

# OF WORD FAMILIES AND LANGUAGE TREES: NEW AND OLD METAPHORS IN STUDIES ON LANGUAGE HISTORY

HANS GEISLER · JOHANN-MATTIS LIST

**ABSTRACT** · For a long time, metaphors have played an important role in depicting language history. In this study, we contrast early metaphors on language history, such as the family tree or the wave model, with recent metaphors that were popularized after the quantitative turn, such as forests of trees or phylogenetic networks. Speculating about metaphors which could become important in the future, we conclude that a vivid discussion about the usefulness and the concrete implications of metaphors plays a key role for the development of models for language history in historical linguistics.

**KEYWORDS** · Language History, Language Evolution, History of Linguistics, Metaphors, Historical Linguistics.

## 1. INTRODUCTION

**M**ETAPHORS are crucial for scientific research, specifically when it comes to describing new phenomena which have not been observed or studied before. Choosing good metaphors may not only decide about the acceptance of a new theory in a scientific community, but also about the way in which the theory is explored by its followers.

An interesting example for the importance of metaphors for scientific research is the field of historical linguistics. Looking back at a century-long history in which the phenomena of language change and language history were investigated and discussed from various intra- and interdisciplinary perspectives, the history of historical linguistics gives us interesting insights into the role which metaphors play in the development of scientific theories. Language history, today also often called *language evolution*, can be understood as the phenomenon by which languages diverge and split into independent lineages over time as a result of independent change. While the phenomenon was detected and described rather early in the history of the language sciences – with first discussions in Europe reaching back to the 17th century – it was only in the 19th century that scholars began to agree that language change reflected the norm rather than the exception and that «the languages change as long as they exist», as August Schleicher wrote in 1863 (Schleicher, 1863, 18f, our translation, original text: «[...] die sprachen sich ändern solange sie leben»).

In this short review, we will first discuss early metaphors on language history and point to the tension which arose from the problem that early family tree models were neither able to capture convergence nor lateral transfer (§ 2). We

geisleh@hhu.de, Heinrich-Heine-Universität Düsseldorf, Deutschland.

mattis\_list@eva.mpg.de, Max Planck Institute for Evolutionary Anthropology, Leipzig, Deutschland.

[HTTPS://DOI.ORG/10.19272/202201902005](https://doi.org/10.19272/202201902005) · «MODERNA» · XXIV · 1-2 · 2022

[HTTP://MODERNA.LIBRAWEB.NET](http://moderna.libraweb.net)

RECEIVED : 13. 6. 2022 · REVISED : 20. 6. 2022 · ACCEPTED : 22. 6. 2022

will then discuss recent metaphors which evolved after the quantitative turn in the early 21st century in historical linguistics (§ 3), and speculate about metaphors that could play an important role in future research (§ 4).

## 2. EARLY METAPHORS ON LANGUAGE HISTORY

In the European context, the earliest way to describe language change was to present it as some form of decay. Scholars either assumed that languages had lost certain properties due to contact with other varieties (compare the *Kratylos* dialogue by Plato, see Eigler 2001), or they assumed that a language was losing its earlier perfection due to misuse by younger language users or novel technologies, such as writing (compare the *Phaidros* dialogue by Plato, see Eigler, 2001). This assessment persists in intellectual circles until today (see Sick, 2004 for German). The catastrophic attitude on language change can already be found in Greek and Latin philosophy and was later perpetuated and extended in the medieval ages, where language decay was justified with the myth of the Tower of Babel, according to which all languages could be ultimately traced back to Hebrew, before the Confusion of Tongues had let them drift apart (Klein, 2004, Allen, 1953, pp. 55-57). It was only later that scholars began to realize that language change is an integral, inevitable part of the history of all languages, which occurs at all times, independent of where and by whom the language is spoken.

Catastrophic views on language change made it easy for scholars to find similarities among unrelated languages. Without a coherent model of sound change that would prevent wild speculations, scholars did not have trouble finding similar words in arbitrary language pairs when comparing languages such as German with Hebrew (Münster, 1523) or Chinese with Ancient Greek (Webb, 1787). Occasionally, however, scholars began to detect more regular, more striking, and more recurring similarities when comparing languages which by now have been proven to be historically related (Campbell, Poser, 2008, pp. 18-23). Although it was not the first language family to be detected and systematically compared (*ibidem*), it was the Indo-European language family, represented by geographically distant languages like Ancient Greek, Latin, Sanskrit, and Gothic, which contributed to the broad acceptance of historical language comparison techniques among scholars of all kinds and also paved the way for evolutionary theories in other disciplines.

The detection of the Indo-European language family led to the establishment of a new methodological framework for the comparison of languages, which was no longer based on the search for spurious similarities that could be explained within a catastrophic framework of language change, but rather emphasized the importance of finding regularly recurring – or systematic – similarities, which would allow scholars to prove that languages are historically related. While this framework, nowadays known under the name «comparative method» (Meillet, 1925, Weiss 2015), was not really established until the end of the 19th century, the idea that language history proceeds in the form of diversification events that can be depicted «by the image of a branching tree» (Schleicher, 1853, pp. 787, our

translation, original text «unter dem Bilde eines sich verästelnden Baumes») was already discussed much earlier.

