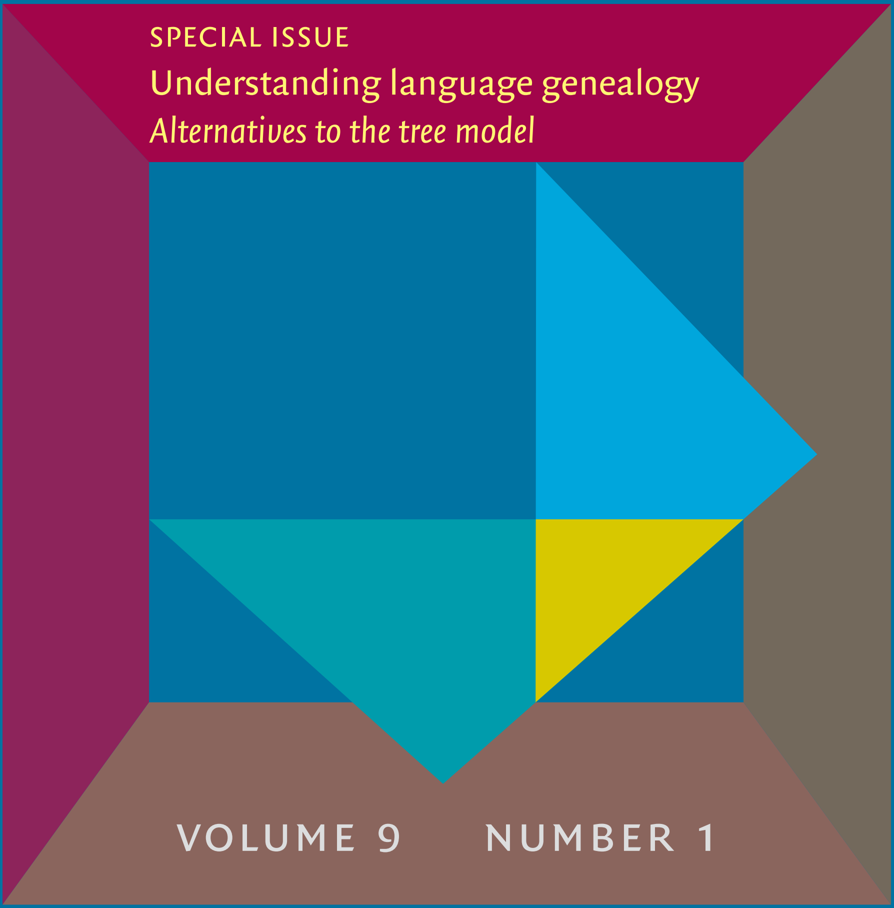
Journal of Historical Linguistics



SPECIAL ISSUE

Understanding language genealogy

Alternatives to the tree model

VOLUME 9 NUMBER 1

2019

John Benjamins Publishing Company

Journal of Historical Linguistics

VOLUME 9 NUMBER 1 2019

John Benjamins Publishing Company
Amsterdam/Philadelphia

Journal of Historical Linguistics

SPECIAL ISSUE

Understanding language genealogy
Alternatives to the tree model

Edited by Siva kalyan, Alexandre François and
Harald Hammarström

Australian National University / LaTTiCe, CNRS, École Normale Supérieure, Univ.
Paris 3 Sorbonne nouvelle, Australian National University / Uppsala University

Table of contents

INTRODUCTION

ARTICLES

INTRODUCTION

Problems with, and alternatives to, the tree model in historical linguistics

Siva Kalyan[1], Alexandre François[1,2] and
Harald Hammarström[3]
[1] Australian National University |
[2] LaTTiCe (CNRS; ENS-PSL; Paris 3-USPC) | [3] Uppsala University

Ever since it was popularized by August Schleicher (1853, 1873), the family-tree model has been the dominant paradigm for representing historical relations among the languages in a family. There have been many other proposals for representing language histories: for example, Johannes Schmidt's (1872) "Wave Model" (as illustrated, e.g., in Schrader 1883: 99 and Anttila 1989: 305); Southworth's (1964) "tree-envelopes" (which seem to predate the "species trees" of phylogeography, e.g. Goodman et al. 1979; Maddison 1997); Hock's (1991: 452) "'truncated octopus'-like tree"; and, more recently, NeighborNet (Hurles et al. 2003; Bryant et al. 2005) and Historical Glottometry (Kalyan & François 2018). However, none of these representations reaches the simplicity, formalization, or historical interpretability of the family tree model.

