along that branch. This is known as *mid-point rooting*. This method is valid if the rate of change across the tree is constant; for sequence data, that means they are evolving according to a molecular clock (see above). This is not generally applicable to texts where the rate of change is not constant and depends on the frequency of copying, the exemplar used for making a new copy, and the number of alterations introduced into the text in a round of copying.

8.1.3.5 Reliability

A computer program for phylogenetic inference will always produce a tree. However, that does not mean that the tree is reliable. Many scholars find it helpful to use a range of different methods for phylogenetic reconstruction, and place more reliance on groupings that are consistently recovered with different methods. However, it should be realised that this is not infallible!

The technique of *bootstrapping* is often used to estimate the statistical reliability of particular groups. The method takes the original dataset, notes the number of sites in it, and picks the same number of sites randomly from within the dataset. That is to say, for a dataset of 200 sites, 200 sites would be picked at random from it, and the character states recorded; some sites would not be sampled, and others would be sampled once or more than once. This is described as *sampling with replacement*, and generates a subsidiary dataset. The process is repeated using the original dataset to generate a large number of subsidiary datasets (perhaps 100 or 1,000). They are all different from each other, but also all derived from the original. The chosen phylogenetic inference method is then applied to each of the datasets. Bootstrapping then considers the original tree and asks in what percentage of the trees derived from the subsidiary datasets a particular grouping from the original tree is seen. These percentages are then displayed on the original tree, adjacent to the nodes that define the groups. In effect, this method gives an indication of whether a particular grouping is independent (thus with a high bootstrap value) of the particular sites used to calculate the tree. Using each dataset to calculate a tree takes the same amount of time as calculating the initial tree, so for large datasets or computationally demanding methods, bootstrapping may take a large amount of time. Bayesian analysis automatically generates values indicating support for particular groupings shown in the tree, so bootstrapping is unnecessary.

It should be noted that these methods provide an estimate of how robust groupings are to the consequences of stochastic effects (e.g. whether a particular feature was included in the data). They do not provide a protection against systematic errors, for example in cases of model misspecification (see above).

8.1.3.6 Extant ancestors

Typical interpretations of phylogenetic trees with biological data assume that ancestors of extant species (which would be placed on internal nodes) no longer exist, and place all taxa at the ends (*terminal nodes*) of edges in the tree. Clearly, the assumption that ancestors no longer exist need not be valid for trees based on texts. Experiments with “artificial” manuscript traditions (i.e. where the copying history is known a priori) indicate that these extant ancestors are placed at the terminal nodes of very short branches emerging from their “true” position in the tree (Spencer, Davidson, et al. 2004, 507). This needs to be kept in mind when interpreting outputs.

8.1.3.7 Horizontal gene transfer and recombination

The construction of phylogenetic trees representing evolutionary relationships is underpinned by the assumption that genetic information passes down through the generations, with accumulated mutations allowing the evolution of new species. However, this vertical transmission of information is not the sole means of inheritance. It is estimated that some 20–30 % of the genome of bacteria is acquired by horizontal gene transfer (HGT), also known as lateral gene transfer (LGT). This process can be mediated by transmissible DNA molecules (plasmids, responsible, for example, for the spread of antibiotic resistance between bacterial strains), viruses, or the acquisition of naked DNA from the environment. The result of this lateral transfer is a molecule that has two or more different ancestors, rather than the single ancestor implied by a tree-like propagation. Other processes may generate products with two ancestors, including hybridisation between different taxa (which is especially significant in plant evolution), or recombination between individuals within the same taxon. Depending on the details of the process, the resulting molecule may show a single distinct break-point, with material from different ancestors on either side, or may be a patchwork (fig. 8.1-4). Recombination increases the amount of variation generated as a result of sexual reproduction.

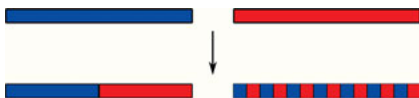


Fig. 8.1-4: Consequences of horizontal gene transfer, hybridisation, or recombination. The parental DNA sequences (red and blue) may undergo one or more recombination events to generate molecules that are composed either of two sections or of a patchwork.

These processes in biological systems are broadly analogous to contamination (see 4.4) in a textual tradition where a scribe used multiple exemplars, switching from one exemplar to another either at a single location (successive contamination) or at multiple locations throughout a text (simultaneous contamination). Computational methods have been developed in the biological sciences that allow the representation of multiple affiliations between species. These methods show the relationships between taxa as networks rather than as simple branching trees. One such method is NeighborNet (D. Bryant and Moulton 2004).

The most likely location of a recombination event between two DNA strands (or texts) can be determined by statistical methods such as maximum chi-squared analysis (Maynard Smith 1992). Although this is not strictly a phylogenetic method, we nevertheless discuss it here. In this method, a putative break- (or hybridisation) point is moved stepwise along the two sequences selected for comparison, and at each point the number of differences between the two sequences on either side of the putative break-point is determined and compared with the number of differences that would be expected if the two sequences were equally related along their lengths. The chi-squared value for the deviation from the expectation is then calculated. If there is an instance of recombination or change of exemplar, it can be expected that the sequences will become more (or less) closely related after the break-point, and the site at which there is the greatest discrepancy between the observed and expected number of differences (and thus the maximum chi-squared value) is indicated as the most likely break-point. The chi-squared values at each location may be plotted, in which case the peak of the graph represents the most likely break-point. A statistical measure can be used to determine if the chi-squared value is significant, taking into account the number of taxa in the dataset, with a significant result indicating that the distribution of differences arose by recombination or exemplar change rather than simply by chance variation, at the specified level of statistical significance. Chi-squared analysis is not directly applicable where there is a mosaic of recombinant DNA or where there is more than one exemplar used throughout the generation of a given text, with multiple break-points throughout the sequence. In the context of textual analysis, the maximum chi-squared method was first applied to the prologue to *The Wife of Bath's Tale*, and gave results that were consistent with those obtained by conventional scholarship (Windram, Howe, and Spencer 2005, 202).

8.1.4 Ancestral state reconstruction

The aim of phylogenetic inference using sequence data is frequently to understand more of the evolutionary history of the taxa under consideration, rather than to reconstruct the DNA or protein sequence(s) of the ancestor of all the taxa. In some circumstances, however, reconstructing ancestral states may be useful. For exam-

ple, it has been suggested that vaccines based on ancestors of present-day viral strains may be more useful than vaccines based on specific strains. A number of programs are available for ancestral state reconstruction using sequence data. They are typically based on principles of maximum parsimony, maximum likelihood, and Bayesian analysis, that is, methods dealing with individual character states rather than distance matrices. These methods have not been widely used for ancestral state reconstruction with textual data, although some examples have been described (e.g. Hoenen 2015b; 5.4.6 above). Instead, scholars have typically focused on consideration of variant readings at particular sites in a text.

8.1.5 Appropriateness of application of phylogenetic methods

Some scholars have expressed reservations over the application of phylogenetic methods to textual analysis. These reservations have been discussed elsewhere (Howe, Connolly, and Windram 2012) including in this book (see 5.5), so we will not consider them in detail here. Some of them are due to misunderstandings of the methods. Some of them are well founded, but usually they are equally applicable to conventional textual analysis as well – for example, the problem of dealing with contamination in traditions or with convergent changes (the same change occurring independently in different sequences or witnesses). It should be remembered, though, that when carrying out phylogenetic analysis in biology it is important to assess whether the methods and (often implicit) assumptions about how the data have evolved are correct. Otherwise, an incorrect phylogenetic tree may be inferred, and may even have a high statistical robustness. Similarly, analysis of texts should not be an exercise in using computer-based methods blindly to produce the single “correct” tree of relationships. Rather, it should be seen as a useful tool that allows textual scholars to focus attention most usefully on particular groups of witnesses or particular sections of a text. The process of encoding witnesses for phylogenetic analysis can be time-consuming. However, once that has been done, the ability to answer “what if” questions (What if we just look at this set of witnesses? What if we compare the copying history of this chapter with the following chapter?) may be very helpful indeed. Throughout, though, it is essential for textual scholars to use their experience in interpreting the results of phylogenetic analysis.

Further reading

Lemey, Salemi, and Vandamme (2009). Online resource: www.ebi.ac.uk/training/online/course/introduction-phylogenetics/what-phylogenetics.

8.2 Linguistics

Dieter Bachmann

This section discusses the history of the genetic, or phylogenetic, perspective in the study of the history of languages. An overview of the use of computational phylogenetic methods in the field of historical linguistics since the early 2000s is given.

8.2.1 Origins of historical linguistics

In the history of ideas, the discovery of language change is surprisingly recent. It may even be argued that the notion is an entirely modern one, made explicit only in scholarship of the seventeenth to eighteenth centuries. In the grammatical traditions of Antiquity, both in the Graeco-Roman West and in India, there certainly was an awareness of differences between languages and registers, such as Pāṇini's recognition of grammatical rules that apply only *chandasi* (i.e. in the Vedic hymns), but this does not amount to the hypothesis that such differences are due to a process of historical evolution. Such recognition of linguistic change as we find in the Greek classics (e.g. Plato *Cratylus* 432; Aristotle *Poetics* 1457) is concerned with the introduction of error, without the suggestion that the accumulation of such errors over sufficiently long timespans may result in entirely new languages. As noted by Gippert (forthcoming, citing Lentz 1870, 791), some grammarians of the first century BC do, however, seem to assume that Latin was derived from Greek; according to the testimony of Herodianus (second century AD), the grammarian Philoxenus notes the lack of a dual among Aeolians and among Romans, arguing that the Romans are descended from the Aeolians.

Indeed, the main reference to the idea of language change in ancient literature is Genesis 11, the confusion of tongues at Babel, which is also immediately paired with the idea of migration in the form of the “scattering over the face of the whole earth” of the speakers of the individual languages now “confused”. The Babel story provides an explanation for language change implicit in the assumption that (i) all nations are descended from one common ancestor (the biblical Noah) and that (ii) each nation is characterised by its own language. Early Christian literature specified the number of nations as seventy-two (based on the number of grandsons of Noah; cf. Augustine *De civitate Dei* 16.3), and in mediaeval literature this occasionally appears as the number of languages created in the confusion of tongues (e.g. in the Irish *Auraicept na n-Éces*, which holds that the Irish language was created by combining what was best in each of the seventy-two languages).

The study of language change can thus be considered a strictly modern, and strictly Western, innovation. Observations regarding the similarity of words in different languages were made as early as the sixteenth century; notably, Filippo Sassetti, a Florentine traveller to India, noted the striking similarity of certain words in Sanskrit and Italian in a letter dated 1585. More systematic comparisons postulating

the derivation of modern languages from a common source were published in the seventeenth century, notably by Marcus Zuerius Boxhornius (1647), without receiving widespread recognition (Muller 1986). The first exposition of the idea of a genealogical relationship of the world's languages paralleling the history of the world's peoples is probably a short essay by Leibniz published in 1710 (entitled “*Brevis designatio meditationum de originibus gentium, ductis potissimum ex indicio linguarum*”). Leibniz here aimed at what we would today call Proto-World, surveying most known languages of Europe, Asia, and Africa, and suggesting that the linguistic changes accumulating over time might be so extreme as to render the relationship unrecognisable. William Jones, in 1786, noted the similarity of Sanskrit to Greek and Latin, postulating their derivation from “a common source, which, perhaps, no longer exists” (quoted in e.g. Meier-Brügger 2003, 173–174).

8.2.2 Comparative Indo-European linguistics

The systematic comparative study of Indo-European languages developed in the early decades of the nineteenth century, as documented in the *Vergleichende Grammatik* of Franz Bopp (1833–1852). The first graphical representation of a family tree (*Stammbaum*) of Indo-European language families was published by August Schleicher (fig. 8.2-1).

The choice of representation by Schleicher was directly inspired by Darwin's *Origin of Species*, published only two years earlier in 1859. Schleicher would go on to make this explicit in a communication entitled “*Die Darwinsche Theorie und die Sprachwissenschaft*” to his friend, the evolutionary biologist Ernst Haeckel, in 1863.

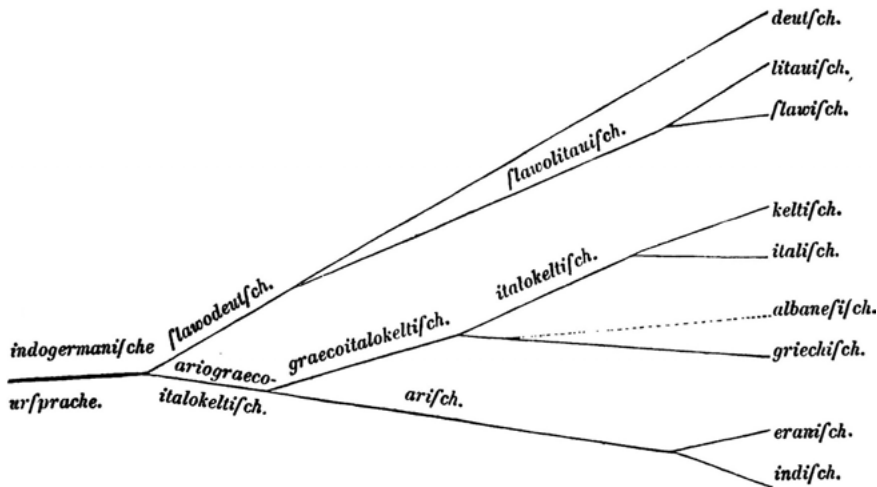


Fig. 8.2-1: Schleicher's genealogical tree. Source: Schleicher (1861, 1:7).

Schleicher's 1861 phylogenetic diagram depicts the genealogy of nine contemporary Indo-European language families with eight binary nodes (Armenian was recognised as an independent branch later, by Heinrich Hübschmann in 1875; the two eccentric branches of Tocharian and Anatolian were discovered still later, in 1908 and 1915 respectively). Branch length in Schleicher's diagram indicates relative time-depth: of his eight intermediate nodes, only the two most recent ones remain generally accepted, Balto-Slavic ("slawolitauisch") and Indo-Iranian ("arisch"); the others, including Italo-Celtic and Germano-Balto-Slavic ("slawodeutsch"), find at best very limited support today.