August Schleicher (1821-1868) was one of the leading proponents of tree thinking in linguistics. He was not the first one to invoke the family tree metaphor to model language history (first tree drawings go back to the 17th century, see Schottel, 1663) and the overviews in List, Cyson *et alii*, 2016a and Hellström, 2018. However, his early study from 1853 (Schleicher 1853), in which he presented the Indo-European language family in the form of a bulky oaktree (see FIG. 1), as well as his later work, in which he published more refined and abstract family trees (Schleicher, 1861, Schleicher, 1868), became very influential among linguists of his time and were likewise controversially discussed.

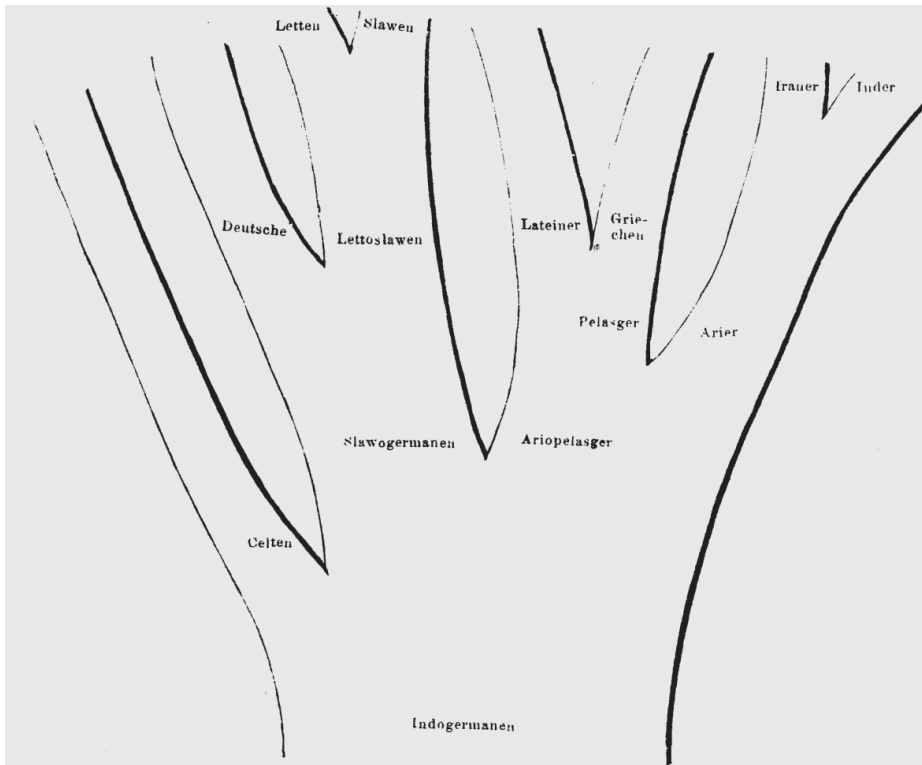


FIG. 1. August Schleicher's early family tree from 1853.

What is interesting about Schleicher's and other linguists' early tree drawings is that they reflect a rather concrete mapping of the image of a tree with the idea of branching evolution. Mathematical concepts like 'branch lengths', which show the amount of time or change that happened since separation, were not yet developed, although scholars like Schleicher began to play with the idea of historical distance in their work (as reflected also in some surprising notions of linguistic distance opposed to distance due to separation in time, see List, 2021). All in all,

early family trees in linguistics showed much more resemblance with actual trees in nature than the phylogenies scholars use today.

That a family tree only shows one aspect of language history was made clear already in Schleicher's work, where he compared cases of language contact with the hybridization of plants (Schleicher, 1863). In order to make up for this shortcoming – the incapacity of the family tree model to show convergence after divergence or to show how languages exchange material after their separation – scholars began to propose alternative models of language evolution. The most famous model that was presented to replace the family tree was the so-called «wave theory» (Wellentheorie), presented first in the work of Johannes Schmidt (1841-1901) in which he suggested to replace the image of the branching tree

[...] by the image of a wave that spreads out from the center in concentric circles becoming weaker and weaker the farther they get away from the center. (Schmidt, 1872, p. 27)<sup>1</sup>

As famous as this 'alternative' to the tree became in the following, as diverse and unclear was the concrete interpretation of the model proposed by Schmidt. This is reflected in numerous and diverse attempts to visualize the wave theory (see Geisler, List, 2013 and Jacques, List, 2019). As a result, the wave theory routinely made and makes it into text books of historical linguistics (Lehmann, 1969, Anttila, 1972), but while modern approaches to phylogenetic reconstruction are almost exclusively based on trees and networks, the wave theory is almost completely ignored when it comes to concrete examples of individual language histories (for an attempt to formalize the wave model, see Sankoff, Sankoff, 1980), and even classical scholarship tends to prefer the tree over the wave (Katičič, 1966, p. 63).

As a third early model of language history which is less frequently mentioned in modern literature, we find attempts to combine the family tree as a general model for the diversification of languages with models that present the history of individual words or linguistic units. The most common representation for such a model of language history would have a family tree showing the major patterns of diversification as a backbone, while being complemented by horizontal lines that indicate processes of borrowing or convergence due to language contact following the diversification. This model finds a very clear description by Hugo Schuchardt, who suggests that

We connect the twigs and branches of the family tree by countless horizontal lines, and it ceases to be a tree. (Schuchardt, 1900, p. 11)<sup>2</sup>

In an extreme notion, where the branching structure of the family tree serving as a backbone would be fully ignored, one would aim to model individual histories for all the words of a language variety, a model for language history, which become popular under the slogan of «chaque mot a son histoire» («every word has

<sup>1</sup> Our translation, original text: [Ich möchte an seine stelle] das bild der welle setzen, welche sich in concentrischen mit der entfernung vom mittelpunkte immer schwächer werdenden ringen ausbreitet.