The family tree model is simple in that it emerges naturally from a small number of assumptions about the diversification of languages. Firstly, it is assumed that every generation of speakers derives their language from the parental generation. Secondly, it is assumed that speakers sometimes modify the language that they acquire. Thirdly, it is assumed that once a language has been modified, it cannot share any further genealogical innovations with its unmodified variant, but must develop in a separate lineage. These assumptions set up the same kind of "descent with modification" scenario that motivates the use of trees for representing the evolution of biological species. Furthermore, tree representations allow for the use of powerful techniques of phylogenetic inference that have been developed in biology (see Greenhill & Gray 2009; Baum & Schmidt 2013), and the stringent assumptions underlying a family tree make it possible to infer the relative age of a linguistic feature by looking at its synchronic distribution within the language family (see Jacques & List this volume: Section 5.3, Kalyan & François 2018: Section 2.1, and Baum & Schmidt 2013: Chapter 10 for parallels in biology).

https://doi.org/10.1075/jhl.00005.kal
Journal of Historical Linguistics 9:1 (2019), pp. 1–8. issn 2210-2116 | e-issn 2210-2124

Yet there are important reasons to be skeptical of the accuracy and usefulness of the family tree model in historical linguistics. When applying that model to a language family, it is assumed that every linguistic innovation applies to a language as an undifferentiated whole (François 2014: 163); in other words, each node in a tree represents a linguistic community as a point with no "width."[1] This assumption makes it impossible to use a tree to model the *partial* diffusion of an innovation within a language community ("internal diffusion" in François 2017: 44) or the diffusion of an innovation across language communities ("external diffusion" in François 2017: 44, or simply "borrowing"). These limitations have long been noticed by historical linguists (Schmidt 1872; Schuchardt 1900), but they become glaringly obvious in the cases discussed by Ross (1988, 1997) under the heading of "linkages," i.e., language families that arise through the diversification, *in situ*, of a dialect network.

Following the discussion in François (2014: 171), a linkage consists of separate modern languages which are all related and linked together by intersecting layers of innovations; it is a language family whose internal genealogy cannot be represented by any tree. Figure 1 shows how innovations (isoglosses numbered #1 to #6) typically spread across a network of dialects (labelled A to H) in intersecting patterns – a configuration encountered both in dialect continua and in the linkages that descend from them.

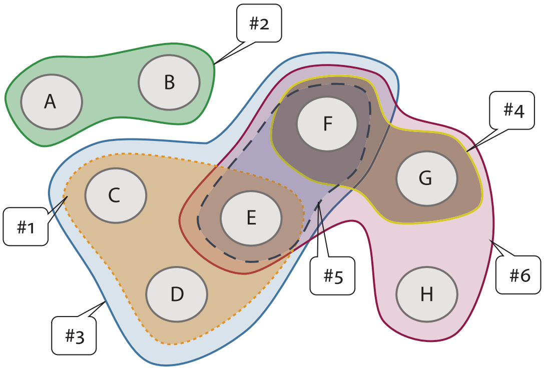


Figure 1. Intersecting isoglosses in a dialect continuum, or a "linkage"

Over the past several decades, linguistic research has revealed numerous examples of linkage phenomena in a broad range of language families: these

---

[1] As noted by Kalyan & François (this volume), this type of assumption is well-justified in biology, where the rate at which innovations spread is far greater than the rate at which populations split, so that for all practical purposes, each innovation affects a species as an undifferentiated whole (Baum & Schmidt 2013: 79).