Schleicher postulated the phylogenetic relationship of Indo-European languages forcefully, to the point of being considered the originator of the *Stammbaumtheorie* (the stemmatological approach to relationships between languages). But Darwin (1859, 422) himself had already adduced the phylogeny of languages as an example. Like Leibniz (1710), he identifies the genealogy of languages and that of mankind as going back to Proto-World, saying that "a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world" and that "the various degrees of difference in the languages from the same stock, would have to be expressed by groups subordinate to groups; but the proper or even only possible arrangement would still be genealogical".

Nevertheless, as had been clear to historical linguists since before Schleicher's publication, the comparative evidence did not naturally result in a picture that favoured a genealogical or tree-like (phylogenetic) representation. Evidence considered in comparative linguistics is of at least three types: phonology (sound laws), lexicon (cognate vocabulary), and morphology (word formation), and, where possible, syntax. The study of these aspects in isolation will immediately result in relationships that are mutually exclusive in naive phylogenetic terms. An important example from the field of phonology is the centum-satem division, which bisects the Indo-European phylum (with the exception of the marginal branches of Anatolian and Tocharian, and possibly of Albanian and Armenian). At first glance, this is a major division of the Indo-European family of languages, neatly producing an eastern and a western group, Balto-Slavic and Indo-Iranian in the east (satem), and Greek, Italic, Celtic, and Germanic in the west (centum). But nothing else about the languages thus grouped necessarily suggests any closer relationship within each group; indeed, Balto-Slavic and Indo-Iranian end up at opposite ends of Schleicher's diagram. Schleicher's fashionable "linguistic Darwinism" was therefore harshly criticised from the outset. Its main opponent was Johannes Schmidt, who, in the assessment of Delbrück (1919, 118), had "done away for good with [endgültig beseitigt] Schleicher's theory" in 1872. Schmidt's *Wellentheorie* replaced the phylogenetic model of language change with the notion of waves of innovation which could travel across linguistic boundaries. It goes without saying that Schmidt's work by no means spelled the final end of attempts to describe the Indo-European family in phylogenetic

ic terms, but the recognition of ubiquitous waves of linguistic innovation (in stemmatological terms, contamination; see 4.4) in the late nineteenth century had the beneficial effect of liberating historical linguistics from worrying over phylogenetic paradoxes, and instead allowed it to be led where the data pointed. Rather than by a phylogenetic structure, linguistic relationships in general are more naturally represented by means of a map of isoglosses. The more crossing of isoglosses is present, the less the situation will be amenable to a phylogenetic representation. Anttila (1989, 305) provides an exemplary isogloss map of the Indo-European phylum, including twenty-four isoglosses in which the centum–satem isogloss crosses eight others.

With the discovery of the Anatolian and Tocharian branches in the early twentieth century, the phylogeny of Indo-European began to look a little more articulated, or tree-like. The hypothesis that the Anatolian branch can serve as an archaic outgroup for the remaining phylum has become known under the somewhat misleading name of the “Indo-Hittite hypothesis”, coined in the 1930s. This proposal long remained controversial, and was contrasted with the *Schwund-Hypothese*, which claimed that the absence from Anatolian of many features which appeared to be common elsewhere in the Indo-European phylum was due to loss rather than representing the archaic stage of the language family. While the question has not been resolved completely, progress in the reconstruction of Anatolian in the 1990s seems to bear out the early separation of Anatolian, albeit perhaps not as far removed from the breakup of non-Anatolian Indo-European as originally envisaged in the Indo-Hittite hypothesis (Melchert 1998, forthcoming). It is also commonly assumed that the Tocharian branch was the next to diverge from Common Indo-European after Anatolian (Watkins 2001). By contrast, a phylogenetic tree of Indo-European published by Hamp (1990), based on a combination of morphology, phonology, and lexicon, placed Tocharian under “Northwest-IE” alongside a number of poorly attested Palaeo-Balkan languages. If we wanted to represent the consensus in the field as to the phylogenetics of the Indo-European phylum, little more than three nodes, Anatolian–“Late PIE”, Balto-Slavic, and Indo-Iranian, could be argued to be uncontroversial, at least for the most part, with a large “bush-like” node of nine branches not amenable to phylogenetic resolution. Figure 8.2-4 contrasts the minimal, or bush-like, phylogeny that can be said to be widely accepted with the more speculative but still fairly mainstream articulation distinguishing “early” (Proto-Indo-Hittite), “middle”, and “late” stages of Proto-Indo-European (PIE).

With the addition of nodes at least widely held to be somewhat plausible – the early divergence of Tocharian, and the Italo-Celtic and Palaeo-Balkan (Graeco-Armenian) groupings – the bush-like node still unites at least five branches. Here, the question arises of what exactly should be understood as a genetic relationship between languages: languages thus grouped together may or may not reflect prolonged linguistic contact between already articulated groups, in the sense of a *Sprachbund*, at a time in prehistory when these languages had not yet diverged very widely.

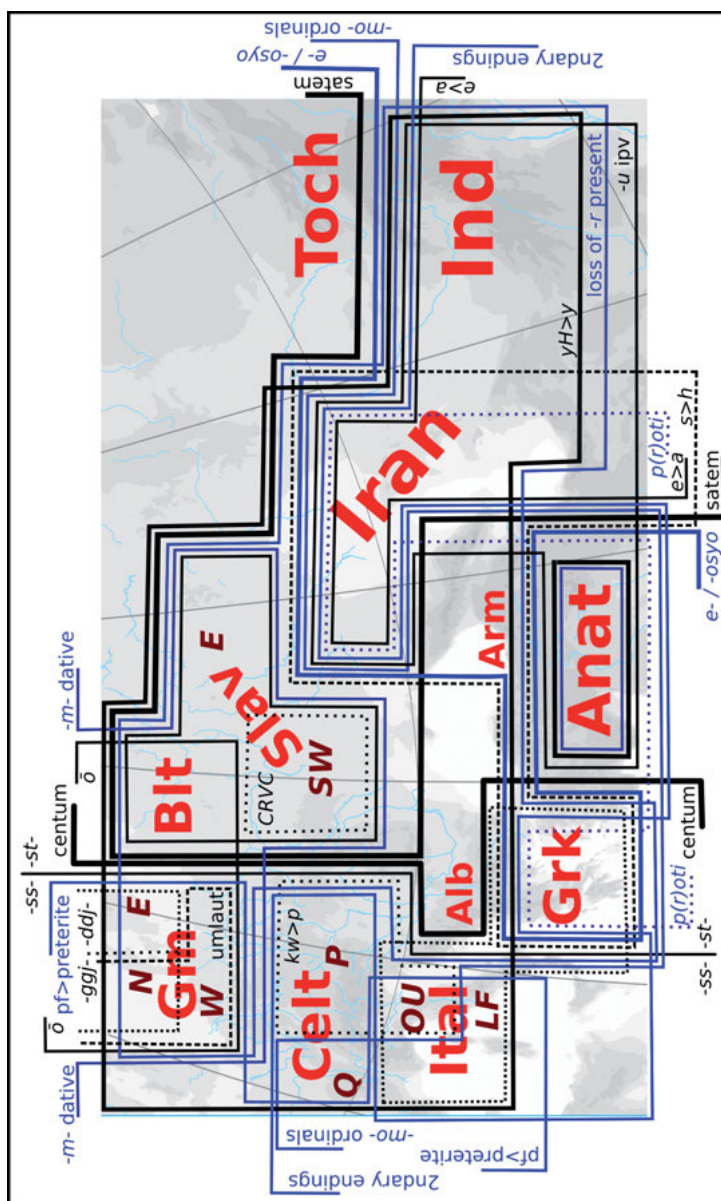


Fig. 8.2-2: A map of major isoglosses separating the twelve main Indo-European language families, after Anttila (1989, 305). All isoglosses shown are morphological (blue) or phonological (black), not lexical. Language families are shown in the approximate geographical location they (or their ancestral dialects) would have had around the late Bronze Age (although some of the later isoglosses may not have developed before the Iron Age). The isoglosses selected in Anttila (1989) divide some of the main families into subgroups: North, East, and West Germanic; East Slavic vs South and West Slavic; Q-Celtic vs P-Celtic; and Osco-Umbrian vs Latino-Faliscan within Italic.

The contrast between *Stammbaumtheorie* and *Wellentheorie* is illustrated in figures 8.2-2-3. The map in figure 8.2-2 shows a selection of major isoglosses separating the twelve main Indo-European language families (after Anttila 1989, 305). Isoglosses were selected for grouping families; each named family has numerous isoglosses unique to itself that are not shown. All isoglosses shown are morphological (in blue) or phonological (in black), not lexical. Any crossing isoglosses indicate the presence

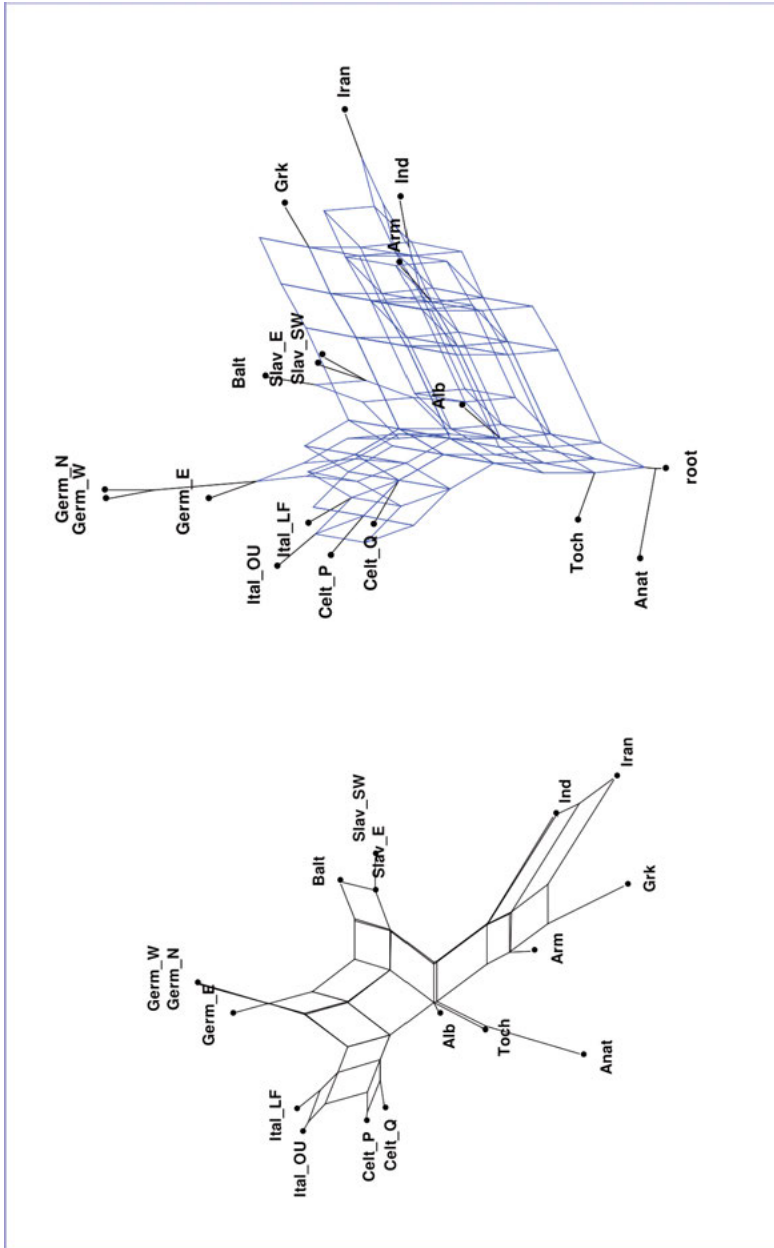


Fig. 8.2-3: Visualisation of the isogloss topology shown in figure 8.2-2 as phylogenetic networks, illustrating the significantly non-tree-like structure of Indo-European. Left: an unrooted network generated using the NeighborNet algorithm (D. Bryant and Moulton 2004); right: a rooted hybridisation network (Huson and Klöpper 2007) with Anatolian as outgroup, calculated using SplitsTree4 (Huson and Bryant 2006). A similar analysis has been published by Heggarty, Maguire, and McMahon (2010).

