Trees, Waves and Friends

Maria Zielenbach

May 2023

In this workshop, we will explore different models of language diversification and the methods they utilize. Models of language diversification try to explain how the many languages on Earth developed (diversified) and how they are related to each other.
More precisely we will discuss:

- The Traditional Tree Model: the *genealogical/genetic* Trees we all know (and many love)
- The Traditional Wave Model: a model intended to be a alternative to the Tree Model but which never really took off (for reasons)
- Historical Glottometry: a revised version of the Wave Model
- Traditional Lexicostatistics: a quantitative method to infer language Trees which has now mostly been abandoned
- Bayesian Phylogenetics: a new quantitative method to infer language Trees based on Bayesian statistics

The learning goals of the workshop are as follows:

- acquire a 'passive reading knowledge' of the different models and methods
- understand that language relatedness and classification is not a theory free issue and that models of language diversification are controversially debated among linguists
- understand that not every language tree is based on the same assumptions and method

For each model or method we will answer the questions:

- What does it do and what does it assume?
- How are the findings depicted and how to evaluate them?
- What are the pros and cons?
- How is the model/method regarded today?

There are many more methods and models that could be discussed in this context. We will quickly look at some of them in the last part of the workshop. You will notice that the focus during the discussion of the models and methods lays on the debates around them instead of their individual details. This is done on purpose because I believe it is easier to read up the facts about a model or method. I have also excluded some arguments of the ongoing debate in order to fit the topic into a 1,5h session. One question you may want to ponder for yourself is how each model defines a 'language'.

1 Important terminology
Before we start, here are some important terms that we will use throughout the workshop.

1.1 Cognates
Cognates are (lexical and grammatical) morphemes in different languages which derive from the same morpheme in the proto-language. Sometimes you will read that cognates are morphemes that "share form and function". While most cognates do that, detecting similarity is only the first step in finding possible cognates. Many linguists require regular sound correspondences in order to establish legit cognacy. Proposed cognates without establish sound correspondences are suspicious of being mere 'look-alikes'. Cognates are establish through the historical comparative method. (cf Campbell (2013: 110), (Trask 2015: 193))

1.2 Proto-language
The language from which all related languages have developed. In one way or the other, all the models and methods discussed in this workshop assume that languages diversify from one language into several *mutually unintelligible* languages ('divergence'). The original *ancestor language* is the Proto-language and the diversified languages are often called daughter languages. If not attested, a Proto-language can be reconstructed through the Historical Comparative Method. (cf Campbell (2013: 107), (Trask 2015: 167))

1.3 The Historical Comparative Method
The Historical Comparative Method (HCM) strives to reconstruct the Proto-language from which all related languages are derived through establishing *regular correspondences* in phonology and morphology between languages. The HCM is closely linked to the Neogrammarians (germ. 'Junggrammatiker') who postulated that *sound laws suffer no exceptions* (Germ. 'Ausnahmslosigkeit der Lautgesetze'). [1] (cf Campbell (2013: 107ff), Trask (2015: 191ff))

1.4 Shared retentions, shared innovations, horizontal diffusion and parallel evolution
Shared retentions, *shared innovations*, *horizontal diffusion* and *parallel evolution* are four explanations for similarities between languages. Shared retentions (*symplesiomorphies*) are features, lexemes etc. which have been directly inherited from an ancestor language. Shared retentions cannot be used for subgrouping since all languages that developed from one ancestor language are per default supposed to have the same features. If the features, lexemes etc. of an ancestor language are not retained in a daughter language, the new features, lexemes are called shared innovations (*synapomorphies*). Shared innovations are the main diagnostic for subgrouping. If two or more languages share a feature, lexeme etc. but they are not genetically related, it may have been a borrowing from one language into the other. This is called (horizontal) diffusion. If a shared feature, lexeme etc. cannot be explained as either a shared retention, shared innovation or diffusion, its occurrence may be due to parallel evolution (*homoplasies*), i.e. the development of the same feature etc. in two unrelated languages by accident. (cf (Jacques & List 2019: 137))

---

[1] You may wonder why the HCM is not discussed as one of main topics of this workshop. This has the following reasons: Firstly, understanding its basic principles usually is not a problem but going into the details would require a workshop of its own and secondly, information on this method is readily available *everywhere*.