examples can be found in (subgroups of) Sinitic (Hashimoto 1992; Chappell 2001); Semitic (Huehnergard & Rubin 2011; Magidow 2013); Western Romance (Penny 2000: 9–74; Ernst et al. 2009), Germanic (Ramat 1998), Indo-Aryan (Toulmin 2009), and Iranian (Korn forthcoming); Athabaskan (Krauss & Golla 1981; Holton 2011); Pama-Nyungan (Bowern 2006); and Oceanic (Geraghty 1983; Ross 1988; François 2011) – to name but a few. However, there is no consensus on how best to analyze or model these situations. At one end of the spectrum, we have a "backbone" tree accounting for vertical transmission with a sprinkle of additional borrowing events (as exemplified by, e.g., Ringe et al. 2002 or Nakhleh et al. 2005 for Indo-European); on the other end, the roles are reversed, with the bulk of linguistic change being due to diffusion and the vertical component reduced to a "star phylogeny" (as exemplified by, e.g., the "rake-like tree" discussed by Pawley 1999 for Austronesian or the "fallen leaves" of van Driem 2001 for Tibeto-Burman). The search for ways of quantifying and representing the diversification of a linkage has antecedents stretching back to at least Kroeber & Chrétien (1937) and Ellegård (1959); still, it remains an open problem.

The articles in the present issue all contribute towards addressing this problem from a range of perspectives. The first three articles present case studies of particular language families that exhibit linkage-like behavior, using methodologies that vary in the degree to which they accept the premises of the family-tree model.

Verkerk, in her article "Detecting non-tree-like signal using multiple tree topologies," addresses the question of how and where non-tree-like behaviour can be diagnosed within the framework of Bayesian phylogenetics. Instead of producing a single tree, her methods infer two trees for each language family – a "majority tree" accounting for the largest possible proportion of the data and a "minority tree" accounting for as much as possible of the remainder. The differences between the trees can then be explored, typically with the hypothesis that the minority tree reflects reticulation on top of the "backbone" provided by the majority tree. It is also possible to explore which specific characters (in this case lexical cognate sets) are more or less responsible for these differences. The approach is applied to existing datasets of the Austronesian, Sinitic, Indo-European, and Japonic families.

Elias, in "Visualizing the Boni dialects with Historical Glottometry," takes on the microgroup of Boni dialects (Cushitic) in Kenya and Somalia. The list of lexical and phonological innovations occurring in this group is carefully surveyed before addressing the question of which features are inherited and which are diffusional in origin. The author finds that the earliest split is fully consistent with a tree-like divergence, while the remaining innovations cross-cut any further tree-like evolutionary scenario. The latter set of innovations are instead quantified and illustrated using the newly-proposed technique of Historical Glottometry (François 2014; Kalyan & François 2018). This helps the human observer to visually appreciate the presence and extent of multiple subgrouping, chaining, and areal spread.

Daniels, Barth & Barth, in "Subgrouping the Sogeram languages: A critical appraisal of Historical Glottometry," investigate the little-studied Sogeram subgroup of Trans New Guinea. They enumerate the 196 relevant innovations that occur in this group, then address the question of which historical scenario(s) could explain them. Using Historical Glottometry, the authors quantify and compare various subgroup hypotheses. Evidence is found both for dialect-chain and tree-like break-ups in the history of this subfamily. Furthermore, some improvements to the Historical Glottometry approach are suggested; these relate to visualization, the handling of missing data, and transparency of data analysis.

While all of the above papers discuss theoretical and methodological issues in the context of particular datasets, the final two articles in this issue are more general in nature; they try to make explicit the differences between the family tree model and its alternatives and discuss the extent to which these may be combined into a unified framework for thinking about language diversification.

Jacques & List, in "Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them)," address skeptics of the tree model. They critique some models that have been brought forward as alternatives, in particular distinguishing "data display" from models that encode an explicit historical scenario. Further, they show how data which at first glance seem incompatible with the tree model can in fact be the result of tree-like diversification, once the phenomenon of "incomplete lineage sorting" is taken into account; thus they remind us that a tree-like history for a given set of data should not be dismissed too quickly. Lastly, they give examples in which an assumption of tree-like language diversification simplifies the task of inferring the histories of particular features.