<sup>2</sup> Our translation, original text: Wir verbinden die Äste und Zweige des Stammbaums durch zahllose horizontale Linien, und er hört auf ein Stammbaum zu sein.

its own history»), often attributed to Jules Gillieron (1854-1926, Campbell, 1999, p. 189), although Jacob Grimm (1819, p. xiv) had emphasized much earlier that «each word has its own history and its own life» (our translation, original text «jedes Wort hat seine Geschichte und lebt sein eigenes Leben»).

Although he described himself as a family tree sceptic, Schuchardt (1900, 21f) accepted the family tree model for those cases in which divergence would have happened without consecutive contact among the diverging varieties, thus admitting that the family tree would have its value as an idealized, albeit simplifying model of language divergence in specific conditions. As a more realistic alternative, he suggested modeling language diversification simultaneously in space and time, by showing how language varieties diverge along with the separation of their speakers with time and successively converge. For the visualization, he described some kind of video animation, in which colors indicate the similarity of language varieties. What makes Schuchardt's model remarkable is that video animations did not exist in his time, so we can only speculate how amazed he would have been if he had seen the technical possibilities we have today (even if we still lack the data to implement what Schuchardt suggested).

I want to replace the image of the family tree, which I reject, by another one. Imagine that all countries of Romance tongues appear in the same color, in white, representing the original Romance language, and that this white color then darkens, taking on different shades, which become stronger and stronger until finally the colors of the rainbow appear in front of our eyes, flowing into each other in a way that one could barely notice. This image is less simple than the other one, because it does show different moments of inspection rather than only one, but it is for this reason that it comes much closer to the original state of affairs. (*Ibid.*)

Our translation, original text: [Ich] möchte Ihnen das Bild des Stammbaums, das ich zurückweise, durch ein anderes ersetzen. Es sei der ganze Länderkomplex romanischer Zunge mit einer und derselben Farbe, mit Weiss, bedeckt, welches die allgemeine Vulgärsprache repräsentire; dieses Weiss verdunkle sich, nehme verschiedene matte Töne an, welche stärker und immer stärker hervortreten, bis endlich die Farben des Regenbogens unmerklich ineinander überfliessend vor unsern Augen stehen. Dieses Bild ist, weil es verschiedene, nicht einen einzigen Moment der Anschauung erfordert, zwar ein weniger einfaches als jenes, kommt aber eben darum dem auch keineswegs einfachen Sachverhalt näher.

Although Schuchardt's model might appear to be in opposition to family tree and wave, it is important to note that he describes the same processes, namely diversification of languages due to the geographic diversification of their speakers, accompanied by convergence due to ongoing language contact. Thus, Schuchardt's model is compatible with an extended network model that combines a base phylogeny with horizontal edges indicating convergence.

To summarize, we find two prominently competing models of language history, the family tree model and the wave theory, of which the former was constantly criticized while still being employed where needed, while the latter was praised without being actively used to model concrete instances of language history. In addition, we find a third proposal that emphasizes the 'mosaic character' of language history (see Baptiste *et alii*, 2013 for the discussion of mosaic evolution in biology), being reflected in the fact that different objects and different aspects of

language evolve in different ways which may often not coincide. This third proposal is both reflected in network models that suggest combining the family tree with horizontal lines indicating convergence, as well as models that concentrate on the individual histories of individual linguistic traits.

### 3. EVOLUTIONARY METAPHORS AFTER THE QUANTITATIVE TURN

With the beginning of the second millennium, the discipline of historical linguistics – as well as many other disciplines of the humanities – experienced some drastic changes. Funding for traditional programs devoted to the traditional investigation of the Indo-European language family had been cut in many European universities (specifically in Germany) during the 1990s, and many institutes closed or were forced to carry out a reorientation of their object of study. At the same time when linguists interested in qualitative studies of Indo-European were looking into a dark future, a new research approach, which was no longer interested in qualitative small-scale analysis but in quantitative large-scale investigations, emerged and became soon popular, both among scholars and a broader public.

At the beginning of the quantitative turn, there were increased attempts to apply methods originally designed for the application in evolutionary biology to linguistic data in order to reconstruct language phylogenies. Evolutionary biology had seen an increase in quantitative applications much earlier than linguistics (Hull, 1988), and although there was some awareness that not all processes in biological evolution strictly follow the model of a branching tree (Dagan, Martin, 2006), methods for phylogenetic tree reconstruction had been dominating the field, specifically also due to the complexity of alternative network approaches.

Apart from the practical advantages of applying biological methods for phylogenetic reconstruction to linguistic data, there was a methodological advantage, since linguists had ceased to discuss the mathematical aspects of phylogenetic tree reconstruction further, not long after some initial attempts to establish more standardized methods to measure linguistic divergence from the 1950s onwards (Swadesh, 1950, 1952, 1955, Sankoff, 1969) had been rejected or ignored by the majority of the field. Thus, while linguists had for a long time given up the idea to develop methods that would allow to estimate the divergence times of languages with the help of mathematical models (as reflected in the early attempts of Swadesh, 1950 and Lees, 1953 to establish their method of glottochronology), the biological toolbox would offer several improved methods for the dating of phylogeny divergence times, which could be directly applied to linguistic data as well.