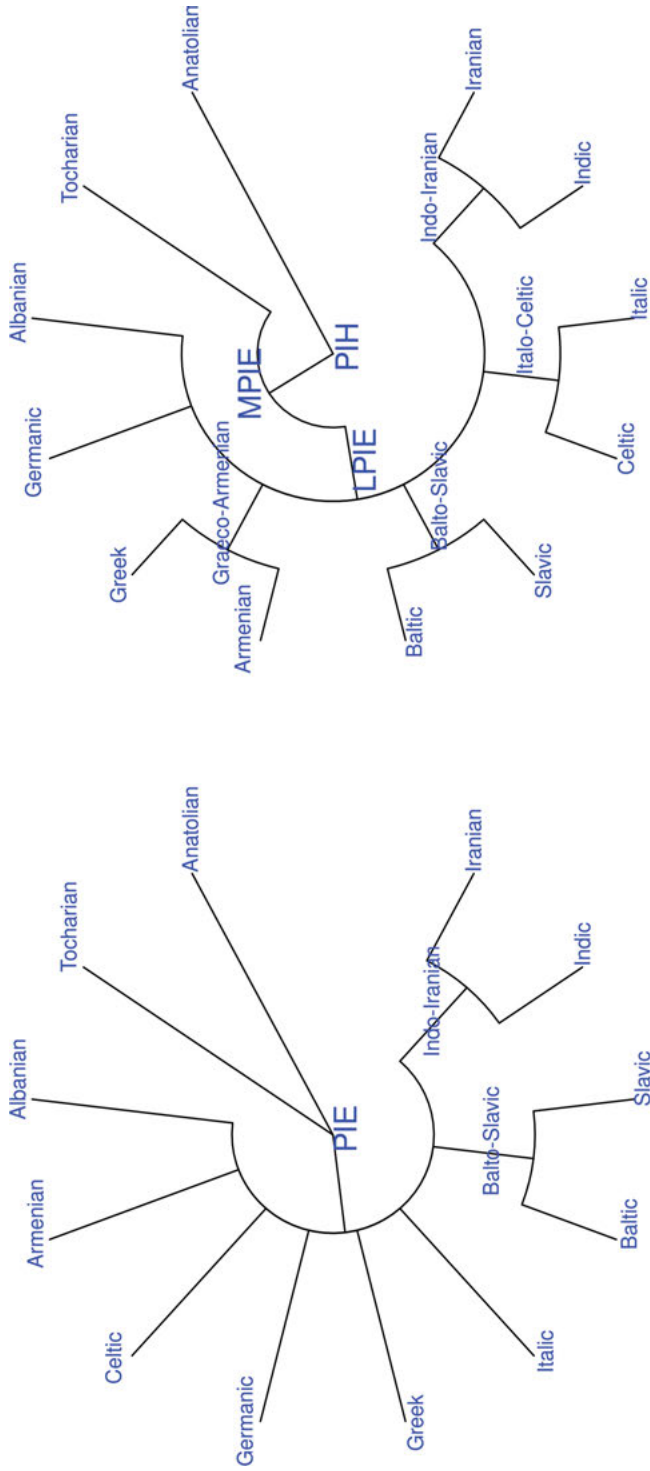


Fig. 8.2-4: Comparison of the unarticulated, high-consensus and the more articulated, low-consensus phylogenetic diagrams of Indo-European language families drawn as rooted phylogenetic trees. The more articulated variant distinguishes “early” (Proto-Indo-Hittite, PIH), “middle” (MPIE), and “late” (LPIE) stages of Proto-Indo-European.

of non-tree-like signals. Figure 8.2-3 uses these isoglosses to construct two phylogenetic networks. This can be taken as a rough visualisation of the significantly non-tree-like structure of Indo-European, evident even from a very limited (albeit not randomly selected) dataset based on phonology and morphology only. By contrast, figure 8.2-5 illustrates the use of purely lexical data typical of the lexico-statistical approach. A separate phylogeny can in principle be derived for each lexeme independently – in the case shown here, the lexeme for “sun”. The task of calculating a phylogeny for entire language families would then amount to the search for a “consensus tree” for the phylogenies assumed for the individual lexemes.

It may be important to remind ourselves, focusing on the question of phylogenetics as we are here, that this inconclusiveness does not by any means represent a failure of the two-centuries-long project of Indo-European reconstruction: Indo-European studies is primarily a philological field tasked with studying the historical development of the recorded languages and with reconstructing their common proto-language. There is no compelling reason for the philologist to assume that any given subset of linguistic history can necessarily be represented by a genealogical tree with any degree of accuracy. The burden is on the phylogeneticists (that is to say, on such historical linguists as advocate phylogenetic methods) to show that their methods are of any use in the field. This point will become important presently in understanding the cool reception which the more recent attempts at introducing methods of computational phylogenetics into historical linguistics have tended to be met with.

8.2.3 Other phyla and proposed macro-families

Indo-European has been something of an ideal case for the development of historical linguistics and the comparative method: its numerous branches, its great time-depth, and the antiquity of the written record of several of its member families were all conducive to a bountiful picking of low-hanging fruit in the early phase of the field’s history. While the phylum does not, as we have seen, present itself as overwhelmingly tree-like, it is still sufficiently “bush-like” to suggest genealogy rather than a history of random or purely areal wave-like exchange of linguistic features.

There are only a handful of candidates for linguistic phyla with similar properties among the world’s languages. First, there are Afro-Asiatic and Sino-Tibetan, perhaps both, or at least the former, likely of an age comparable to that of Indo-European. Austronesian is of more recent attestation, but of no lesser attractiveness in this context, making up for its limited time-depth by the remarkably tree-like structure resulting from the literal isolation of its (island-dwelling) communities of speakers. Several others, such as Uralic, Niger–Congo, Mon–Khmer, Uto-Aztecan, or Tupi, may be of considerable antiquity, but reconstruction is fraught with difficulties due to very late attestation. Most other proposed large phyla are even more

difficult to tackle, and often their very existence as a phylogenetic unit is disputed (Altaic, Nilo-Saharan). Other families are less diversified, comparable perhaps to the divisions within some of the branches of Indo-European (e.g. Dravidian, Tungusic, Eskimo–Aleut). In summary, there are perhaps in the order of half a dozen major linguistic phyla which can reasonably be hoped to yield information on proto-languages spoken in deep prehistory, that is, before around six thousand years ago (the Old World Neolithic or Chalcolithic). The scholarly effort poured into the comparative study of these families is, of course, far beyond the capacity of a single scholar, which may have led to a fragmentation of the field (Kortlandt 1995).

Beyond the reconstruction of the deep history of non-Indo-European phyla, there are proposals which combine already reconstructed phyla into super-phyla, which might then reach down to late Palaeolithic times, ten thousand years ago or earlier. The most notable proposal in this context is Nostratic (named thus by Holger Pedersen in 1903), which would combine at least Indo-European, Altaic, and Uralic into a large Eurasian super-phylum, besides various proposed subgroupings such as Indo-Uralic. The main proponent of this type of reconstruction during the second half of the twentieth century was Joseph Greenberg, who went against the linguistic mainstream in proposing that genealogical relationships can, after all, be extricated for very remote phyla if enough data is available (a notion which became known and criticised under the name of “mass comparison” in historical linguistics). While such proposals have gained somewhat wider popularity since the 1990s, it must be emphasised that they are much more speculative by their nature than conventional linguistic reconstruction, and not based on evidence of comparable reliability to that gained by the philologist’s careful search for archaisms. As pointed out by Kortlandt (1995), comparative linguistics prides itself more in basing its conclusions on a small body of unassailable evidence than trusting in statistics. For this reason, Greenberg’s mass comparison approach never achieved recognition in the mainstream of historical linguistics (Campbell 2003).

Greenberg can still be considered a pioneer of a data-driven, algorithmic approach reminiscent of the phylogenetic methods used in biology. His approach is also the only viable one in the case of large language families without a literary tradition, as in the case of African or New World languages. It is certainly meaningful to use mass comparison to group the hundreds of languages in the Niger–Congo group, quite regardless of whether this group is considered a phylum in the phylogenetic sense; the interpretation of such a grouping as a genetic classification goes to the heart of the question of what exactly we mean when we talk of the genetic relationship of languages. The main worry is essentially that the method is underdetermined, so that an algorithmic approach of this kind will always result in the “discovery” of ever larger “genetic” macro-families regardless of the data, without any objective criteria to judge the reliability of such proposals. For the macro-families discovered in this way, it is intractable whether their shared features (even assuming that pure chance, or onomatopoeia, has been duly ruled out) are in fact

genetic in nature rather than merely areal and the result of prolonged language contact. The case of the Indo-Pacific family proposed by Greenberg himself in 1971, claimed by supporters to be at least forty thousand years old (Ruhlen 1994, 144), is just one example.

A level even more remote from the classical comparative method is represented by theories of monogenesis or glottogony, that is, the derivation of all known human languages from a common origin spoken in the middle Palaeolithic (the time of phylogenetic unity of *Homo sapiens* itself, or at least prior to the long-lasting division of humanity by its dispersal across continents). This debate can no longer pretend to be based on comparative evidence, however sparse. Instead, the argument is one of plausibility: since language is a human universal, it stands to reason that its genealogy goes back to the original dispersal of anatomically modern humans, represented by the split of the hunter-gatherers of southern Africa from the remaining populations of early *Homo sapiens*, at least 150,000, or possibly as early as 300,000 years ago (Schlebusch et al. 2017). The topic of glottogony has long been a taboo in linguistics, for the very good reason that it attracts boundless speculation: when Darwinism was freshly *en vogue* in the 1860s, linguistic debate was inundated with such proposals to the point where, in 1866, the Linguistic Society of Paris banned the entire topic as a hopeless exercise, a prohibition that had a lasting influence in the field of historical linguistics. The question of the first development of the human language faculty is of course a valid topic in contemporary evolutionary anthropology, but it is hardly amenable to the methods of historical linguistics. It may still be worth mentioning in this context that a 2012 study arrived at a plausible middle Palaeolithic estimate for the age of Proto-World based on a simple model of gradual increase of phonemic diversity over time (Perreault, Mathew, and Petraglia 2012).

8.2.4 Time-depth and *Urheimat*

A genealogical view of language is necessarily concerned with the history and pre-history of human migration. The parallelism of the genealogy of languages and the genealogy of the “races of man” were made explicit in the absolute in the statements by Leibniz (1710) and Darwin (1859) discussed above. And, even though the phenomenon of wave-like innovations through language contact is of course real, any postulate of a language family in the genealogical sense implies the geographical separation of a population of speakers of a proto-language over a period of time sufficient for language change beyond mutual comprehensibility.

The question of the original home or *Urheimat*, or homeland, of the Indo-European phylum has received considerable attention ever since the existence of the phylum was recognised. The question does, however, go beyond the field of historical linguistics proper, and requires the evaluation of evidence from archaeol-

ogy and anthropology. The early history of the preferred scholarly hypotheses regarding the Indo-European homeland does not inspire confidence. Opinion has fluctuated based on intellectual fashion in the best case, and based on political ideology in the worst. In the early nineteenth century, enthusiasm for the recent discovery of the surprisingly archaic nature of Sanskrit – combined with the equally recent infatuation of Western romanticism with Buddhism and Hinduism – produced the “Indomania” of which Friedrich Schlegel was the primary proponent. In the mid-nineteenth century, based on more systematic studies of the vocabulary for plants and animals, central Asia became the leading hypothesis, favoured by Max Müller, Franz Bopp, Jacob Grimm, and others (Sasse 2017, 334). During the second half of the nineteenth century, the central Asian hypothesis stood opposed to a northern European one. Northern Europe was strongly preferred by British scholars – such as Robert Latham, Canon Isaac Taylor, and John Rhys – in particular, and the idea was well received in the Scandinavian Romantic nationalism of the time. In the early twentieth century, northern Europe also tended to become favoured among German scholars such as Matthäus Much and Hermann Hirt (Hirt 1905–1907, 1:334). This is the origin of the unhappy association of the term “Aryan” (at the time used as a name for the entire Indo-European phylum) with northern European ethnography, even though Max Müller had warned, already in the 1880s, against the rash conflation of linguistic and anthropological features (“an ethnologist who speaks of Aryan race, Aryan blood, Aryan eyes and hair, is as great a sinner as a linguist who speaks of a dolichocephalic dictionary or a brachycephalic grammar”; Müller 1888, 120).

The question of geographical origin of a language family is of course closely tied to that of its temporal depth. The term “glottochronology” was coined for methods to estimate the age of reconstructed prehistoric languages – the “molecular clock” (see 8.1.1.3), as it were, of language change – by Morris Swadesh in the early 1950s (Jenset and McGillivray 2017, 62). Comparative linguists in the later twentieth century have tended to balk at participating in glottochronology or *Urheimat* debates, emphasising that comparative reconstruction has validity in its own right, but is by its nature capable of reconstructing linguistic features only, not of attaching an absolute time or place to the reconstructions. Indeed, arguments for both geographical and temporal estimates on the basis of reconstructed vocabulary can carry weight only in aggregate: semantic shift prevents the preservation of the type of compelling archaism favoured in purely linguistic reconstruction. For example, the reconstruction of a PIE word **laks* for “salmon” was used to argue for an area of origin in northern Europe where Atlantic salmon is found, first proposed by Otto Schrader (1883). This argument, in isolation, carries no weight, as the word in question could well have referred to trout in the proto-language and might have been independently transferred to salmon by such Indo-European groups as happened to settle near the North Sea (Thieme 1953). Similarly, the reconstruction of a PIE word **hrotha* for “wheel” or “wagon” would necessarily require a date for the proto-

language later than the invention of the wheel about six thousand years ago. But the word might be a derivation from a verbal root meaning “to roll” and thus indicate any “rolling thing” without necessarily implying the presence of an axle and wheel, and, similarly to the “salmon” case, could easily have been independently used as a name for the wheel or chariot once they became available.

In the late twentieth century, a wide-ranging consensus nevertheless emerged, which favours the placement of the Indo-European proto-language on the Pontic–Caspian steppe roughly 6,500 to 5,000 years ago, that is, at the boundary between eastern Europe and central Asia geographically, and between the late Neolithic and the early Bronze Age temporally. This steppe-origin model is also known as the *Kurgan hypothesis*, advocated by archaeologist Marija Gimbutas, beginning in the 1950s. It was to a large extent predicated on the progress made by Soviet archaeology in the early to mid-twentieth century, and centres on the Pit Grave, or Yamnaya, cultural horizon and its immediate predecessors (*kurgan* being the Tatar and Russian word for the burial mounds associated with this and later cultures of the area). The Kurgan hypothesis was widely popularised by Mallory (1989). Kortlandt (1990) discusses the hypothesis in terms of a series of waves of expansion from the steppe area; an Indo-Hittite separation as early as 6,500 years ago; and the Yamnaya context of about 5,000 years ago, which he identifies with those groups affected by satem, that is, the proto-languages of Balto-Slavic and Indo-Iranian.

While the Kurgan hypothesis, among such Indo-Europeanists who do not prefer to remain agnostic as a matter of principle, has eclipsed the early theories of a northern European or Indian homeland, there are two competing hypotheses proposed in the 1980s that may be worth mentioning, both favouring a homeland in Asia Minor. The Armenian hypothesis proposed by Gamkrelidze and Ivanov (1984) proposes a late date (fourth millennium BC), and hinges on an idiosyncratic reconstruction of PIE phonology by these authors which has found little mainstream interest. The Anatolian hypothesis proposed by Renfrew (1987), on the other hand, proposes an extremely early date (seventh millennium BC), and equates the Indo-European expansion with the Neolithic expansion from Asia Minor to Europe. This suggestion has likewise found little support among Indo-Europeanists, but it has become notable in the context of the computational modelling of Indo-European phylogeny in the 2000s (see 8.2.6).