Finally, **Kalyan & François**, in their contribution "When the waves meet the trees: A response to Jacques & List," address the latter authors' critique of Historical Glottometry. They stress agreements between Jacques & List's approach and their own, then turn to the reading of glottometric diagrams. They define a systematic procedure for inferring sequences of historical events from a glottometric diagram, thereby arguing that such diagrams are not limited to static data display. They conclude that Historical Glottometry is in fact compatible with Jacques & List's conception of the tree model, provided that the notion of "incomplete lineage sorting" (i.e., unresolved variation in a proto-language) is extended to the case of dialectal (i.e. geographically-conditioned) variation.

In summary, the articles in this volume provide a sample of possible approaches to analyzing the evolution of a language family in non-cladistic terms. Further, they aim to clarify the assumptions behind the tree model and the extent to which different approaches diverge from these assumptions. We hope that this issue leads to a diversification of methods in historical linguistics, with ample borrowing and diffusion among them.

Acknowledgements

This work contributes to the research program “Investissements d’Avenir,” overseen by the French National Research Agency (ANR-10-LABX-0083): LabEx *Empirical Foundations of Linguistics*, Strand 3 – “Typology and dynamics of linguistic systems.”

References

Anttila, Raimo. 1989. *Historical and Comparative Linguistics (Second Edition)*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.6

Baum, David A. & Stacey D. Smith. 2013. *Tree Thinking: An Introduction to Phylogenetic Biology*. New York: Macmillan.

Bowern, Claire. 2006. Another Look at Australia as a Linguistic Area. *Linguistic Areas* ed. by Yaron Matras, April McMahon & Nigel Vincent, 244–265. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230287617_10

Bryant, David, Flavia Filimon & Russell D. Gray. 2005. Untangling Our Past: Languages, Trees, Splits and Networks. *The Evolution of Cultural Diversity: Phylogenetic Approaches* ed. by Ruth Mace, Clare J. Holden & Stephen Shennan, 69–85. London: UCL Press.

Chappell, Hilary. 2001. Language Contact and Areal Diffusion in Sinitic Languages. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics* ed. by Alexandra Aikhenvald & R. M. W. Dixon, 328–357. Oxford: Oxford University Press.

van Driem, George. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region*. Leiden: Brill.

Ellegård, Alvar. 1959. Statistical Measurement of Linguistic Relationship. *Language* 35:2.131–156. https://doi.org/10.2307/410530

Ernst, Gerhard, Martin-Dietrich Gleßgen, Christian Schmitt & Wolfgang Schweickard. 2009. *Romanische Sprachgeschichte – Ein internationales Handbuch zur Geschichte der romanischen Sprachen/Histoire linguistique de la Romania – Manuel international d’histoire linguistique de la Romania*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110211412.3

François, Alexandre. 2011. Social Ecology and Language History in the Northern Vanuatu Linkage: A Tale of Divergence and Convergence. *Journal of Historical Linguistics* 1:2.175–246. https://doi.org/10.1075/jhl.1.2.03fra

François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. *The Routledge Handbook of Historical Linguistics* ed. by Claire Bowern & Bethwyn Evans, 161–189. London: Routledge.

François, Alexandre. 2017. Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues. *Diffusion: implantation, affinités, convergence* ed. by Jean-Léo Léonard, 43–82. (= *Mémoires de la Société de Linguistique de Paris*, XXIV.) Louvain: Peeters.

Geraghty, Paul A. 1983. *The History of the Fijian Languages* (= *Oceanic Linguistics Special Publication*, 19.) Honolulu: University of Hawaii Press.

Goodman, Morris, John Czelusniak, G. William Moore, A. E. Romero-Herrera & Genji Matsuda. 1979. Fitting the Gene Lineage into its Species Lineage: A Parsimony Strategy Illustrated by Cladograms Constructed from Globin Sequences. *Systematic Biology* 28:2.132–163. https://doi.org/10.1093/sysbio/28.2.132

Greenhill, Simon J. & Russell D. Gray. 2009. Austronesian Language Phylogenies: Myths and Misconceptions About Bayesian Computational Methods. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust* ed. by Alexander Adelaar & Andrew K. Pawley, 1–23. (= *Pacific Linguistics*, 601.) Canberra: Pacific Linguistics.