In contrast to the early family trees drawn by linguists, biological phylogenies no longer show any resemblance with massive oak trees but have developed into detailed models of divergence which follow specific mathematical properties. This means that certain properties, like the evolutionary distance between two points in the tree (such as, for example, extant species or language varieties), can be derived by summing up the lengths of the branches connecting them. Since phylogenetic trees are idealized models of divergence in evolution, they should not be interpreted as a direct reflection of reality, but an idealized abstraction that helps to shed light on some major aspects of evolutionary processes, or – as

Atkinson and Gray (2006, p. 94) put it: family tree models «are lies that lead us to the truth».

When scholars started to apply biological phylogenetic approaches to linguistic data, they received very mixed reactions. While the new analyses were fascinating for the broader public and received a broad interest by scientists working in non-linguistic historical fields, such as genetics and archaeology, it was met with a high amount of scepticism by practitioners of historical linguistics. The reasons for this scepticism were manifold. On the one hand, there was a moment of envy and anger about the new methods which were highly appraised by the public and featured in newspapers, although they were mostly based on hand-coded datasets. These datasets in turn were based on the application of the traditional comparative method. Thus, the insights for which the new methods received appraisal were ultimately based on the century-long application of the traditional techniques for historical language comparison, which had long since ceased to be noted by the larger public. As a result, many traditional linguists, who had witnessed how their field was gradually declining, were upset that the biological approaches, often applied by outsiders with no direct training in historical linguistics, received more attention than they had ever hoped for, although they were only presenting the same old data in a new quantitative dress.

On the other hand, however, the scepticism of linguists was due to a misunderstanding of the role of formal models in science accompanied by a limited understanding on the biological methods and concepts themselves. Had family trees in the field of historical linguistics mainly served to provide a rough account of the subgrouping of language families, without providing much detail regarding divergence times, the mathematical models which biologists use to infer phylogenies from biological datasets have been developed since the 1970s, reaching a considerable amount of sophistication already in the beginning of the 2nd millennium.

When first phylogenetic analyses of linguistic data appeared, many linguists lacked the detailed knowledge about the way in which data were coded and the algorithms that were used in the early applications. As a result, early criticism against the use of phylogenetic methods from biology was often not very profound, typically concentrating on very general problems of phylogenetic reconstruction, such as lexical borrowing, or uncertainties with respect to cognate judgments, which apply to all attempts of subgrouping. As an example for the lack of knowledge of data structures and algorithms, consider the book-long critique of phylogenetic approaches by Pereltsvaig and Lewis (2015, pp. 64-69), in which the authors claim that phylogenetic methods were not able to distinguish innovations from retentions, which they exemplify, among others, by pointing to the development of the root for the concept 'whole' in Slavic languages (Ivi, p. 69), which shows some peculiar sound changes. This example is misleading for several reasons. First, the concept 'whole' did not appear in the traditional concept lists used by early pioneers of phylogenetic reconstruction (Gray, Atkinson, 2003), which were almost exclusively derived from concept lists proposed by Morris Swadesh in the 1950s (Swadesh, 1950, 1952, 1955, see also the comparison in the Concepticon resource by List, Pathmanathan *et alii*, 2016, <https://concepticon.cldd.org>). Second, most phylogenetic approaches (with extremely few exceptions,

see e.g. Chacon, List 2015) were explicitly ignoring sound changes as evidence for subgrouping, but instead derived phylogenies from lexical data coded for shared cognate words alone. Given that there is extreme uncertainty with respect to retentions and innovations of individual lexical items (see Jacques, List, 2019 for a detailed discussion of retentions and innovations), as reflected in the small number of attempts to arrive at an explicit onomasiological (concept-based) reconstruction of words in individual proto-languages (see Jäger, List, 2018), it was ignored by early critics that phylogenetic approaches could in fact provide us with valid scenarios that we could test against alternative hypotheses.

Another aspect of phylogenetic approaches, which was also ignored for a long time by their critics, was the handling of uncertainty. While uncertainty was always a problem of historical language comparison – which is also reflected in the numerous versions of individual proto-languages reconstructed by different scholars – critics of the new phylogenetic approaches often did not see that these methods were actually dealing with uncertainty quite elegantly. Thus, instead of producing only one phylogeny and proposing it as the best and only possible one, phylogenetic analyses are typically based on the computation of millions of alternative phylogenies of which a subset is later retained which would explain the data under consideration best. As a result, the family tree model of August Schleicher was replaced by a language family forest that represents our uncertainty with respect to the ‘true phylogeny’ – if this ever existed.

Fewer critical accounts discussed the problems resulting from the transfer of methodology from one scientific domain to another. Prominent problems discussed in this context include the comparability of linguistic and biological data, and the adequacy of evolutionary models to describe evolution in the linguistic domain. The former is specifically reflected in discussions about the coding procedure by which linguistic data was reflected in numerical formats. The coding of data, which consists in the identification of cognate words across different languages, is most of the time still carried out manually. As a result, human coders – typically linguists with knowledge in the history of the languages under comparison – play an important role for automatic analyses. Surprisingly, however, agreement among different coders was only rarely investigated, although there are many steps in the workflow, in which coders may disagree. These can all have an impact on the results of an analysis (Geisler, List, 2010). Examples are the translation of a concept into a word in the target language, the identification of borrowed items, and the assignment of words to cognate classes. The latter problem is related to the underlying model of *cognacy* or *relatedness*. Typically, phylogenetic analyses require a binary decision regarding the cognate status of individual words. According to this model of cognacy, words are either cognate or not, similar to genes which would be either homologous or not. Lexical change processes, however, can easily lead to situations where entire words are not related, while individual word parts are. These cases cannot be handled by methods which are currently applied to linguistic data (List, 2016; Hill, List, 2017).