2 The Traditional Tree Model

The Tree Model (Germ. *Stammbaummodell*) is *the* model in historical linguistics. It is usually attributed to German philologist August Schleicher (1821-1868), however, there are language trees that precede Schleicher's (1853) paper (cf. List et al. (2016)).

> "Aus der Art und Weise, wie sämmtliche indogermanische Sprachen unter einander verwandt sind, schloss man nun mit Recht, dass sie aus einer Urspache entsprungen seien, dass eine Nation, das indogermanische Urvolk, sich mit der Zeit in jene acht Völker getheilt habe, von denen jedes in ähnlicher Weise sich später wieder differenziirte, bis endlich die Mannigfaltigkeit unserer Epoche erstand." (Schleicher 1853: 786)

2.1 What does it do and what does it assume?

- The Tree Model assumes that language diversify through abrupt splits from an ancestor language and isolated development of the new varieties.
- Related languages are sorted into smaller genetic groups ('subgrouping') based on *shared innovations* established through the *Historical Comparative Method*. The languages in these subgroups are more closely related.
- Contact phenomena (e.g. borrowing) are treated as a kind of anomaly since the Tree Model only considers inherited material ('vertical descent').

2.2 How are the findings depicted and how to evaluate them?

- Traditional Trees usually proceed from top to bottom or left to right, with the proto-language (German *Ursprache*) at the top or left side.
- Every node in the tree represents a split. The language that heads the node is the proto-language from which the languages at the end of each line below the node developed. All languages below a node form a *subgroup*.
- Authors should tell you which shared innovations their Tree is based on. A subgroup should be based on more than one shared innovation.

2.3 What are the pros and cons?

- Tree diagrams are easy to read and understand.
- Opponents claim that the Tree Model oversimplifies the history of the languages, leaving out contact phenomena and hence half of the story.
- Opponents say languages do not usually diversify in isolation and that the Tree Model over-conventionalizes the rare cases in which this happens as the general model of language diversification: "In a nutshell, cladistic (tree-based) representations are entirely based on the fiction that the main reason why new languages emerge is the abrupt division of a language community into separate social groups. Trees fail to capture the very common situation in which linguistic diversification results from the fragmentation of a language into a network of dialects which remained in contact with each other for an extended period of time" (François 2014: 162)
- The Tree Model relies on bundles of *shared innovations* for subgrouping. These are often difficult to find for closely related languages that have always been in close contact ('dialect continuum', 'linkage'). Network-like representations have been proposed as an alternative (cf Jacques & List 2019).
- Advocates of the Tree Model usually reply to these issues that the Tree Model is not meant to capture every part of language history ('models are simplifications per definition')
- Subgroupings are subjective: they are based on the decisions of individuals (Bayesian Phylogenetics is supposed to help out here).

2.4 How is it regarded today?

- the Tree Model remains as the standard model of historical linguistics
- when linguists talk about 'related' they usually mean 'related in the sense of the tree model' and the goal of many historical endeavours is to find the best tree for a language family

2.5 Literature

- Campbell (2013), chap. 6.4, 7
- Trask (2015: 169-172)
- Carling et al. (2022)
- François (2014): critical assessment of the Tree Model (also see the other papers under Historical Glottometry)

3   Traditional Lexicostatistics

Traditional Lexicostatistics is an early quantitative approach to genetic classification based on shared lexicon. It was developed in the 1950s by Morris Swadesh (1909-1967) with the aim to introduce a means for dating branching events.