Hashimoto, Mantarō J. 1992. Hakka in *Wellentheorie* Perspective. *Journal of Chinese Linguistics* 20.1–49.

Hock, Hans Henrich. 1991. *Principles of Historical Linguistics, Second Edition*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110219135

Holton, Gary. 2011. A Geo-Linguistic Approach to Understanding Relationships Within the Athabaskan Family. Paper presented at Language in Space: Geographic Perspectives on Language Diversity and Diachrony, Boulder, Colorado, USA, June 23–24, 2011.

Huehnergard, John & Aaron Rubin. 2011. Phyla and Waves: Models of Classification of the Semitic Languages. *Semitic Languages: An International Handbook* ed. by Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson, 259–278. (= *Handbücher zur Sprach- und Kommunikationswissenschaft*, 36.) Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110251586.259

Hurles, Matthew E., Elizabeth Matisoo-Smith, Russell D. Gray & David Penny. 2003. Untangling Oceanic Settlement: The Edge of the Knowable. *Trends in Ecology and Evolution* 18:10.531–540. https://doi.org/10.1016/S0169-5347(03)00245-3

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. *Let's Talk about Trees: Problems in Representing Phylogenic Relationships Among Languages* ed. by Ritsuko Kikusawa & Lawrence A. Reid, 59–89. (= *Senri Ethnological Studies*, 98.) Ōsaka: National Museum of Ethnology.

Korn, Agnes. Forthcoming. Isoglosses and Subdivisions of Iranian. *Journal of Historical Linguistics* 9:3.

Krauss, Michael E. & Victor Golla. 1981. Northern Athapaskan Languages. *Handbook of North American Indians, Vol. 6: Subarctic* ed. by June Helm & William C. Sturtevant, 67–85. Washington, DC: Smithsonian Institution.

Kroeber, Alfred L. & C. Douglas Chrétien. 1937. Quantitative Classification of Indo-European Languages. *Language* 13:2.83–103. https://doi.org/10.2307/408715

Maddison, Wayne P. 1997. Gene Trees in Species Trees. *Systematic Biology* 46:3.523–536. https://doi.org/10.1093/sysbio/46.3.523

Magidow, Alexander. 2013. Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic Dialect Diversity. University of Texas at Austin PhD dissertation.

Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81:2.382–420. https://doi.org/10.1353/lan.2005.0078

Pawley, Andrew. 1999. Chasing Rainbows: Implications of the Rapid Dispersal of Austronesian Languages for Subgrouping and Reconstruction. *Selected Papers from the Eighth International Conference on Austronesian Linguistics* ed. by Elizabeth Zeitoun & Paul Jen-Kuei Li, 95–138. (= *Symposium Series of the Institute of Linguistics: Academia Sinica*, 1.) Honolulu: University of Hawaii Press.

Penny, Ralph John. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164566

Ramat, Paolo. 1998. The Germanic Languages. *The Indo-European Languages* ed. by Anna Giacalone Ramat & Paolo Ramat, 380–414. London: Routledge.

Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and Computational Cladistics. *Transactions of the Philological Society* 100:1.59–129. https://doi.org/10.1111/1467-968X.00091

Ross, Malcolm D. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. Canberra: Pacific Linguistics.

Ross, Malcolm. 1997. Social Networks and Kinds of Speech-Community Event. *Archaeology and Language I: Theoretical and Methodological Orientations* ed. by Roger Blench & Matthew Spriggs, 209–261. London: Routledge.

Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* ed. by Johann Gustav Droysen & G. W. Nitzsch, 786–787. Braunschweig: C. A. Schwestchke & Sohn.

Schleicher, August. 1873. *Die Darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Häckel, o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena, Second Edition*. Weimar: Hermann Böhlau.

Schmidt, Johannes. 1872. *Die Verwantschaftsverhältnisse der indogermanischen Sprachen*. Hermann Böhlau.

Schrader, Otto. 1883. *Sprachvergleichung und Urgeschichte: linguistisch-historische Beiträge zur Erforschung des indogermanischen Altertums*. Jena: Hermann Costenoble.

Schuchardt, Hugo. 1900. *Über die Klassifikation der romanischen Mundarten. Probe-Vorlesung, gehalten zu Leipzig am 30. April 1870*. Graz.