While critical accounts were very frequent in the first two decades of the second millennium (Holm, 2007, 2008; Pereltsvaig, Lewis, 2015; Donohue, 2012), their number decreased drastically in the past years, and phylogenetic approaches



based on «Bayesian phylolinguistics» (Greenhill *et alii* 2020) have by now become the norm rather than the exception in historical linguistics. This is witnessed by a large number of phylogenetic analyses on different language families which were conducted in the past decade, including, among others, the Dravidian languages spoken in South India (Kolypakam *et alii*, 2018), the Sino-Tibetan languages spoken in the Himalayan region and South-East Asia (Sagart *et alii*, 2019), and the Pama-Nyungan languages spoken in Australia (Bouckaert *et alii*, 2018).

Apart from the forest of trees representing the possible evolutionary scenarios with all their uncertainty of a language family, the quantitative turn also opened the doors for the revival of the idea of network-like evolution, which – as we have seen in the previous section – was already mentioned by Schuchardt (1900). Here again, methodological transfer was done from computational biology to the field of historical linguistics, but while tree-based phylogenetic methods tended to compare languages with eukaryotic species in biology (Pagel, 2009), network approaches from biology were originally designed to study bacterial evolution (Nelson-Sathi *et alii*, 2011, List *et alii*, 2014). While enjoying popularity for some time, being framed as an alternative to phylogenetic tree reconstruction methods or as a better version of the wave theory (Heggarty, 2010), network approaches could not be established as a true alternative to tree-reconstruction methods. One reason is that most network approaches still need a tree-like phylogeny as a backbone against which horizontal edges that do not conform to the major phylogenetic scenario are plotted (see List *et alii*, 2014 for details). Furthermore, the fact that network approaches provide very detailed accounts on potential borrowings between the languages in a given sample showed that the approaches could rarely hold what they promised but would instead often assume borrowings where classical research had excluded them (List, 2019). While many scholars still agree that a network representation in which the tree-like part displays inheritance and horizontal edges display borrowing is more apt to model language history, it seems that the computational methods inspired from evolutionary biology are not yet sufficient to handle linguistic data.

#### 4. METAPHORS OF TOMORROW

The quantitative turn in the beginning of the 21st century has caused quite a stir in the field of historical linguistics. Unlike the attempts initiated by Morris Swadesh (1909-1967) in the 1950s, quantitative approaches are gaining ground now, and more and more researchers demand quantitative expertise and openness to quantitative studies from their peers. As a result, it is very likely that the role which quantitative approaches play in historical linguistics will become more prominent. It is also likely that biologists and linguists will keep inspiring each other, even if it is likely that the inspiration will typically go into one direction, from biology to historical linguistics.

But what new metaphors will be important in the next decades, and what new findings might these new metaphors help us to make? Based on recent developments in the field of evolutionary biology, we see one methodological framework from evolutionary biology which could be fruitfully adapted to future work in linguistics. This framework is the so-called *reconciliation* of gene trees with species trees.

In gene-tree-species-tree reconciliation, one tree (or network) depicting the development of a single gene family is mapped onto an overarching phylogeny representing the major evolution of the species (Nakhleh, 2013). The idea is that individual gene histories may well differ from the overarching histories of the species in which the genes occur. By reconciling two trees with each other or mapping one tree onto the other one, the histories can be compared and specific inferences about specific processes, such as lateral transfer or gene duplication events can be made. Conceptually, gene-tree reconciliation in biology has a direct counterpart in the idea above-mentioned that all words have individual histories. In linguistics, however, scholars have rarely tried to depict the history of single words in a schematic way, not to speak of trying to map the history of individual words onto the overarching history of a language family, and while biologists have started to develop complex models and algorithms by which trees which represent the evolution of different domains can be compared and combined to form complex evolutionary scenarios, linguistic research has so far merely discussed the possibility or reconciling word trees with language trees (Gray *et alii*, 2007) and provided first ideas regarding the formal representation of word trees (Schweikhard, List, forthcoming), while attempts to transfer reconciliation methods directly from biology have not been convincing (Willems *et alii*, 2016).

Theoretically, however, the framework of tree reconciliation offers many advantages for the modeling of language history, as shown at least in anecdotal form in some studies, which point out that the phenomenon of incomplete lineage sorting in biology, which yields patterns of shared genes which seemingly contradict a given phylogeny due to variation in ancestral species accompanied by differential loss of certain alleles in later generations, has a direct counterpart in language history, where variation can also be encountered in different ways (Jacques, List, 2019). In FIG. 2, an example is shown in which different individual word histories are reconciled with an overarching language history. If it will ever be possible to carry out these analyses automatically, instead of collecting the information through manual comparison of etymological sources which provide the accumulated knowledge about language history in a condensed form is not clear by now. For the future of historical linguistics, however, we are convinced that a more formal treatment of individual word histories would be very fruitful.



FIG. 2. Reconciliation of word trees with language trees.