> "[T]he fundamental everyday vocabulary of any language – as against the specialized or "cultural" vocabulary – changes at a relatively constant rate. The percentage of retained elements in a suitable test vocabulary therefore indicates the elapsed time." (Swadesh 1952: 452)

3.1   What does it do and what does it assume?
- Lexicostatistics measures relatedness between two languages based on lexical similarity.
- Standardized word lists ('Swadesh lists') are use which cover is allegedly 'stable' vocabulary which is unlikely to be borrowed ('basic vocabulary').
- Lexicostatistics assumes that lexical similarity of basic vocabulary is lost at a consonant rate, this is used to date branching events = glottochronology (in analogy to carbon dating, cf Swadesh (1952: 452, 454))

3.2   How are the findings depicted and how to evaluate them?
- Sometimes the results of a lexicostatistic survey is depicted in a tree, these can be read like traditional trees (but be careful: they are based on a completely different methods).
- Another means of depicting lexicostatistic results are matrices that show the similarity between each compared language.
- Authors should tell you how they chose the languages they compare and how they identified the 'cognates'.

3.3   What are the pros and cons?
- Well, it's easy.
- Lexicostatistics can give a first impression of how similar languages are.
- There are several problems with using standardized concept word lists: "concept fuzziness" (some concepts are hard to translate, e.g. 'love', 'know' etc.), "synonymous differentiation" (concept is divided among several lexemes in a language, e.g. many languages don't have a word for 'bird (in general)' but only for specific species), "linguistic diversity" ("different translations for the basic concept, due to dialectal or sociolinguistic variation", e.g. *watch* vs *observe* vs *monitor*) (Geisler & List 2010: 3)
- 'Stability' of vocabulary can vary, every part of the lexicon and grammar can be borrowed.
- In a perfect world, lexicostatistics would be used for languages whose relatedness has already been establish, based on real cognates. But in the past, linguists have often used superficial look-alikes from superficially similar languages to establish relatedness through lexicostatistics (also cf Geisler & List (2010: 3-4) for more problems with cognacy).

3.4   How is it regarded today?
- Glottochronoly was generally rejected: languages don't change at the prescribed constant rates.
- Swadesh's lists are now living a second life as elicitation tools in language documentation.
- "The finding of substantial rate variation in languages was so damaging to glottochronology that historical linguistics largely rejected quantitative methods. Today, lexicostatistics and glottochronology are seen as textbook examples of bad linguistic methodology, and we are told that 'linguists do not do dates'" (Greenhill et al. 2021: 227)

3.5   Literature
- Swadesh's original method: Swadesh (1950), Swadesh (1952), Swadesh (1955)
- Greenhill et al. (2021): historical overview from lexicostatistics to Bayesian phylogenetics
- Geisler & List (2010): critical assessment of the method

4 Bayesian Phylogenetics

Bayesian Phylogenetics is a new method of doing lexicostatistics which is based on methods from biology and Bayesian statistics. It is claimed to balance out the shortcomings of Traditional Lexicostatistics and Traditional Trees. It is very important to understand that it is not just a 'modern' method to infer a genetic Tree but that it relies on different theoretical and methodological assumptions (cf Carling et al. (2022)).

> "The approach infers a sample of trees that have a high probability of explaining the patterns in the data under a specified model of character change." (Greenhill et al. 2021: 229)

4.1 What does it do and what does it assume?
- Bayesian Phylogenetics can process large data sets (better than humans).
- It can: "evaluate subgrouping hypothesis", "date language divergences", "estimate ancestral states", "infer rates of change in lexical and grammatical traits", "test hypotheses of functional dependencies in linguistic features", "infer geographic homelands and migration routes" (Greenhill et al. 2021: 228-229)
- Instead of one tree, a a bunch of trees (forest) is generated. The trees in the forest are rated due to their probability. There are different ways to generate forests.
- Like lexicostatistics, Bayesian phylogenetics can be misused, especially when non-specialists take data from random languages without establishing relatedness or cognacy first.
- The data comes in the form of word lists (e.g. Swadesh's 200-meaning list), coded for cognacy beforehand.
- Bayesian phylogenetics does not directly use sound changes as diagnostics for subgrouping, instead cognates are assumed to entail these innovations.
- Splits are dated based on historical information which is then extrapolated to parts of the tree for which no historical records exist, instead of using set rates for all language families ('relaxed clock').