Southworth, Franklin C. 1964. Family-Tree Diagrams. *Language* 40:4.557–565. https://doi.org/10.2307/411938

Toulmin, Matthew. 2009. *From Linguistic to Sociolinguistic Reconstruction: The Kamta Historical Subgroup of Indo-Aryan*. (= *Pacific Linguistics*, 604.) Canberra: Pacific Linguistics.

Address for correspondence

Siva Kalyan
Department of Linguistics
School of Culture, History and Language
College of Asia and the Pacific
Australian National University
9 Fellows Road
Acton, ACT 2601
Australia
siva.kalyan@anu.edu.au

Co-author information

Alexandre François
CNRS-ENS-LaTTiCe
alexandre.francois@ens.fr

Harald Hammarström
Department of Linguistics and Philology
Uppsala University
harald.hammarstrom@lingfil.uu.se

Detecting non-tree-like signal using multiple tree topologies

Annemarie Verkerk
University of Reading &
Max Planck Institute for the Science of Human History

Recent applications of phylogenetic methods to historical linguistics have been criticized for assuming a tree structure in which ancestral languages differentiate and split up into daughter languages, while language evolution is inherently non-tree-like (François 2014; Blench 2015: 32–33). This article attempts to contribute to this debate by discussing the use of the multiple topologies method (Pagel & Meade 2006a) implemented in *BayesPhylogenies* (Pagel & Meade 2004). This method is applied to lexical datasets from four different language families: Austronesian (Gray, Drummond & Greenhill 2009), Sinitic (Ben Hamed & Wang 2006), Indo-European (Bouckaert et al. 2012), and Japonic (Lee & Hasegawa 2011). Evidence for multiple topologies is found in all families except, surprisingly, Austronesian. It is suggested that reticulation may arise from a number of processes, including dialect chain break-up, borrowing (both shortly after language splits and later on), incomplete lineage sorting, and characteristics of lexical datasets. It is shown that the multiple topologies method is a useful tool to study the dynamics of language evolution.

Keywords: Bayesian phylogenetic inference, Austronesian, Sinitic, Indo-European, Japonic, language contact, reticulation

1. Introduction

In the last eighteen years, phylogenetic methods from evolutionary biology have made inroads into historical linguistics: these methods are applied both to building phylogenetic trees (Ben Hamed & Wang 2006; Gray, Drummond & Greenhill 2009; Lee & Hasegawa 2011; Bouckaert et al. 2012; Grollemund et al. 2015) and to inferring the process of evolution of typological characteristics on trees (Dunn et al. 2011; Verkerk 2014; Zhou & Bowern 2015). The reception of these studies has been mixed, with studies criticizing data type (generally cognate-coded lexical

https://doi.org/10.1075/jhl.17009.ver
Journal of Historical Linguistics 9:1 (2019), pp. 9–69. issn 2210-2116 | e-issn 2210-2124

data), data quality, the applicability of methods and models from another discipline to linguistic data, and the limited possibilities for incorporating previous knowledge into the phylogenetic analysis (Eska & Ringe 2004; Heggarty 2006; Holm 2007; Blench 2015; Pereltsvaig & Lewis 2015).

Nevertheless, phylogenetic methods have been adopted by a wide range of linguists to answer questions regarding the diversification of language families and of typological features (Bowern & Atkinson 2012; Bouchard-Côté et al. 2013; Galucio et al. 2015; Macklin-Cordes & Round 2015; Meira, Birchall & Chousou-Polydouri 2015; Jaeger & Wichmann 2016; Widmer et al. 2017). Dunn (2014: 191), in his chapter in Bowern & Epps' (2014) The Routledge Handbook of Historical Linguistics, notes that they are "increasingly well received within linguistics." This positive reception is for good reason, given that evolutionary biologists have been considering statistical approaches to the study of species diversification for over fifty years now: the first maximum likelihood analysis of (human) genetic material came from Edwards & Cavalli-Sforza (1964). Dunn (2014) describes how linguists began considering quantitative approaches to language history only when Swadesh (Swadesh 1952, 1955) first developed his methods of lexicostatistics and glottochronology, with recent growth having occurred after the adoption of methods from evolutionary biology by Gray & Jordan (2000). Thus, the methods of evolutionary biology are likely to complement those of traditional historical linguistics, especially when it comes to quantitative methods for inferring language genealogies and changes in cultural and linguistic features in these genealogies (Pagel 2009; Levinson & Gray 2012).