(A) is a single word-tree for terms for 'moon' in four Chinese dialects.

(B) is the language tree, and (C) is the reconciled tree.

## 5. CONCLUSION

In this short review, we have discussed metaphors that were used to describe how languages evolve across different times, starting from early accounts, preceding the rise of historical linguistics as a scientific discipline via recent accounts that

have been flourishing along with the quantitative turn in historical linguistics up to metaphors and framework which we consider fruitful for future research on language history and language evolution. Our overview was quite selective and cannot do justice to the complex history of the discipline of historical linguistics, which evolved a lot over the past centuries through the interaction with other historical disciplines. We hope, nevertheless, that our remarks might inspire the one or the other reader to dive deeper into the fascinating history of historical language comparison.

## REFERENCES

- ALLEN, WILLIAM S. 1953, *Relationship in Comparative Linguistics*, «Transactions of the Philological Society», 52, 1, pp. 52-108.
- ANTTILA, RAIMO 1972, *An Introduction to Historical and Comparative Linguistics*, New York, Macmillan.
- ATKINSON, QUENTIN D.; RUSSELL D. GRAY 2006, *How Old Is the Indo-European Language Family? Illumination or More Moths to the Flame?*, in *Phylogenetic Methods and the Prehistory of Languages*, edited by Peter Forster, Colin Renfrew, Cambridge-Oxford-Oakville, McDonald Institute for Archaeological Research, pp. 91-109.
- BAPTESTE, ERIC; VAN IERSEL, LEO; JANKE, AXEL; KELCHNER, SCOT; KELK, STEVEN; MCINERNEY, JAMES O.; MORRISON, DAVID A. et alii 2013, *Networks: Expanding Evolutionary Thinking*, «Trends in Genetics», 29, 8, pp. 439-441.
- BOUCKAERT, REMCO R.; BOWERN, CLAIRE; ATKINSON, QUENTIN D. 2018, *The Origin and Expansion of Pama-Nyungan Languages Across Australia*, «Nature Ecology and Evolution», 2, 3, pp. 741-749, <https://doi.org/https://doi.org/10.1038/s41559-018-0489-3>.
- CAMPBELL, LYLE 1999, *Historical Linguistics. An Introduction*, Edinburgh, Edinburgh University Press (ii ed.).
- CAMPBELL, LYLE; POSER, WILLIAM JOHN 2008, *Language Classification: History and Method*, Cambridge, Cambridge University Press.
- CHACON, THIAGO; LIST, JOHANN-MATTIS 2015, [9] *Improved computational models of sound change shed light on the history of the Tukanoan languages*, «Journal of Language Relationship», 13, 3, pp. 177-204.
- DAGAN, TAL; MARTIN, WILLIAM 2006, *The Tree of One Percent*, «Genome Biology», 7, 118, pp. 1-7.
- DONOHUE, MARK; DENHAM, TIM; OPPENHEIMER, STEPHEN 2012, *New Methodologies for Historical Linguistics? Calibrating a Lexicon-Based Methodology for Diffusion Vs. Subgrouping*, «Diachronica», 29, 4, pp. 505-522.
- EIGLER, GUNTHER, 2001, *Platon. Kratylos*, 3, Darmstadt, Wissenschaftliche Buchgesellschaft.
- GEISLER, HANS; LIST, JOHANN-MATTIS 2010, *Beautiful Trees on Unstable Ground. Notes on the Data Problem in Lexicostatistics*, in *Die Ausbreitung des Indogermanischen. Thesen aus Sprachwissenschaft, Archäologie und Genetik*, hrsg. Heinrich Hettrich, Wiesbaden, Reichert.
- GEISLER, HANS; LIST, JOHANN-MATTIS 2013, *Do Languages Grow on Trees? The Tree Metaphor in the History of Linguistics*, in *Classification and Evolution in Biology, Linguistics and the History of Science. Concepts-Methods-Visualization*, edited by Heiner Fangerau, Hans Geisler, Thorsten Halling, William Martin, Stuttgart, Steiner, pp. 111-124.
- GRAY, RUSSELL D.; GREENHILL, SIMON J.; ROSS, ROBERT M. 2007, *The Pleasures and Perils of Darwinizing Culture (with Phylogenies)*, «Biological Theory», 2, 4, pp. 360-375.
- GREENHILL, SIMON J.; HEGGARTY, PAUL; GRAY, RUSSELL D. 2021, *Bayesian Phylolinguistics*, in *The Handbook of Historical Linguistics. II*, edited by Richard D. Janda, Brian D. Joseph, Barbara S. Vance, West Sussex, Blackwell, pp. 226-253.
- GRIMM, JACOB 1819, *Deutsche Grammatik*, I, Göttingen, Dieterichsche Buchhandlung.