4.2 How are the findings depicted and how to evaluate them?
- *Consensus trees* show the common denominator of the forest, the little numbers next to the branches show the probability of each subgroup ('clade') aka the percentage of generated trees in the forest that include this subgroup
- *Densitrees* show all probably generated trees on top of each other and hence the uncertainties
- Authors should tell you how they chose the languages they compare and how they identified the 'cognates'. There should be an in-depth discussion of the methodology (including the code!).

4.3 What are the pros and cons?
- Well, it's not easy.
- Opponents say Bayesian Phylogenetics is just fancy lexicostatistics and can't make up for the latter's problems – advocates usually claim that opponents don't understand the method.
- Opponents say that the method invites non-specialist to process huge amounts of data that they understand which leads to wrong findings (which are published in fancy journals anyways) – advocates usually agree with this point (data must therefore be pre-curated by specialists).
- The word lists in Bayesian phylogentics show the same problems as the word lists used in lexicostatistics (cf above)
- Bayesian phylogenetics shares the main assumptions of Traditional Trees: contact phenomena are seen as noise not as an fundamental part of language history. If you have a problem with Traditional Trees you will also have a problem with trees produced by Bayesian Phylogenetics.
- Advocates claim that matches between result of Bayesian phylogenetics and traditional trees prove that the method works.

4.4 How is it regarded today?
- Bayesian Phylogenetics has gained extreme popularity in the past years.

4.5 Literature

- Greenhill et al. (2021): comprehensive overview of theory and methodology
- Evans et al. (2021)
- `https://www.youtube.com/watch?v=KOZDaFM_MnY`: lecture about Bayesian Phylogenetics
- cf Greenhill et al. (2021) and Simon Greenhill's Website for more publications

5 The Traditional Wave Model

The Wave Model (German *Wellentheorie*) was developed as an alternative to the Tree Model. It is usually attributed to Johannes Schmidt (1843-1901, Schmidt (1872)) but was actually proposed a little earlier by Hugo Schuchardt (1842-1927, Schuchardt (1868), cf also Schuchardt (1885)).

> "Ich möchte an [stelle des baumes] das bild der welle setzen, welche sich in concentrischen mit der entfernung vom mittelpunkte immer schwächer werdenden ringen ausbreitet." (Schmidt 1872: 27)

5.1 What does it do and what does it assume?
- The Wave Model assumes that dialect continuums are the most common state of languages (languages are never uniform, huge intra-language variation is the norm), therefore the Wave Model claims to have greater historical accuracy than the Tree Model
- Innovations are assumed to start in one variety and then 'wave' over adjacent varieties which adopt the innovations, likewise innovations within a variety (e.g. sound changes) start in one word and then 'wave' over to other words. Diversification evolves through slow divergence of varieties first into dialects and then into languages.
- "Since later changes may not cover the same area, there may be no sharp boundaries between neighbouring dialects or languages; rather, the greater the distance between them, the fewer linguistic traits dialects or languages may share." (Campbell 2013: 188)

5.2 How are the findings depicted and how to evaluate them?
- 'Wave diagrams' usually show a number of languages/varieties, displayed according to their geographic location. This depiction was popularized by Bloomfield (1933)316.
- Lines are used to show the overlapping shared features, all varieties within a circle share a specific feature.
- Authors should demonstrate why they don't assume that a branching event has taken place.