There is no shortage of treatments of the conflation of the Indo-European homeland question with nationalist ideology, which has indeed occurred, not exclusively but most notoriously in Germany. It is less common to point out that the history of the now most widely accepted scenario likewise has an ideological component, one of “eco-feminism” (Gimbutas, at least in her later years, painted a picture of the Indo-Europeans as patriarchal warriors invading peaceful matrilineal cultures of Neolithic Europe). As noted by Anthony (1996), such “interpretive abuse” of prehistoric migration in the light of present-day ideological preferences has caused many archaeologists to view the project of comparing linguistic and archaeological evidence with suspicion.

Urheimat debates for other linguistic families are plagued by similar uncertainty. For Afro-Asiatic, the debate is divided between the hypotheses of an African (most likely in or close to the Horn of Africa) and an Asiatic (Levantine) hypothesis (Blench 2006, 144). For Austronesian, the homeland question is tied to the speed of expansion, the “slow boat” hypothesis suggesting a deep prehistory in Melanesia as opposed to the “express train to Polynesia” hypothesis suggesting more recent origin in Taiwan. This latter question of Austronesian origins, at least, has a reasonable chance of being amenable to conclusive resolution based on population genetics (Oppenheimer and Richards 2001).

Kortlandt (1990, 1) states the following caveat on the difficulty of aligning archaeology and language: “Speculations about the linguistic affinity of a prehistoric culture are futile because it is reasonable to assume that the vast majority of prehistoric linguistic groups have vanished without leaving a trace.” There is, that is to say, an asymmetry between archaeology and linguistics because the archaeological record is potentially far more complete than the linguistic one. This is a sentiment often expressed in comparative linguistics: since most of the history of languages is lost without trace, such vestiges as we have are due to the accidents of preservation and cannot be regarded as a representative sample of anything. The same view is expressed from the point of view of archaeology by Blench (2008). It is certainly well advised to avoid the rash conflation of archaeology, ethnography, and linguistics. Nevertheless, there is no reason to disregard independent evidence that may increase or decrease the plausibility of a hypothesis. And, in any case, speculation is necessary if progress is to be made, as in the case of the Kurgan hypothesis, which was originally proposed as an archaeological hypothesis and gained its mainstream status later due to cumulative evidence which happened to favour it quite independently of a possible ideological backdrop to its original formulation.

Unfortunately, during the 1960s to 1980s, scepticism towards such proposals was extended beyond the reasonable by an ideological fashion which rejected “migrationism” as nationalistic or imperialistic in favour of a “diffusionism” based on Marxist archaeology. The 1970s adage of “pots are not people” is trivially true, of course, but it entailed the prohibition of any attempt to link the two as politically irresponsible. This ideological reluctance to view material and linguistic prehistory in combination most strongly affected the English-speaking world, where it reached its apex in the fully postmodernist “post-processual archaeology” of the 1980s before it began to wane during the 1990s (Härke 1998).

It is partly for such reasons that progress regarding the age and phylogenetics of linguistic phyla, primarily of Indo-European, was not revived, and did not attract mainstream interest, until the 1990s. Other, more pragmatic reasons lie in technological advances that only then became available: the rapid progress in DNA sequencing (8.2.6) and the availability of computational methods capable of tackling complex phylogenetics (8.2.5).

8.2.5 Computational estimates

Comparative linguistics is traditionally concerned with the reconstruction of a proto-language without necessarily opining on the phylogenetic relationship between daughter languages, let alone on absolute time-depth. To this end, the most valuable information comes from archaisms, for example in the form of grammatical irregularities. A famous example is the exact equation of Sanskrit *vr̥kīs* and Old Norse *ylgr*, “she-wolf”, discovered by Karl Verner in 1877. This not only establishes the existence of a word for “she-wolf” in the parent language of Germanic and Indic, right down to its accentuation and inflection; it also establishes the presence of a grammatical feminine, or at least a morphological derivation for the female of an animal species. In contrast to stemmatology, perhaps, the ubiquitous possibility of language contact (or stemmatological contamination) reduces the value of what look like common innovations (or stemmatological *Leitfehler*). For these reasons, computational phylogenetic methods have so far played a limited role in comparative linguistics. Attempts to generate phylogenetic trees of language families have, of course, been made since the 1960s, but these have always been of limited or no use in adding information to what has already been worked out by the experts: the quality of the results achieved by the new method has been judged based on what was already known rather than in terms of previous knowledge being evaluated or verified in light of the new results. The main difficulty in such approaches lies in the choice of what type of information to use as input data. Most of the time, such material will already implicitly contain the judgement of the philologists who, for example, provided the etymological dictionaries used to produce a table of cognates used as the input to be analysed by the phylogenetic algorithm. In his overview of the history of phylogenetic methods applied to language history, Dunn (2014, 192) emphasises that “correct cognate classification is no trivial matter”.

By far the most popular approach is the use of Swadesh lists, the comparison of languages based on the number of cognates they share in a short list of core vocabulary. As an example, let us consider a single lexeme of the Indo-European Swadesh-100 list in Dyen, Kruskal, and Black (1992). The word for “sun” in Indo-European, reconstructed as nom. **sahwl*, gen. **shwens* (a heteroclitc stem with alternating *-l-* and *-n-* suffix), can be traced throughout most of the branches of Indo-European, and in some cases undergoes characteristic evolution which is not only phonological but also morphological (see fig. 8.2-5). A Swadesh list-based phylogenetic study of the type of Gray and Atkinson (2003) will retain from this complex reconstruction only the binary information that the original lexeme is present in all branches other than Anatolian, Tocharian, Armenian, and Albanian (and is further absent in Gorkhali, an Indic language which appears to have adopted the same non-Indo-European lexeme found in Tocharian, and in Irish, which has retained as the unmarked word for “sun” what seems to be an epithet derived from the Indo-European word for “warm”).

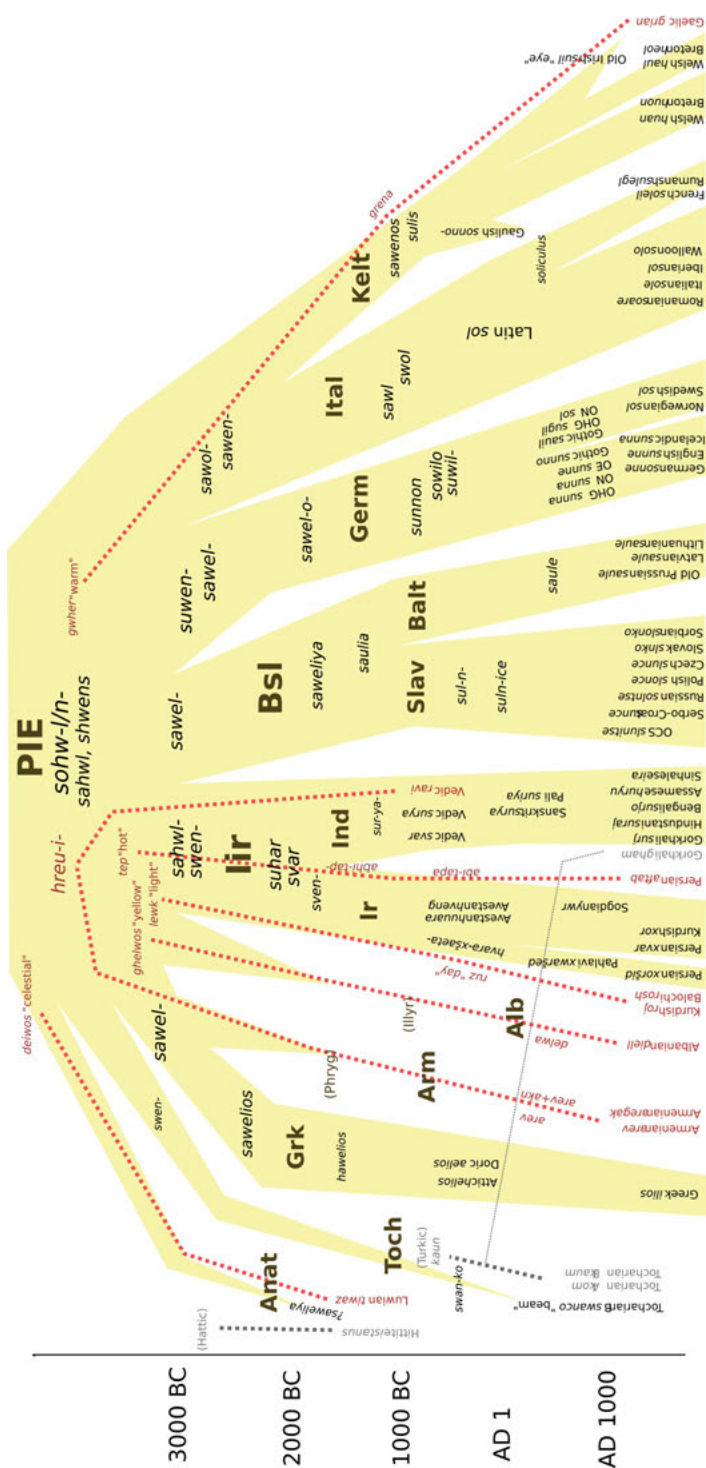


Fig 8.2-5: This figure shows a simplified phylogeny for the lexeme for “sun” in Indo-European. The forms given in attested languages are based on the *Indo-European Lexical Cognacy Database (IExlex; iexlex.mpi.nl; originally from Dyen, Kruskal, and Black 1992)*, with simplified orthography. The time scale on the left is intended as a rough approximation only. The phylogenetic diagram shown in outline traces derivations from the PIE word for “sun”, nom. **sahwl*, gen. **shwens* (a recent publication on the Indo-European lexeme is Simms 2017). Dotted lines trace cases where the word for “sun” was replaced with a different lexeme. A lexico-statistical analysis of this word in a Swadesh list will treat all words derived from the **sahwl*, **shwens* paradigm as etymologically identical.

Considering the extreme simplicity of the Swadesh list approach, it is capable of producing surprisingly solid results. But the method is abused if it has to carry epistemic weight of the kind of “English and Russian share 34 percent of lexemes in the Swadesh-100, which would imply that their common ancestor split [...] approximately 4000 years ago” (Pereltsvaig and Lewis 2015, 94). On the other hand, it is very difficult to go beyond the Swadesh list approach, as the more parameters are introduced, the more subjective the decision about which to include and how to weight them against one another becomes.

In the early 2000s, a series of publications presented reconstructions of language phylogenies calculated on the basis of Swadesh lists. Dunn (2014, 206) dates the beginning of the “modern phylogenetic era in linguistics” to the publication of Gray and Jordan (2000), who argued for the rapid model of Austronesian expansion based on a phylogenetic tree constructed from vocabulary lists.

An early such study dedicated to Indo-European, published in 2002, was a collaboration of two historical linguists and a computer scientist. Marked as “work in progress”, the paper discusses methodology and presents a number of best trees based on a database including phonological, morphological, and lexical features. The paper’s main takeaway is that “we need to devise appropriate methods for inferring non-treelike networks of linguistic diversification”, echoing the conclusions drawn from the original *Stammbaumtheorie* approach in the 1870s (Ringe, Warnow, and Taylor 2002, 112).

The publication by Gray and Atkinson (2003) was less modest in its conclusions, and much more widely received. Published in *Nature*, this paper announced that “language-tree divergence times support the Anatolian theory of Indo-European origin”. Based purely on Swadesh lists of modern Indo-European languages (with the addition of Hittite and Tocharian lists), the authors used a phylogenetic method which allowed for variable “molecular clocks” to estimate divergence times in their consensus tree. The claimed support for the Anatolian theory (Renfrew 1987) was derived from an estimated age of the tree root of close to nine thousand years, which, the authors concluded, favoured the theory of a spread of Indo-European in the course of the Neolithic revolution over the Chalcolithic steppe origin proposed by the Kurgan hypothesis. The main criticism that has to be raised against the 2003 paper is clearly that it was overselling itself with sensationalist PR. The strength of the study was the presentation of a surprisingly credible phylogeny with branch lengths, and its use of a variable clock genuinely addressed the most prevalent criticism of the failings of naive glottochronology. It was not concerned with Anatolia in any way, and the estimated age of the root was just one feature, and indeed the least reliable one in the reconstructed tree (because, at the root of the tree, there is no longer any data to suggest variation of clock speed). This result may serve as an illustration of the observation made more than a decade earlier by Kortlandt (1990, 2), who diagnosed “a general tendency to date proto-languages farther back in time than is warranted by the linguistic evidence”. The problem addressed

by Kortlandt is at least partly a terminological one: should, for example, “proto-Romance” refer to the time of the breakup of the varieties of Latin spoken in the late Roman Empire, or should it refer to a – possibly ahistorical – complete unity of the Latin language in early Rome about a millennium earlier? And, by analogy, should “PIE” refer to the period of Indo-European expansion and breakup, or should it refer to a theoretical dialect-free complete unity of Indo-Hittite? It is this exaggerated postulate of complete linguistic unity in a proto-language, never observed in any living language, that causes the excessive age estimates. In addition, the absence of any “methods for inferring non-treelike networks” forces the algorithm to assign temporal depth to what were likely wave-like or areal innovations across the early phase of IE expansion.