- HEGGARTY, PAUL; MAGUIRE, WARREN; MCMAHON, APRIL 2010, *Splits or Waves? Trees or Webs? How Divergence Measures and Network Analysis Can Unravel Language Histories*, «Philos. Trans. R. Soc. Lond., B, Biol. Sci.», 365, 1559, pp. 3829-3843.
- HELLSTRÖM, PETTER 2019, *Trees of Knowledge: Science and the Shape of Genealogy*, Doctoral dissertation Uppsala University, Uppsala Studies in History of Ideas, p. 51.
- HILL, NATHAN W.; LIST, JOHANN-MATTIS 2017, *Challenges of Annotation and Analysis in Computer-Assisted Language Comparison: A Case Study on Burmish Languages*, «Yearbook of the Poznań Linguistic Meeting», 3, 1, pp. 47-76.
- HOLM, HANS J. 2007, *The New Arboretum of Indo-European "Trees": Can New Algorithms Reveal the Phylogeny and Even Prehistory of Indo-European?*, «Journal of Quantitative Linguistics», 14, 2-3, pp. 167-214.
- HOLM, HANS 2008, *The Distribution of Data in Word Lists and Its Impact on the Subgrouping of Languages*, in *Data Analysis, Machine Learning, and Applications*, edited by Christine Preisach, Hans Burkhard, Lars Schmidt-Thieme, Reinhold Decker, Heidelberg-Berlin, Springer, pp. 629-636.
- HULL, DAVID L. 1988. *Science as a Process. An Evolutionary Account of the Social and Conceptual Development of Science*, Chicago, The University of Chicago Press.
- JACQUES, GUILLAUME; LIST, JOHANN-MATTIS 2019, *Save the Trees: Why We Need Tree Models in Linguistic Reconstruction (and When We Should Apply Them)*, «Journal of Historical Linguistics», 9, 1, pp. 128-166, <https://doi.org/https://doi.org/10.1075/jhl.17008.mat>.
- JÄGER, GERHARD 2018, *Global-Scale Phylogenetic Linguistic Inference from Lexical Resources*, «Scientific Data», 5, 180189, pp. 1-16.
- KATIČIĆ, RADOSLAV 1966, *Modellbegriffe in der vergleichenden Sprachwissenschaft*, «Kratylos», 11, pp. 49-67.
- KLEIN, WOLF PETER 2004, *Was wurde aus den Wörtern der hebräischen Ursprache? Zur Entstehung der komparativen Linguistik aus dem Geist etymologischer Spekulation*, in *Gottes Sprache in der philologischen Werkstatt: Hebraistik vom 15. bis zum 19. Jahrhundert*, hrsg. Giuseppe Veltri, Gerold Necker, Leiden, Brill («Studies in European Judaism», 11), pp. 3-23.
- KOLIPAKAM, VISHNUPRIYA; JORDAN, FIONA M.; DUNN, MICHAEL; GREENHILL, SIMON J.; BOUCKAERT, REMCO; GRAY, RUSSELL D.; VERKERK, ANNEMARIE 2018, *A Bayesian Phylogenetic Study of the Dravidian Language Family*, «Royal Society Open Science», 5, 171504, pp. 1-17.
- LEES, ROBERT B. 1953, *The Basis of Glottochronology*, «Language», 29, 2, pp. 113-127, <https://www.jstor.org/stable/410164>.
- LEHMANN, WINFRED PHILIPP 1969, *Einführung in die Historische Linguistik*, translated by Rudolf Freudenberg, Heidelberg, Winter.
- LIST, JOHANN-MATTIS 2016, *Beyond Cognacy: Historical Relations Between Words and Their Implication for Phylogenetic Reconstruction*, «Journal of Language Evolution», 1, 2, pp. 119-136, <https://doi.org/10.1093/jole/lzwo06>.
- LIST, JOHANN-MATTIS 2019, *Automated Methods for the Investigation of Language Contact Situations, with a Focus on Lexical Borrowing*, «Language and Linguistics Compass», 13, e12355, pp. 1-16.
- LIST, JOHANN-MATTIS 2021, *Evolutionary Aspects of Language Change*, «Humanities Commons», preprint, currently under review.
- LIST, JOHANN-MATTIS; PATHMANATHAN, J.; LOPEZ, P.; BAPTISTE E. 2016a, *Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics*, «Biology Direct», 11, 39, pp. 1-17.
- LIST, JOHANN-MATTIS; CYSOUW, M.; FORKEL R. 2016b, *Concepticon. A resource for the linking of concept lists*, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pp. 2393-2400.
- LIST, JOHANN-MATTIS, CYSOUW, MICHAEL, FORKEL, ROBERT 2016, *Concepticon. A Resource for the Linking of Concept Lists*, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari (Conference Chair), Khalid

- Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, Stelios Piperidis, Luxembourg, European Language Resources Association (ELRA), pp. 2393-2400.
- LIST, JOHANN-MATTIS; NELSON-SATHI, SHIJULAL; GEISLER, HANS; MARTIN, WILLIAM 2014, *Networks of Lexical Borrowing and Lateral Gene Transfer in Language and Genome Evolution*, «Bioessays», 36, 2, pp. 141-150.
- LIST, JOHANN-MATTIS; PATHMANATHAN, JANANAN SYLVESTRE; LOPEZ, PHILIPPE; BAPTESTE, ERIC 2016, *Unity and Disunity in Evolutionary Sciences: Process-Based Analogies Open Common Research Avenues for Biology and Linguistics*, «Biology Direct», 11, 39, pp. 1-17.
- MEILLET, ANTOINE 1954, *La méthode comparative en linguistique historique*, Paris, Honoré Champion [1925].
- MÜNSTER, SEBASTIAN 1523, *Dictionarium hebraicum, nunc primum editum et typis excusum, adiectis chaldaicis vocabulis non parum multis*, apud Frobenium.
- NAKHLEH, LUAY 2013, *Computational Approaches to Species Phylogeny Inference and Gene Tree Reconciliation*, «Trends in Ecology and Evolution», 28, 12, pp. 719-728.
- NELSON-SATHI, SHIJULAL LIST; JOHANN-MATTIS, GEISLER HANS; FANGERAU, HEINER GRAY; RUSSELL, D. MARTIN; WILLIAM, DAGAN TAL 2011, *Networks Uncover Hidden Lexical Borrowing in Indo-European Language Evolution*, «Proceedings of the Royal Society of London B: Biological Sciences», 278, 1713, pp. 1794-1803.
- PAGEL, MARK 2009, *Human Language as a Culturally Transmitted Replicator*, «Nature Reviews. Genetics», 10, pp. 405-415.
- PERELTSVAIG, ASYA; LEWIS, MARTIN W. 2015, *The Indo-European Controversy. Facts and Fallacies in Historical Linguistics*, Cambridge, Cambridge University Press.
- SAGART, LAURENT; JACQUES, GUILLAUME; LAI, YUNFAN; RYDER, ROBIN; THOUZEAU, VALENTIN; GREENHILL, SIMON J.; LIST, JOHANN-MATTIS 2019, *Dated Language Phylogenies Shed Light on the Ancestry of Sino-Tibetan*, «Proceedings of the National Academy of Science of the United States of America», 116, pp. 10317-10322.
- SANKOFF, DAVID 1969, *Historical Linguistics as Stochastic Process*, Dissertation, Montreal, McGill University.
- SANKOFF, DAVID; SANKOFF, GILLIAN 1980, *Wave Versus Stammbaum Explanations of Lexical Similarities*, in *The Social Life of Language*, edited by Gillian Sankoff, Philadelphia, University of Pennsylvania Press, pp. 143-151.
- SCHLEICHER, AUGUST 1853, *Die ersten Spaltungen des indogermanischen Urvolkes*, «Allgemeine Monatsschrift Für Wissenschaft Und Literatur», 3, pp. 786-787.
- SCHLEICHER, AUGUST 1861, *Compendium der vergleichenden Grammatik der indogermanischen Sprache*, 1, Weimar, Böhlau.
- SCHLEICHER, AUGUST 1863, *Die Darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Haeckel*, Weimar, Böhlau.
- SCHLEICHER, AUGUST 1868, *Eine Fabel in indogermanischer Sprache*, in *Beiträge zur vergleichenden Sprachforschung auf dem Gebiete der arischen, celtischen und slawischen Sprachen*, 5, hrsg. Adalbert Kuh, August Schleicher, Berlin, Dümmler, pp. 206-208.
- SCHMIDT, JOHANNES 1872, *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*, Weimar, Böhlau.
- SCHOTTEL, JUSTUS GEORG 1663, *Ausführliche Arbeit von der Teutschen HauptSprache*, 5 vols, Braunschweig, Zilligern, [http://reader.digitale-sammlungen.de/de/fs1/object/display/bsb10523348\\_00001.html](http://reader.digitale-sammlungen.de/de/fs1/object/display/bsb10523348_00001.html).
- SCHUCHARDT, HUGO 1900, *Über die Klassifikation der romanischen Mundarten. Probe-Vorlesung, gehalten zu Leipzig am 30. April 1870*, Graz [1870], <https://archive.org/details/berdieklassifiko1schugoo/page/n4/mode/2up>.
- SCHWEIKHARD, NATHANAE E.; LIST, JOHANN-MATTIS forthcoming, *Modeling Word Trees in Historical Linguistics. Preliminary Ideas for the Reconciliation of Word Trees and Language Trees*, in *Sprach(en)forschung: Disziplinen und Interdisziplinarität. Akten der 27. Fachtagung der Gesellschaft für Sprache und Sprachen*, hrsg. Bela Brogyanyi, Reiner Lipp, Hamburg, Dr. Kovač.

- SICK, BASTIAN 2004, *Der Dativ ist dem Genitiv sein Tod. Ein Wegweiser durch den Irrgarten der deutschen Sprache*, Köln, Kiepenheuer und Witsch.
- SWADESH, MORRIS 1950, *Salish Internal Relationships*, «International Journal of American Linguistics», 16, 4, pp. 157-167, <https://www.jstor.org/stable/1262898>.
- SWADESH, MORRIS 1952, *Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos*, «Proceedings of the American Philosophical Society», 96, 4, pp. 452-463.
- SWADESH, MORRIS 1955, *Towards Greater Accuracy in Lexicostatistic Dating*, «International Journal of American Linguistics», 21, 2, pp. 121-137, <https://www.jstor.org/stable/1263939>.
- WEBB, DANIEL 1787, *Some Reasons for Thinking, That the Greek Language Was Borrowed from the Chinese*, in *Notes on the Grammatica Sinica of Mons. Fourmont*, London, Dodsley.
- WEISS, MICHAEL 2015, *The Comparative Method*, in *The Routledge Handbook of Historical Linguistics*, edited by Claire Bower, Nicholas Evans, New York, Routledge, pp. 127-145.
- WILLEMS, MATTHIEU; LORD, ETIENNE; LAFOREST, LOUISE; LABELLE, GILBERT; LAPOINTE, FRANÇOIS-JOSEPH; DI SCIULLO, ANNA MARIA; MAKARENKOV, VLADIMIR 2016, *Using Hybridization Networks to Retrace the Evolution of Indo-European Languages*, «BMC Evolutionary Biology», 16, 1, pp. 1-18, <https://doi.org/10.1186/s12862-016-0745-6>.