5.3 What are the pros and cons?
- The Wave Model targets areas where the Tree Model fails (or has been claimed to fail), i.e. dialects and other closely related varieties.
- The Wave Model does not distinguish between variety internal changes and changes that are due to contact (the processes are not distinguished).
- Sound changes are claimed to show exceptions but opponents say that exceptions can only be found because we know the regular sound changes.
- 'Wave diagrams' may be more faithful to the *synchronic* distribution of linguistic features but do not have a *diachronic* depths (they show that features diffused but not when and how), they are also more messy than trees.

5.4 How is it regarded today?
- The Wave Model never received the same popularity as the Tree Model in historical linguistics but you may encounter it from time to time
- Other linguistic disciplines adopted the Wave Model, or at least Wave Model-like representations, e.g. dialectology and areal linguistics

5.5 Literature
- Campbell (2013), chap. 7
- Trask (2015: 172-174)
- François (2014): positive assessment of the Wave Model (also see all papers listed under Historical Glottometry)
- Podcast about the criticism against the Neogrammarians (you may want to listen to episode 9 about the Neogrammarians first)

6   Historical Glottometry

Glottometry was devised by Alexandre François and Siva Kalyan as a modern, Wave Model-based alternative to the Tree Model.

> "The objective of Historical Glottometry is to identify genealogical subgroups in a language family, and measure their relative strengths so as to assess their historical patterns of distribution across social networks." (François 2014: 173)

6.1   What does it do and what does it assume?
- Historical Glottometry adopts the basic assumptions of the Wave Model that innovations 'wave' over varieties and languages develop from dialect continuums. On the other hand, it also assumes that 'related' varieties are derived from the same common ancestor. It therefore mostly deals with 'waves' within related (and mutual intelligible) varieties.
- The Historical Comparative Method is used to find shared features between languages.
- Varieties can belong to several diffusion areas (unlike in Trees where every variety or subgroups is headed by one node). "[I]t is often the case [...] that an innovation only spreads partway through a population before that population splits. In this situation, an innovation need not be passed on to all of the descendants of the language it occurs in, but only to some of them." (Kalyan & François 2019: 168)
- Historical Glottometry accepts that languages may diversify due to splits, but such cases are regarded as rare and a special case of 'waves'.

6.2   How are the findings depicted and how to evaluate them?
- The diagrams are similar to isogloss maps but the thickness and darkness of the lines is used to show degrees of relationship.
- The darkness of the lines shows the 'cohesiveness' of a subgroup: "the proportion of supporting evidence with respect to the entire set of relevant evidence" (how many innovations does a subgroup share compared to the total number of scrutinized innovations?)
- The thickness of the lines shows the 'subgroupiness': "the product of the cohesiveness rate (k) with the number of exclusively shared innovations" (how many innovations does a subgroup share compared to the total number of scrutinized innovations and how many does it share exclusively?)
- Authors should explain you how they differentiated between inherited and diffused features.

6.3   What are the pros and cons?
Several of the counterarguments discussed here also apply to the Wave Model in general.

- The strongest argument against Historical Glottometry is that it only depicts the synchronic distribution of features and lacks diachronic depths. Advocates claim that diachrony can be read from the isoglosses.
- Opponents claim that Historical Glottometry does not distinguish between shared innovations and parallel evolution. They also argue that the concept of shared innovations only makes sense within the Tree Model because it does not merely describe a phenomenon (languages share a feature) but also explains it (because they constitute a split event from an ancestor language). Advocates retort that shared innovations can be distinguished from parallel innovations based on their distribution pattern.

6.4   How is it regarded today?
- while they are some studies that have used Glottometry in the past it hasn't really set off

6.5   Literature
- François (2014): first paper that introduced Glottometry
- Kalyan & François (2018): similar to the previous paper but with a stronger focus on the actual methodology
- Jacques & List (2019): critical reply to Kalyan & François (2018) from a Tree Model perspective
- Kalyan & François (2019): reply to Jacques & List (2019)
- Agee (2018): Glottometric subgrouping for Germanic

6.6 Conclusion

Some linguists regard the Tree Model and Wave Model not as alternatives but as complementary models that cover different aspects of language history (cf Carling et al. (2022)), others view the shortcomings of one of them as too serious to use both models (cf Jacques & List (2019)).