In 2012, Gray and Atkinson along with a larger author collective (Bouckaert et al. 2012) doubled down, this time with a publication in *Science*, repeating their result of an age for PIE of close to nine thousand years, but the phylogeny based on Swadesh lists had now been augmented by a Bayesian phylogeographical framework developed in 2010 to locate the origin of virus outbreaks. The authors did simulate the shape of the Eurasian landmass, but no topographical features, and determined Anatolia as the most likely geographical origin of the tree. The result is highly interesting but ultimately unconvincing. All topographical features are ignored, yet, as the historical record tells us, the steppes are vastly favoured as an invasion route. In addition, the random walk algorithm at the core of the method had little incentive ever to “set foot” near the Pontic steppe, as the language families (other than Anatolian and Tocharian) were given their mediaeval ranges as priors, with the entire steppe area, which we know was once Scythia, and populated by East Iranian-speakers, being treated as non-Indo-European. In effect, the historical circumstance of the Turkic expansion of the first millennium thus affected the likelihood estimate for a Pontic homeland several millennia before. On top of this, the exaggerated age estimate of proto-Anatolian discussed above might also have favoured Anatolian origin, as suggested by the reported loss of confidence for the Anatolian homeland by an order of magnitude once only living languages were taken into account (Bouckaert et al. 2012, 959, report a Bayes factor for “Anatolian vs. steppe II” reduced from 159.3 for “all languages” to 11.4 for “contemporary languages only”).

For better or worse, the two papers (Gray and Atkinson 2003; Bouckaert et al. 2012) were widely advertised in the popular press as establishing the Anatolian origin of the Indo-Europeans, and caused substantial comment. Phylogenetic studies of numerous language families other than Indo-European have been performed since the early 2000s, and several conferences on the topic have been held since 2010 (summarised in Pereltsvaig and Lewis 2015, 55–57), but the Indo-European case has continued to attract the most attention and controversy.

Pereltsvaig and Lewis (2015, 3), a monograph mostly inspired by the Gray and Atkinson papers, is emphatic in pointing out the “spectacular failure” of the computational approach from the point of view of historical linguists. Unfortunately, the past decade appears to have deepened the chasm between historical linguists and

computational phylogeny, as linguists accuse the phylogeneticists of lack of respect for their field and “unjustified and unjustifiable simplification” (Pereltsvaig and Lewis 2015, 127), while often lacking an understanding of the strengths of the methods proposed. On the other side of the divide, Atkinson and Gray (2006) and Greenhill and Gray (2009) attempt to explain their methodology with palpable frustration. The blame for this decade of unease in interdisciplinary exchange is most likely shared. On the one hand, the phylogeneticists, at least in their early publications, have been over-confident. On the other hand, critics have largely failed to engage with understanding the methods used and their genuine strengths. In an ideal world, Gray and Atkinson would have presented their approach as a preliminary study showing the promising potential of Bayesian phylogeny, to be revised in future studies with better linguistic data and better models of migration. Instead, the debate was immediately reduced to the Anatolian hypothesis. The complaint by Greenhill and Gray (2009, 390) regarding the “vexing misconception about phylogenetic linguistics: ‘this method is not giving anything new’” may be instructive: it is true that phylogenetic methods may bring new, rigorously objective tools for assessing the degree of relationship between languages. On the other hand, it is quite obvious to the historical linguists that the quality of the results produced so far is not comparable to what has long been achieved by traditional philology: the credibility of the trees produced by the phylogeneticists is still judged by their congruence with traditional results, not vice versa (Greenhill, Drummond, and Gray 2010).

Many more examples of computational phylogenetics applied to historical linguistics were published from the 2000s to the early 2010s. Dunn et al. (2005) attempted to compare Papuan languages in the complete absence of etymology, purely based on typological similarity. Dunn (2009) is a similar study of Melanesian languages, notable as an early use of likelihood (Bayesian) algorithms rather than simple distance-based methods. These are examples of computational methods being used to arrive at an assessment of the phylogeny in groups of languages in the complete absence of any historical record or any etymologies. While it is not inconceivable that a phylogenetic signal may be detected using such brute-force methods, it is very difficult to quantify the confidence that should be placed in the results. The main problem of phylogenetic algorithms is, of course, that they will *always* result in a best-estimate tree, regardless of whether any actual phylogenetic signal was present in the data. Careful statistical analysis is especially important here in order to be able to quantify the probability that the signal discovered by the algorithm might be due to chance alone. Perhaps an instructive example of this effect is Fortunato (2011), a study which attempts to reconstruct PIE monogamy using phylogenetic methods (applied to languages) in combination with sociological data from an ethnographic atlas. The result presented is that polygyny was a Proto-Indo-Iranian innovation. While proper caveats are included in the article text, the result was nevertheless cast in terms of “evidence in support of PIE monogamy; this pattern likely extended back to PIH” (Fortunato 2011, 99). Without wishing to

embark on too much of a tangent, this appears to be quite a *bold* conclusion from the mere distribution of modern-day polygyny in Iran and India without considering other, areal effects, such as the spread of Islam, which may have contributed to it, and in the complete absence of any knowledge as to the age, cultural stage, or economic situation of Proto-Indo-Hittite.

Unfortunately, the use of more advanced classes of algorithm, such as maximum parsimony or Bayesian inference, has not resulted in a noticeable improvement of the quality of results achieved in language phylogenies compared to what is quite easily derivable using distance-based methods. Dunn (2014, 197) names as one possible reason for this the phenomenon of long-branch attraction, where the parsimony algorithm tends to unduly combine areas of the phylogeny with a lot of change (long branches), fulfilling, as it were, the requirement of being maximally parsimonious by reducing the total number of mutation-heavy areas in the phylogeny.

Use of computational methods has also been attempted in the quite different context of dialectometry. In this case, the close historical relationship of the species under consideration is taken as a given, and the algorithm serves as an aid in giving a visualisation of the structure of the dialect group, for example in the form of network graphs (Dunn 2014, 195). Manni, Guérard, and Heyer (2004) used a boundary-detecting algorithm (Monmonier's algorithm) to automatically detect language barriers in dialect data.

Phylogenetics software has become more readily available and more sophisticated since the early 2000s. BEAST, in particular, is software developed in part for the purpose of modelling linguistic phylogenies (Drummond et al. 2012). It is unclear, however, whether increased sophistication in method will be able to substantially improve on current possibilities. Ultimately, the entire comparative method of historical linguistics can be seen as a single vast exercise in phylogenetics. An automated method approximating this programme would need to fully combine lexical, phonological, and morphological information into a single model. While this is perfectly possible in principle (Dunn 2014, 204), it is a formidable task, which can perhaps only be envisaged once fully machine-readable historical grammars and morphologically analysed text corpora become available for the languages in question, and even then it is questionable whether the computational power required to successfully run a Bayesian tree search on such an amount of data is within the realm of the possible, even in the longer term of technological progress.

8.2.6 Evidence from human genetics

With the rapid development of technological possibilities in genetics during the 2010s, many long-standing intractable questions related to the correlation of changes in material culture or language and migration now seem amenable to objective, quantitative analysis. Human population genetics has the potential of serving as

arbiter in the old “pots vs people” debate. Genetic studies of the 2000s were mostly not autosomal, limiting themselves to patrilineal and matrilineal descent, leading to inconclusive results. Since the 2010s, both autosomal sequencing technology and ancient DNA have become much more readily available, in many cases leading to surprisingly detailed insights into prehistoric migrations.

For the historical period, in cases where a linguistic shift is a matter of record, it has become possible to estimate the extent of population movement associated with it. A prominent example of this kind is the question of the Anglo-Saxon settlement of Great Britain, a long-standing controversy of the “migrationist vs diffusionist” type. The gradual influx of West Germanic-speaking settlers in the fifth and sixth centuries was sufficient to replace the Celtic British language with the Germanic Anglo-Saxon one in what is now England. In the mediaeval period, the arrival of a comparatively small Norman-French elite was not quite sufficient to displace Anglo-Saxon, although a substantial impression from the French superstratum has been left on the English language. In the case of Anglo-Saxon migration, a 2016 study found evidence of early intermarriage between Anglo-Saxons and Britons (rather than a “genocidal” or “apartheid” scenario), with a genetic contribution in the order of 40 % from the Anglo-Saxon settlers to the modern population of eastern England (Schiffels et al. 2016). The accumulation of this and similar examples of migration and linguistic change during the historical period may potentially serve to inform the study of prehistoric migration and language change.

Studies focused on the period of European prehistory relevant for the Indo-European question are bound to shed further light on the Anatolian vs Kurgan scenarios. Haak et al. (2015) is a study of the DNA of sixty-nine European individuals of the Neolithic and Bronze Age. The study was able to determine that the Neolithic farming population arrived from the Near East about 8,000 years ago, and that there was a gradual intermixture with people of local hunter-gatherer ancestry, followed by the sudden appearance of high amounts of “steppe ancestry” on the German plains at 4,500 years ago. This corresponds almost exactly to the classical scenario connecting the Beaker people with the Kurgan expansion and early Indo-European presence in western Europe. This is, obviously, by no means a linguistic result, and it cannot be ruled out, for example, that early forms of Indo-European were present in Europe before this (and Haak et al., abstract, are careful enough to postulate “a steppe origin” only for “at least some of the Indo-European languages of Europe”), but it does allow us for the first time to estimate the population movements associated with shifts in prehistoric culture such as the arrival of the Bronze Age in western Europe. Olalde et al. (2018) is a more extensive analysis of individuals associated with the Beaker culture, finding that the expansion of the Beaker complex was mostly, but not always, paired with substantial migration: in Britain, there was a substantial demographic transformation due to the influx of steppe-related ancestry, while the Beaker culture appears in Iberia with only limited presence of steppe ancestry. Reich (2018) summarises the recent confirmation of the appearance of

“steppe” or “Yamnaya” genetic markers in European ancient DNA at 4,500 years ago, precisely as would be expected from the hypothesis of Indo-European expansion from the steppe at that time. The harsh impact of the new arrivals on genetic lineage, even to the point of extinction or near-extinction of the indigenous male line in the case of the Iberian peninsula, goes some way towards vindicating Gimbutas’s views of the arrival of the Indo-Europeans as a warlike invasion of a small patriarchal elite (which resulted in a bottleneck in paternal lineages, while maternal lineages remained more diverse).

Even if these results are subject to further revision, their availability presents an entirely new foundation for the old “pots vs people” debate: instead of an ultimately ideological dispute over the relative importance of cultural transmission and human migration, we are now beginning to gather quantitative information on the balance between these two modes depending on time and geographical region. It might be a promising avenue for future Bayesian language phylogeny studies to include in their models such quantitative information about prehistoric population movements as a guide.

8.3 Anthropology

Jamshid Tehrani

Phylogenetic methods were originally developed to study the evolutionary relationships among biological species, but in recent years they have been adopted by a growing number of researchers in the humanities and social sciences to investigate the historical development of cultural traditions. As Howe and Windram point out in an influential review of this literature,

in principle, phylogenetic methods can be applied to model the history of any system in which (i) elements can be replicated with the incorporation of changes and (ii) any change between a progeny element and its parent is stably transmitted in subsequent generations. (C. J. Howe and Windram 2011)

Following the convention of referring to elements of cultural transmission as “memes”, Howe and Windram propose the term “phylomemetics” to describe cultural applications of phylogenetic analysis. While the phylomemetic approach in stemmatology has been covered elsewhere in this volume, this section will discuss the development of phylomemetics in other disciplines, focusing in particular on anthropology.

8.3.1 Historical precedents

As in stemmatology, the philosophical roots of phylomemetics run deep in anthropology, linguistics, folklore, and other disciplines. Indeed, the fundamental idea that “descent with modification” is a general evolutionary process that applies as

much to cultural diversity as the natural world is explicit in Darwin's own works. In *The Descent of Man*, Darwin suggests that "the formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously the same" (Darwin 1871, 1:59).

This idea was taken up – or possibly even anticipated – by August Schleicher (1861), the founder of modern historical linguistics. Schleicher hypothesised that relationships among the Indo-European languages could be directly modelled with the kind of tree diagrams used by Darwin to depict the phylogeny of biological species (fig. 8.3-1). Thus, he suggested that they were all derived from a single common ancestral language that gradually differentiated into separate branches such as Romance, Germanic, and so on.

Many contemporaries of Darwin and Schleicher believed that the analogy between organisms and languages could be extended to other cultural domains. For instance, one of the pioneers of material culture studies in anthropology, Henry Augustus Pitt Rivers (1875, 1906), collected and organised artefacts from all over the world with the express intention of demonstrating how the principles of evolution are borne out in tools, weapons, and craft objects. As he explained,

human ideas, as represented by the various products of human industry, are capable of classification into genera, species, and varieties, in the same manner as the products of the vegetable and animal kingdoms, and in their development from the homogeneous to the heterogeneous they obey the same laws. (Pitt Rivers 1875, 307)

As with species and languages, Pitt Rivers believed that it was possible to trace the development of artefacts that were widely distributed throughout the globe to their original "root forms". For example, he argued that similarities among crossbows made in different societies across Europe and Asia suggested that these traditions were all derived from a single proto-crossbow. His theory was tested by Henry Balfour (1889), who produced the first phylogeny of a material culture tradition that was explicitly based on the branching family-tree models employed by biologists and historical linguists (fig. 8.3-1).

In folklore studies, there is similarly a long and rich tradition of evolutionary theory that stretches from Julius Krohn, who founded the historio-geographical school of comparative folklore in the nineteenth century, through to the work of Carl von Sydow and Stith Thompson (1977) in the mid- and late twentieth century. These writers believed that folk tales shared among different cultures could be traced back to an original "archetype" tale, developing into locally distinct forms in each location as they adapted to culturally specific norms and preferences. Such theories often drew directly on biological models – as encapsulated by Stith Thompson's observation that "biologists have long since labelled their flora and fauna by a universal system", and that "the need for such an arrangement of narrative has been realized for a long time" (1977, 413, 414).

In all these fields, reconstructing lineages of cultural "descent with modification" presents the same kind of methodological challenges confronted by biologists.

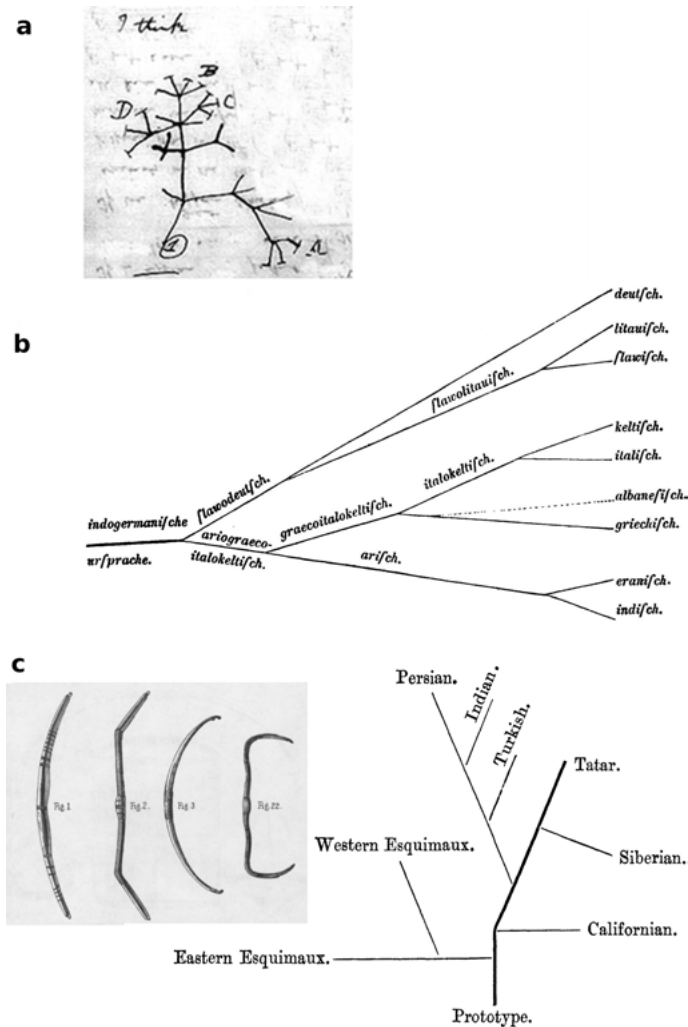


Fig. 8.3-1: Branching lineages drawn (a) by Darwin (1837, 36) for species, (b) by Schleicher (1861, 1:7; discussed in 8.2 above) for Indo-European languages, and (c) by Balfour (1889, 244) for crossbows.

Foremost among them is the problem of missing links, that is, the lack of direct physical evidence to trace relations of common ancestry back through time. Just as the fossil record bears witness to a tiny percentage of extinct ancestral species, data on ancient languages, material culture, and narratives is extremely scarce. The origins of most language families predate the invention of writing, and the archaeological record is very patchy and subject to numerous biases (not least the poor preservation of most material), while folk narratives have mainly been transmitted orally and rarely written down. Given these limitations, the characteristics of miss-

ing links can only be inferred from the traits exhibited by their (presumed) descendants. This brings us to another problem that anthropologists, linguists, and folklorists share with biologists, which is how to distinguish true family resemblances that are the result of hereditary transmission from similarities that arise through other kinds of processes. For example, similar morphological adaptations can be observed in species that almost certainly evolved independently (e.g. wings in bats and birds), just as similar cultural traits can arise in completely unrelated cultural contexts (e.g. writing in the Middle East, China, and Meso-America; or pyramids in ancient Egypt and Mexico). In evolutionary terms, this process is known as *convergence*. Another confounding factor is the transmission of traits across separate lineages, sometimes known as lateral transfer or horizontal transmission in biology and anthropology, or contamination in stemmatology. Horizontal transmission results in similarities among taxa that are only distantly related to one another. This can occur in many biological species, particularly in plants and microbes, and is likely to be even more common in cultural evolution, where trade and other forms of exchange can potentially lead to the widespread borrowing and blending of cultural traits (see e.g. Terrell 1988; Moore 1994).

In the past, anthropologists, linguists, and others had to rely on their own (often highly subjective) judgements to try and solve the problems of missing links and family resemblances. Below, we will see how phylomemetics provides a more rigorous and systematic means to address them.

8.3.2 Phylomemetic analysis

The first step in a phylogenetic/phylomemetic analysis is to define a set of characters, that is, the basic units of information that are transmitted from ancestral taxa to descendant taxa. In biology, these may be gene sequences or morphological traits. In linguistics, they may be lexical items or syntactic features. In material culture, they may be specific designs, craft techniques, and so on. In the case of traditional narratives, phylogenetic characters can be derived from narrative motifs – characters, objects, and episodes that withstand repeated transmission and which folklorists use to identify related versions of the same story (S. Thompson 1977). Once the characters have been defined, the state of every character in each variant of a tale is recorded in a matrix. Characters can be coded as either binary characters that take only two states – 0 (absent) or 1 (present) – or as multistate characters that take a variety of expressions.

While there are numerous types of phylomemetic analysis, here I will focus on two of the most widely used in anthropology and linguistics, one of which is fairly well established in stemmatology, and one of which is not (but in which, in the spirit of cross-fertilisation between disciplines, this section hopes to encourage greater interest). They are, respectively, cladistic analysis and Bayesian inference.

Cladistic analysis clusters entities (e.g. species, languages, or variants of a tale) into hierarchically nested branches known as clades. A clade represents a group of entities that share evolutionarily novel traits (known as derived character states) inherited from an exclusive common ancestor (and which are therefore lacking in other, more distant relatives). If the evolution of a cultural tradition conformed exactly to the accumulation of innovations within branching lineages of descent, the task of sorting the variants into clades would be straightforward. However, for the reasons explained above, evolution is rarely straightforward, and not all similarities among entities are true family resemblances (known in phylogenetic jargon as homologies). In most cases, we would expect the neat hierarchical pattern of inheritance to be disrupted by independent evolution and/or borrowing (similarities due to these processes are classed as homoplasies). Cladistic analysis deals with conflicting patterns in a dataset by searching for the tree (or trees) that minimise the number of evolutionary changes that are required to explain shared character states among the taxa. This approach invokes the philosophical principle of parsimony, which states that scientific explanations should never be more complicated than necessary.

Although cladistic analysis is logically appealing, there are some important limitations to the approach. In particular, the assumption of parsimony may be overly simplistic in many cases, such as when there is considerable variance in rates of evolution in different traits and/or different lineages. For example, folklorists (e.g. S. Thompson 1977) have suggested that motifs related to events in a story are more stable than motifs related to characterisation (e.g. the gender or species of the protagonists), and that storytellers are more likely to alter the beginning and end of a tale than the core middle section of a narrative. In these instances, a less parsimonious reconstruction that allows some motifs (e.g. characterisation, episodes that occur in the beginning or end of a story) to switch between states more freely might be more accurate than one which minimises the overall number of character changes.

For these reasons, cladistic methods have in recent years been largely superseded in anthropological phylomemetics by an alternative approach, known as Bayesian phylogenetic inference (P. O. Lewis 2001), which is better able to deal with these issues (A. M. Wright and Hillis 2014; see fig. 8.3-2 below). Bayesian inference proceeds by calculating the likelihood of the data (i.e. the chance of obtaining the observed distribution of character states) given an initial, randomly chosen tree topology, a set of branch lengths, and a model of character evolution (i.e. the substitution rates for character states). The state of each parameter is then modified (i.e. clades are re-sorted, branches get lengthened/shortened, variance in rates of character change is increased/decreased), and the likelihood of the data is recalculated. This process is then repeated hundreds of thousands of times using a Markov Chain Monte Carlo (MCMC) algorithm. In this analysis, moves that improve the likelihood of the data are always accepted, while those that do not are usually rejected – although some may occasionally be accepted within a certain threshold. This is

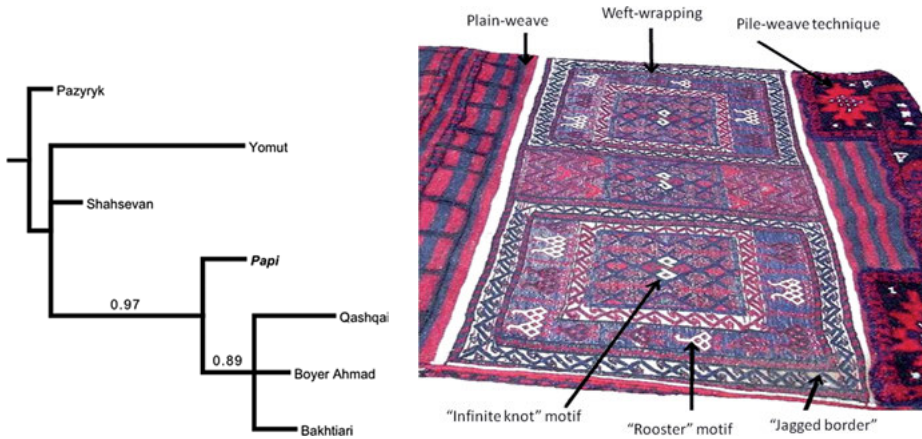


Fig. 8.3-2: Example of a Bayesian consensus tree based on analyses of Iranian textile traits (such as the example illustrated on the right). The values beside the nodes represent posterior probabilities for the corresponding clades. Source: Matthews et al. (2011).

because the search for the best parameter states (i.e. the most likely trees) is similar to walking through a mountain range to find the highest peak: sometimes one needs to go down in order to get higher up. However, in this case, the ground is constantly shifting below one's feet as each parameter gets adjusted simultaneously. Thus, the tree that seems to be the best one (i.e. that maximises the likelihood of the distribution of character states) under one set of conditions may turn out to be suboptimal when the branch lengths or variance in rates of character change are adjusted slightly (while, at the same time, the best values for these parameters will vary with different tree topologies). Bayesian phylogenetic inference integrates the uncertainty associated with alternative evolutionary scenarios by sampling trees at regular intervals in the MCMC chain to compile a posterior distribution of trees. Since the analysis usually favours moves that increase the likelihood of the data, it revisits higher peaks in the likelihood landscape more frequently than lower peaks, meaning that trees with higher probabilities get sampled more often than ones with lower probabilities.

Once the posterior distribution of trees has been compiled, phylogenetic relationships among tale variants can be represented by a consensus tree that shows the posterior probabilities of individual clades, which correspond to the percentage of posterior trees in which they occurred (fig. 8.3-2). The latter provides a useful indication of the robustness of these relationships under a range of plausible evolutionary models, rather than just a single optimality criterion such as parsimony. Moreover, since rates of change and branch lengths are also explicitly modelled, Bayesian inference has other useful features besides estimating phylogenetic relationships. For example, if the analysis includes taxa sampled from different historical periods (e.g. ancient languages in a linguistic analysis, archaeological artefacts

in a material culture analysis, or early literary versions of folk tales in a folklore analysis), they can serve as reference points for calibrating an evolutionary clock which calculates the average number of character-state changes that would be expected over fixed intervals of time (e.g. centuries, millennia). The clock model can then be used to estimate the root age of the tradition under study, as well as all the internal nodes (ancestors) postulated in the tree.

8.3.3 Current applications of phylomemetics in anthropology

Cladistic and Bayesian methods provide modern-day academic descendants of Schleicher, Pitt Rivers, and Krohn with a powerful set of tools for testing long-standing hypotheses concerning the origins and taxonomic relationships among various cultural traditions, from prehistoric stone tools (e.g. O'Brien, Darwent, and Lyman 2001; Lycett 2009) and languages to textiles (Tehrani and Collard 2002, 2013; Matthews et al. 2011; Buckley and Boudot 2017), musical instruments (Tëmkin 2004), and folk tales (Tehrani 2013; Tehrani, Nguyen, and Roos 2016). For example, Tehrani (2013) investigated whether traditional classifications of similar folk tales from different cultures can be classified into distinct international types based on common origins. He carried out a cross-cultural phylomemetic analysis of a famous but controversial tale: *Little Red Riding Hood*. Tehrani's results established that, while European versions of *Little Red Riding Hood* form a phylogenetically distinct group, versions from Africa are actually more closely related to another international folk-tale type, *The Wolf and the Kids*, while East Asian versions are a hybrid of motifs from both types. In a follow-up study (Tehrani, Nguyen, and Roos 2016), Tehrani and colleagues used cladistics, Bayesian analysis, and phylogenetic networks to explore the origins of *Little Red Riding Hood* in Europe. Their analyses found strong evidence that the literary tradition, which dates back to the seventeenth century, is descended from an older oral tradition that probably goes back to at least mediaeval times and not to a literary tradition, as was assumed by some (e.g. Husing 1989).

In addition to reconstructing origins of cultural traditions, applications of phylomemetics in anthropology have also been used to draw wider inferences about population histories. For example, Gray, Drummond, and Greenhill (2009) carried out a Bayesian phylomemetic analysis of Austronesian languages to test the two major hypotheses about the peopling of the Pacific. The “pulse–pause” hypothesis suggests that Austronesians spread from their ancestral homeland in Taiwan in a series of rapid expansions interspersed with longer, more settled periods (see 8.2.3). The “slow boat” hypothesis proposes instead a more gradual and consistent process of diffusion from Wallacea. Gray, Drummond, and Greenhill (2009) found that the pattern and tempo of lexical evolution in their Bayesian Austronesian language trees was more consistent with the “pulse–pause” model than the “slow boat” model. Similar kinds of study have been carried out with material culture data, for

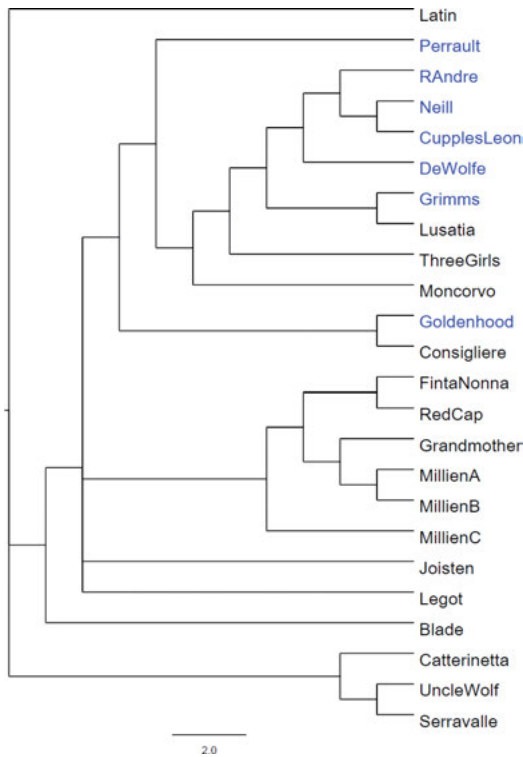


Fig. 8.3-3: Maximum parsimony tree of *Little Red Riding Hood* tales. Oral versions are labelled in black, literary versions in blue. The results suggest that the literary tradition is rooted in the oral one. Source: Tehrani, Nguyen, and Roos (2016, 615).

instance by using phylogenies of early stone tool assemblages to infer patterns of human migration out of Africa (Lycett 2009) and into the Americas (Buchanan and Collard 2007).