"The tree model, based on the principle of divergence, classifies according to splits, continually producing new subclades. The wave model illustrates the effects of convergence and advergence over time, i.e., how similarities between geographically adjacent languages may evolve, creating a gradual, dynamic diversity. Both models are equally important in explaining variation and classifying languages, and both models can be used to describe the evolution of languages and dialects, as well as linguistic subdomains, down to the most fine-grained level of individual traits" (Carling et al. (2022))

Literature

Agee, Joshua. 2018. *A Glottometric Subgrouping of the Early Germanic Languages*. San Jose, CA, USA: San Jose State University Master of Arts. https://doi.org/10.31979/etd.69cp-8xz2.

Bloomfield, Leonard. 1933. *Language*. London: Allen & Unwin.

Campbell, Lyle. 2013. *Historical linguistics: an introduction*. 3. ed. Edinburgh: Edinburgh Univ. Press. 538 pp.

Carling, Gerd, Chundra Cathcart & Erich Round. 2022. Reconstructing the origins of language families and variation. In Nathalie Gontier, Andy Lock & Chris Sinha (eds.), *The Oxford Handbook of Human Symbolic Evolution*, 1st edn. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198813781.013.34.

Evans, Cara L., Simon J. Greenhill, Joseph Watts, Johann-Mattis List, Carlos A. Botero, Russell D. Gray & Kathryn R. Kirby. 2021. The uses and abuses of tree thinking in cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1828). rstb.2020.0056, 20200056. https://doi.org/10.1098/rstb.2020.0056.

François, Alexandre. 2014. Trees, waves and linkages: Models of Language Diversification. In Claire Bowern & Bethany Evans (eds.), *The Routledge Handbook of Historical Linguistics* (Routledge Handbooks in Linguistics), 161–189. London; New York: Routledge. https://doi.org/10.4324/9781315794013.ch6.

Geisler, Hans & Johann-Mattis List. 2010. *Beautiful Trees on Unstable Ground: Notes on the Data Problem in Lexicostatistics*. https://hal.science/hal-01298493 (11 May, 2023).

Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2021. Bayesian Phylolinguistics. In Richard D. Janda, Brian D. Joseph & Barbara S. Vance (eds.), *Handbook of historical linguistics* (Blackwell Handbooks in Linguistics volume 2), 226–253. Hoboken, NJ, USA: Wiley.

Jacques, Guillaume & Johann-Mattis List. 2019. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1). 128–167. https://doi.org/10.1075/jhl.17008.mat.

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. *Senri Ethnological Studies* 98. 59–89.

Kalyan, Siva & Alexandre François. 2019. When the waves meet the trees: A response to Jacques and List. *Journal of Historical Linguistics* 9(1). 168–177. https://doi.org/10.1075/jhl.18019.kal.

List, Johann-Mattis, Jananan Sylvestre Pathmanathan, Philippe Lopez & Eric Bapteste. 2016. Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics. *Biology Direct* 11(1). 39. https://doi.org/10.1186/s13062-016-0145-2.

Schleicher, A. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3. 786–787. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2381174 (15 November, 2022).

Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: H. Böhlau.

Schuchardt, Hugo Ernestus Mario. 1868. *Der Vokalismus des Vulgärlateins*. Leipzig: Teubner.

Schuchardt, Hugo Ernestus Mario. 1885. *Ueber die Lautgesetzte. gegen die Junggrammatiker*. Berlin: Robert Oppenheim.

Swadesh, Morris. 1950. Salish Internal Relationships. *International Journal of American Linguistics* 16(4). 157–167.

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). 452–463.

Swadesh, Morris. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21(2). 121–137. `http://www.jstor.org/stable/1263939`.

Trask, Larry. 2015. *Trask's Historical Linguistics*. Robert McColl Millar (ed.). 3rd. London; New York: Routledge.