Finally, phylomemetic analysis has been used to address more general theoretical debates in anthropology about the genesis and diversification of cultures. Whereas Darwin, Schleicher, Pitt Rivers, and so on believed cultural evolution and biological evolution followed fundamentally similar principles, in the twentieth century many anthropologists came to reject this view. Among the most influential of them was Franz Boas, who argued:

Animal forms develop in divergent directions, and an intermingling of species that have once become distinct is negligible in the whole developmental history. It is otherwise in the domain of culture. Human thoughts, institutions, activities may spread from one social unit to another. As soon as two groups come into close contact their cultural traits will be disseminated from one to the other. (Boas 1940, 251)

To test these claims, a number of studies have used phylomemetic techniques to assess the roles of branching vs blending evolution in the generation of cultural

traditions (e.g. Tehrani and Collard 2002, 2013; Jordan and Shennan 2003; Collard, Shennan, and Tehrani 2006; Cochrane and Lipo 2010; Gray, Bryant, and Greenhill 2010). This involves estimating how well patterns of diversity in languages and material culture assemblages fit a tree model, that is, measuring the extent to which the distributions of shared character states among a set of cultural entities are homologous (similarities that are due to common descent, which are compatible with a tree-like model of descent) or homoplastic (similarities with are due to processes other than descent, and are thus incompatible with a tree model). These studies show that cultural evolutionary processes are strongly shaped by ethnic boundaries, patterns of conflict and marriage, migration rates, linguistic and ecological barriers, as well as a number of other factors (Tehrani and Collard 2013). Overall, however, these studies suggest that branching descent with modification plays a vitally important role in generating cultural diversity, and may even be as significant as it is in biological evolution (Collard, Shennan, and Tehrani 2006). These findings go some way to justifying Howe and Windram's claim that the "process of replication with the incorporation of changes is a fundamental one in human cultural activity" (C. J. Howe and Windram 2011), and highlight the wide scope and potential of phylogenetics in the humanities and social sciences.

8.4 Musicology

Cristina Urchueguía

The first striking fact about the relationship between stemmatology and musicology is how late the two disciplines met. First encounters cannot be traced before the end of the 1960s; eventually, musicologists became aware of stemmatological methods, discussed them, and tried using them for different purposes. From the beginning, "filiation" was the most popular term for referring to this philological discourse among German-speaking and Anglo-American musical editors and musicological scholars. Following an enthusiastic reception of stemmatology applied to music from the twelfth to seventeenth centuries, the 1970s and 1980s critically examined its capacity to find the "original" text, but this resulted in a certain degree of disenchantment. Today, the use of stemmatological methods continues to be taught and discussed, but its functionality has been reformulated. Editors of music from the eighteenth to twentieth centuries seldom adopt stemmatology as the principal method; instead, various versions of a best-text method are the norm. In this context, it was the very problematic term "urtext-edition" that functioned as catalyst in the general discussion.

The second specific element concerning this relationship is the attitude of some – sometimes even overt – scepticism that musicologists have shown towards stemmatology. The fact that musicology itself was accepted as an academic disci-

pline rather late is one, but not the most important, reason for this attitude. It is true that musicology was not established in academia before the end of the nineteenth century, but musical editing based on some kind of rational principles can be traced back to as early as the beginning of the nineteenth century. The most important fields of editorial work on historical texts were the recuperation of vocal polyphony from the sixteenth century and the collected works of Johann Sebastian Bach, both begun in the middle of the eighteenth century. Even before the establishment of the discipline, there had been an awareness of the importance of defining methodological frameworks for editorial tasks.

The main objection scholarly music editors have raised against stemmatology is based on the conviction that the transmission of musical texts followed patterns completely different in nature from the transmission of literary texts, even in special cases such as historical chronicles (which were also often updated and recontextualised throughout the centuries). These differences have their roots in the structure of the musical works, their hybrid semiotic composition and the historical development of musical notation, the methods of compiling sources, and, last but not least, the inherently performative essence of music as a human activity. The methods developed for letter-based texts had to undergo a thorough adaptation in order to cope with the semiotic environment of music. Orality and performativity challenged musicological stemmatology from the beginning. The second reason for the misgivings of musicologists regarding stemmatology is the lack of relevant musical sources dating from before the ninth century. The few extant fragments of ancient Greek notation (Pöhlmann and West 2001) do not constitute a corpus comparable in size and relevance to the corpus of texts that were targeted by classical and biblical philologists to develop their methods and prestige. What musical scholars could infer from the scattered, fragmentary bits of clay or stone with ancient notation was so primitive in style and structure that any comparison with classical literature was pointless. These fragments were not perceived as being paradigmatic for musical repertoires in any sense, and had absolutely no impact on the development of musicological stemmatology. Thus, musicology could not rely on the arguments and goals that classical and modern philological disciplines had adopted to establish their editorial principles and textual traditions.

8.4.1 Stemmatology for music editors

Notwithstanding the common perception of music as a phenomenon of an aural nature, scholarly editorial work on music did not initially determine the sonic reconstruction of music as one of its main tasks. On the contrary, the basic material used by editors was the written text, and their first objective the constitution of an edited text. Nonetheless, performance practice and orality were not completely excluded from the critical discussion about the principles to be used in the constitu-

tion of musical editions when it came to *recensio* of the transmitted text; in fact, they became one of the main obstacles to an implementation of stemmatology in analogy to classical philology. The tension between music's various forms of medial representation and their contradictory implications for editions constituted a substantial methodological issue (Bent 1990, 1995).

The notation of Western music began as neumatic notation (see fig. 8.4-1), which served as a political tool in order to document and canonise Gregorian chant and further the hegemonic ambitions of the new rulers during the Carolingian period. Although descriptions of ancient Greek notation were rediscovered at the same time, the neumatic notation of the Middle Ages was not derived from the ancient principles but was created from scratch. This rupture in the written tradition of music is essential to understanding some of the characteristics of musical editing mentioned above. Musical philology was a latecomer within the philological family, lacking roots in a humanistic tradition of its own.

What we call Western music – and this has been almost the only style considered relevant to scholarly musical editions – was a changing, fashionable musical repertoire used first in church and court between the middle of the seventeenth and the end of the nineteenth century that entered the concert repertoire from the beginning of the nineteenth century. In this timeframe, musical styles changed with amazing speed: until the beginning of the nineteenth century, Catholic church music still used the corpus of Gregorian chant together with modern styles, whereas court musicians were always seeking fashionable novelties. A consciousness of the importance of the music of the preceding generation, and even of earlier periods, did not arise until the nineteenth century, together with the awareness of national cultural values. This is the moment this music became an object of scholarly and performative interest. The present contribution is not the place to outline these discourses in detail, but the essential facts are (i) the lack of humanistic roots for music editing, and (ii) the contextualisation of music editing within modern national states and the construction of their cultural identities. The most prominent examples of these historical origins are the importance of the edition of Johann Sebastian Bach's works for Germany, and the focus on Franco-Flemish polyphony for the Netherlands. It is worth mentioning that musicological stemmatology usually ignored the music of the Middle Ages. At the end of the nineteenth and the beginning of the twentieth century, it was rather the idea of the infinite multiplicity of versions due to the performative situation that inspired editors like Pierre Aubry in his collaboration with Bédier (see 2.3.4.1). In their edition of the *Chansons de Croisade* (1909), they reject the possibility of reconstructing an original melody from different versions, and thus also the applicability of textual criticism to music:

La méthode critique, suivie par les éditeurs de textes littéraires, ne nous semble point d'une application possible, en raison de la différence qui existe entre la matière philologique, où tout est stable, et la matière musicale, où tout est fortuit et changeant. (Bédier 1909, xxxiii)

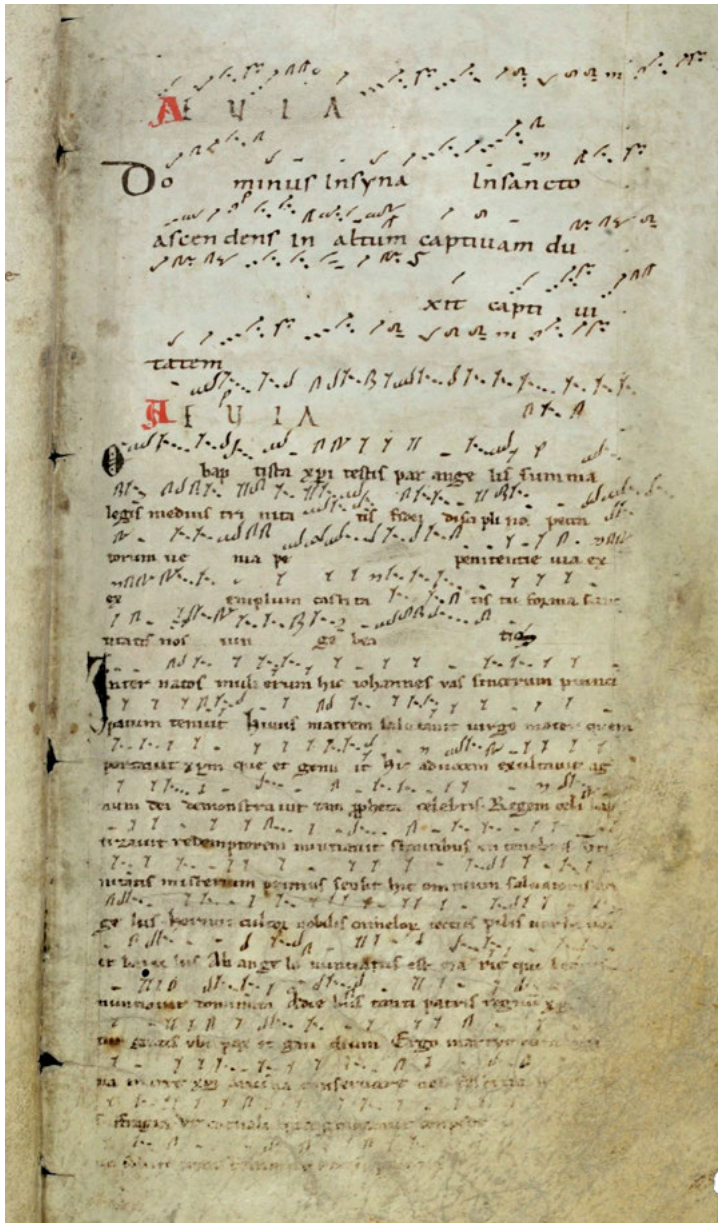


Fig. 8.4-1: Neumatic notation was used as a means to document the general shape of a melody, but it lacks information about the exact pitch, the intervals, and the rhythm. St. Gallen, Stifftsbibliothek, Cod. Sang. 359, p. 7. Source: e-codices.unifr.ch/de/searchresult/list/one/csg/0359. Image: CC-BY-NC.

[The text-critical method, pursued by the editors of literary texts, does not at all seem applicable to us, because of the difference that exists between the philological material, where all is stable, and the musical material, where all is casual and changing.]

The editorial expression of this view is a synoptic reproduction of variant melodies. This radical embrace of variation was abandoned in the case other repertoires that could be used to support nationalistic arguments.

The complete edition of Bach's *œuvre* (1851–1899) initiated the series of large-scale editions of complete works that constitute the paramount achievements of German-language musicology (von Dadelsen 1967; Georgiades 1971; Bannwitz et al. 1975; B. R. Appel and Veit 2000; Emans and Krämer 2015). Focusing on repertoires from the late seventeenth to the beginning of the twentieth century – with only few exceptions predating this period – editors often had access to autographs and authorial witnesses that made any stemmatological analysis of their transmitted texts secondary or even unnecessary. The most common approach was (and continues to be) a case-by-case best-text method, aiming at producing a version intended by the authors, authorised by them, or legitimated by witnesses of reception sanctioned by the author. This method first reached prominence under the label “urtext” at the end of the nineteenth century, becoming the only editorial concept to transcend the scholarly realm and become popularised as a marketing feature (Feder and Unverricht 1959; Dahlhaus 1973; Badura-Skoda 1986). Today, however, scholarly music editors have questioned this highly problematic concept in response to the charges of eclecticism made against it. Strikingly, the most recent handbook on the topic issued in Germany – *Musikphilologie: Grundlagen, Methoden, Praxis* (B. R. Appel and Emans 2017) – does not address the concept of the urtext, or even include the term itself in the glossary.

Stemmatology was, thus, not the first choice in this editorial context, although editors did take note of it and discussed its potential capacity, marginalising it as a tool supplementary to their day-to-day business (Feder 1983; Wade 1988; Strohm 1995). The above-mentioned handbook deals with stemmatology in a rather shallow and pro forma subchapter of no more than five pages (Scheideler 2017).

The natural nurturing ground for stemmatology was the musicological and editorial work on early music, beginning with Franco-Flemish polyphony. Josquin des Prez became the paradigmatic composer because of the richness of his output, the complexity of its transmission patterns, and the fact that the transmission of his works crossed the threshold between manuscript and print (van Benthem 1969–1970; Porter 1976; Blackburn 1976; Hoffmann-Erbrecht 1976; Just 1983; Frobenius 2001). Thus, Josquin's *opera omnia* became a privileged object of study that combined the history of music and the history of music printing. The *New Josquin Edition*, a collaborative international project realised by the most prominent scholars in the field, set path-breaking standards based on filiation (Josquin 1987–2016). The use of stemmatic methods caused a real paradigm change in the perception of the music of Josquin's lifetime, as it provided evidence about the pre-eminence of some

manuscript copies over printed copies, thus turning the scholarly consensus on its head. The bibliographical analysis of early music printing was another fruitful field for stemmatology in a less text-oriented medium (Boorman 1977, 1981b; Noblitt 1981; Hamm 1983; Staehelin 1998; Drake 1999; Mouser 2003, 2004). The scholarly debate about stemmatology largely focused on these two interrelated fields, which were considered alongside one another in special journal issues devoted to the music of Josquin and his time that established the state of the art as we find it in handbooks and monograph publications on musical editing today (Noblitt 1983; Just 1983; Noblitt 1995; for handbooks, cf. Caraci Vela 1995, 2005–2013; Grier 1996).

Although stemmatology proved to be helpful in this context, its results were not conclusive. The emphasis of scholarship has consequently shifted to the transmission process, scribal habits, cultures of compilation, and musical style. Taking into account the hypothesis that scribes of music themselves had to be proficient musicians, if not composers, led to a radical questioning of the validity of the written text with respect to the entity to be edited, the “music itself”. It is needless to point out the lack of consensus about the ultimate aim of music editors, ranging from a religious attitude to the extant notation to the search for a modern transcription in order to represent a hypothetical, lost sound. Musicological editors have always faced diachronic changes in the way notation represents performance practice and historical sound, assuming the role of translators between trained musicological scholars and musicians used to deciphering only conventional modern notation. As the specificities of historical musical notation, scribal competence, and the perception of mutual influence between sound and writing became a central matter of scrutiny and debate, the written text itself began to be perceived as an early stage within the editorial process and not as the final goal. Transcriptions of historical notation into the editor’s standard notation imply a high degree of speculation about the difference between the capabilities of the original and target notations to represent performance habits and sound quality. Stemmatology failed to fill the gap between the visible notation and the hypotheses about the represented sound, and was thus subject to criticism or rejection by musicologists.

From an aural point of view, editorial work on music of the past meant groping in the dark because the dynamics of the evolution of notation, historical instruments, and performance practice had veiled these elements of the musical work in all but absolute obscurity. Editors thus focused on the possibilities textual criticism offered for establishing an authentic written representation of the musical entity.

Even among the group of editors devoted to early music, the enthusiasm of having discovered stemmatology as a powerful tool to cope with the written text (Grier 1995, 1996) soon subsided, and a certain disillusionment took hold (Boorman 1981a, 1981c; Bent 1981; Brett 1988; Grier 1995; Boorman 1995). The *locus classicus* summarising the standard caveats is Stanley Boorman’s unsurpassed discussions about filiation for musical editors (Boorman 1981a, 1981c). He enumerates three aspects characteristic of music that together constitute the main obstacles for an adap-

tation to music of the practice of stemmatology as developed by classical philology (Boorman 1981a, 320–321): (i) “the small size of most individual pieces [...], resulting in relatively little evidence”, (ii) “the fact that most surviving sources were probably copied from more than one exemplar”, and (iii) “the possible confusion between editorial and accidental changes to both substantive and non-substantive elements of the text”. These problems may, sporadically and occasionally, appear also in some sources transmitting literature or other texts, and trouble their editors, but in musical texts they constitute the rule, not the exception, and they always appear together. Boorman’s first issue concerns quantity, while the others have methodological and semiotic implications. Last but not least, musical works of large-scale dimensions that present a rich layer of variants, such as the Romantic symphonies, normally do not require a stemmatological approach because editors can use authorial witnesses.

Boorman’s rather pessimistic affirmation that the “pursuit of an ‘authentic’ text for music is almost that of seeking the chimera, and it is also [...] essentially irrelevant” (Boorman 1981b, 168) should be contextualised within this semiotic frame. Music editors have not only been concerned about the period notation they edit but also about the sound these signs inspire when read by non-contemporary musicians who lack the knowledge and the training to read them properly. Reconstructing the original notation does not guarantee that musicians will understand the original musical text, and even when it comes to the reconstruction of a written text, stemmatology can offer only partial help.

8.4.2 Stemmatology for musicologists

This does not mean that stemmatology is useless. Musicologists have used stemmatological methods in various contexts: notated music, literary texts (mainly the lyrics of musical works), theoretical treatises (Bernhard 1979; Solomon 1983, 1986), the relationship of different versions of a work (see fig. 8.4-2), and even reconstructing the transmission of iconographic elements in music-related sources (Teviotdale 1988). Yet these cases do not require methods that differ from those applied by an editor of language-based material. Moreover, the most uncontested and fruitful contributions of stemmatology to musicology have occurred when scholars focused on a by-product of the process, namely its capacity to trace patterns of transmission within musical repertoires and unique, complex sources. If the combinatorial, or, to use the terminology of philology, the contaminated nature of musical sources was an obstacle to the reconstruction of every single piece in a compilation, it proved to be a blessing for mapping the cultural context and origin of the collection as a distinct entity.

Many repertoires have been studied using stemmatological methods: mediaeval plainchant repertoires, Gregorian chant, Aquitanian and French tropes and versaria

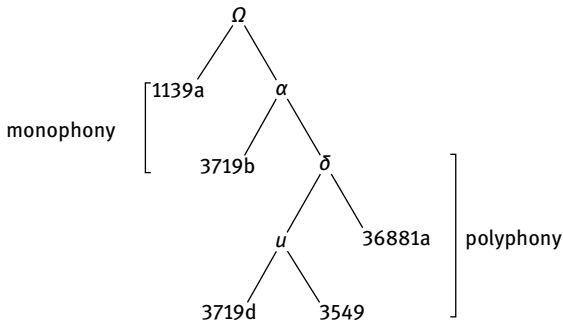


Fig. 8.4-2: This stemma shows the relationship between a monophonic and a polyphonic transmission of the same melody, thus conveying the emergence of polyphony from monophonic repertoires. Source: Grier (1995, 82).

(G. Weiss 1964; Hughes 1969; Grier 1988; Levy 2003; Procter 2006); vernacular monophony of the *trouvère*, *troubadour*, *cantiga*, and *Minnesang* traditions (Karp 1964; van der Werf 1972; Räkel 1973; Schaffer 2000; Wulstan 2000; Räkel 1977); and polyphony of the thirteenth to sixteenth centuries (Atlas 1977; Cook 1978; Atlas 1981; Bent 1981; Frobenius 1987; d'Alvarenga 2011). Even the very complex transmission of instrumental music has been tackled with stemmatological zeal (Memelsdorff 2010). Josquin was not the only composer whose works were scrutinised by stemmatological studies: the scholarly work on Machaut (Dömling 1969), and the editions of Jacob Obrecht (Hudson 1988) and other composers of the fifteenth and sixteenth centuries, included this method in its toolboxes but insisted upon its limitations rather than its accomplishments.

Although stemmatology did not achieve the results editors had envisaged as a result of the method's importance in other disciplines, its combination with theories about the transmission and compilation of music (Hamm 1962; Just 1981) in the period between the fourteenth and sixteenth centuries enabled stemmatology to contribute substantially to reaching new and promising insights. This was achieved by linking repertoires and compilation strategies to political relationships and knowledge of cultural exchange. This approach rendered the seemingly random content of sources historically consistent and plausible. A pattern of research was established by the pioneering work of Allan Atlas on the Capella Giulia *Chansonier* (Atlas 1975–1976), a composite Renaissance source. This edition did not use stemmatology as a means of constituting an archetypal text; instead, assembling the scattered stemmatological evidence, Atlas succeeded in tracing the origin of the different chansons and the cultural contexts and intercultural relationships that made the formation of this collection possible. Atlas's approach inspired other colleagues, and a series of single-source studies appeared in the following decades (Just 1991b; Ros-Fabrégas 1992; Noblitt 1987–1996), all of which focus on the history of the manuscripts involved and the fascinating entanglement between Renaissance sources.

Other musicologists used stemmatological approaches to face problems that had been troubling the discipline for a long time, such as the phenomenon of reworking and contrafacture (Blackburn 1976; Schmid 1999), style (Kirsch 1981; Bent 1987), and scribal habits (Hortschansky 1981; Boorman 1983; H. M. Brown 1983; Bent 1990). Stemmatology featured as one of many tools, but its basic effect was to establish a methodological link with textual studies in other disciplines, thus preventing the complete detachment of musicological speculation from the written evidence. Another important achievement of stemmatology concerns its use as a method to establish hard criteria for the assessment of authenticity, authorial intentionality, and the textual status of musical transmission (Staehelin 1983; Brett 1988; Just 1991a). Having abandoned the aspiration of creating an archetype, the assessment of authenticity still served as a means of containment for pure speculation.

8.4.3 Stemmatology for performing artists

The importance of stemmatology for performing artists is twofold. On the one hand, performers often express a sense of belonging to a genealogy of preceding musicians. Their musical training relies heavily on a “master–disciple” relationship and oral transmission, due to the fact that many aspects of musical interpretation have not been or cannot be recorded as written instructions. Considering oneself the “grandson” of Franz Liszt is not only a metaphor, but also means having inherited a style, a pedigree, a responsibility, and a sense of mission. Yet this understanding of a stemma lies beyond the scope of the present volume.

Central to our concern is the preoccupation of musicologists with rendering the findings of conventional stemmatology fruitful for performing musicians (Boorman 1995; Göller and Mazzola 2002). The crucial element in these considerations is that performing artists should also focus on those elements of the notation which are unprofitable or useless to the music editor. While the differentiation between substantial and accidental aspects of notation has always been suspicious to editors because editorial practices obliged them to make decisions, musicians can use the ephemeral space of performance as a place for experimentation. This might be considered one methodological supplement that musicology has contributed to the stemmatological tradition.

8.4.4 Stemmatology for music ethnographers

Music ethnography, a discipline that frequently deals with orally transmitted music, has interpreted stemmatology in a completely different way than historical musicology. I mention here only two related lines of research in which stemmatology, understood in a broader sense, can be found as a methodological tool. The first is

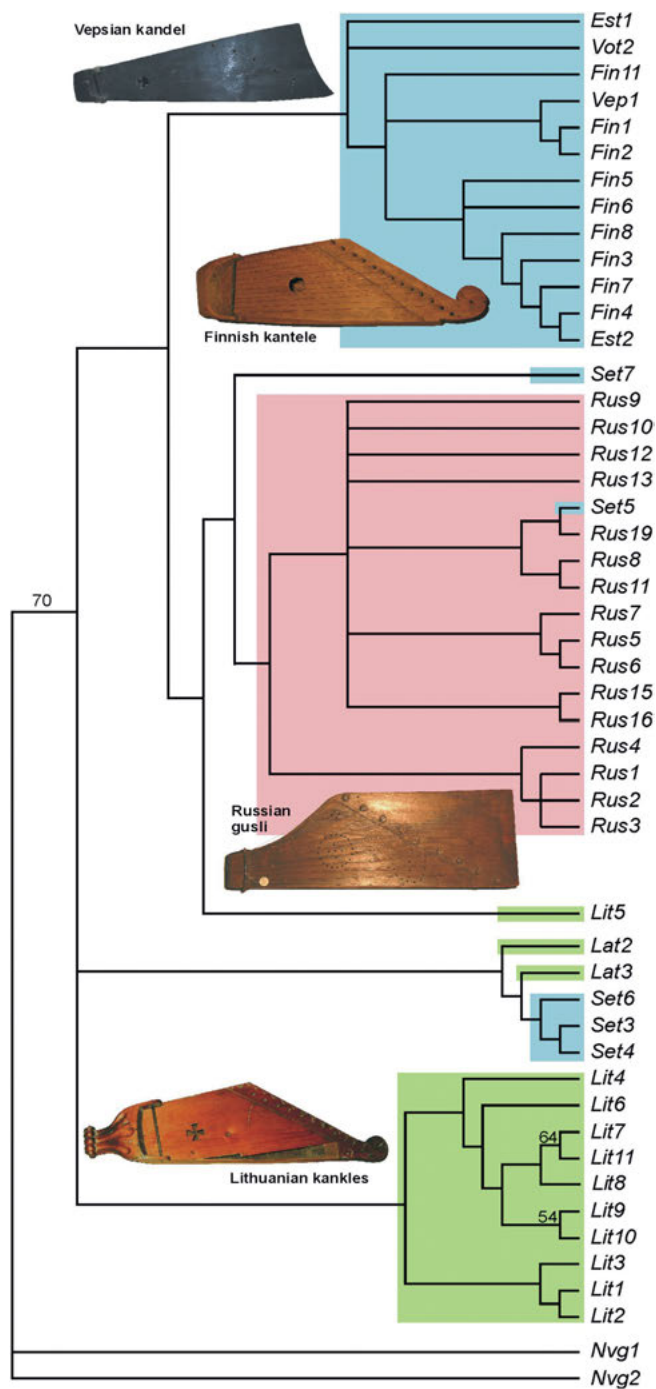


Fig. 8.4-3: Cladogram of the Baltic psalter. Majority-rule consensus tree.
Source: Tëmkin and Eldredge (2007, 147).

the structural analysis of folkloric melodies (Steingrímsson 1975; Zamfir 1972). The hypothesis behind this method is that a structural similarity between melodies mirrors their ethnological relationship. This hypothesis has been strengthened in recent decades by biological research. Some studies have found correlations between DNA and the structural analysis of folkloric music, its structure, and its instruments (Tëmkin and Eldredge 2007; S. Brown et al. 2014); see figure 8.4-3. Moreover, some results establish a parallelism between the development of musical patterns and phylogenetics, claiming that musical patterns seem to match DNA more closely than linguistic relationships.