In the current article, a phylogenetic inference technique taken from biology is applied to linguistic data, and its usability is reviewed. The technique is called "multiple tree topologies." It is implemented using the software *BayesPhylogenies*, which provides a Bayesian framework for inferring trees for a variety of data types, including binary cognate-coded lexical data (Pagel & Meade 2004). The multiple tree topologies method does not assume that the data have a single evolutionary history, but instead fits multiple independent tree topologies to the data. If, for whatever reason, certain sites (columns in the dataset, which here constitute the cognate sets) point towards a different cladistic grouping than others do, these different signals are picked up and then reflected by displaying statistical support for two or more different tree topologies which are estimated simultaneously. This method is applied to lexical datasets from four language families: Austronesian (Gray, Drummond & Greenhill 2009), Sinitic (Ben Hamed & Wang 2006), Indo-European (Bouckaert et al. 2012), and Japonic (Lee & Hasegawa 2011). This article contributes to the use of phylogenetic methods in historical linguistics by applying a specific phylogenetic method to linguistic data. It is especially relevant to the question of whether phylogenetic tree inference, which models change in contemporary entities in terms of descent with modification from common ancestors, can deal with non-tree-like aspects of language change.

To perform a strict test of this method, we must apply it to a dataset that we know includes both (i) genealogically inherited characteristics that can be traced to a process of descent with modification and (ii) features that have arisen through other processes, including dialect chain break-up, borrowing (from a substrate, superstrate, or adstrate), or incomplete lineage sorting. After applying the multiple topologies method, we can then assess how much of the tree-like and non-tree-like signal is recovered and correctly characterized. However, this required dataset does not exist, as the bundles of data that are studied by historical linguistics contain features that could have arisen via any of these processes (Heggarty, Maguire & McMahon 2010: 3829), and quantitative methods for exploring a dataset in terms of these processes are in their infancy or non-existent (Huson & Scornavacca 2010: 29). In some cases, it is possible to distinguish different sources of divergent signals on the basis of knowledge about which features are more likely to arise through a particular process or through detailed knowledge of surrounding languages or families, but these are not strict tests. A strict test of the multiple topologies method or of any method that attempts to incorporate multiple historical signals in some way is therefore not possible at this point in time.

In this light, this study should be seen as an exploratory experiment for the purpose of observing and reporting on the behavior of the multiple topologies method. However, as this article is part of this special issue exploring non-cladistic approaches to language genealogy, the study is also embedded in a larger context of investigating reticulation patterns in language and culture. Within the context of this discussion, after presenting the results from each of the four language families, I speculate on the processes that might have given rise to the evidence found for multiple topologies in three of the language families. My speculation is based on the assumption that if evidence for multiple topologies is found, and if a language X is found in different language groupings in each tree, this implies that the lexical dataset of language X contains evidence for its affiliation with different languages through different historical processes. If the only relevant process of language change was genealogical descent with modification, such mixed signals would not be found. (This does not mean that absence of evidence for multiple topologies means that genealogical descent with modification is the only relevant process of language change.) Therefore, I draw on horizontal transmission processes as well as incomplete lineage sorting to explain evidence for multiple topologies. Whether this assumption is valid or not is a matter of further theorizing and strict tests that are beyond the limits of the current article.

2. Reticulation in language and culture

Despite the growing acceptance of phylogenetic methods in historical linguistics, it is important to keep in mind their limitations in order to improve upon them if possible. One criticism of phylogenetic methods with both practical and theoretical significance is that language change is not tree-like (Blench 2015; François 2014; Geisler & List 2013; see also references in Atkinson & Gray 2005: 523 and Croft 2008: 228–229). Languages, like any part of culture, do not evolve only through an ancestral language differentiating and splitting up into daughter languages which then diverge and split further; rather, they are also characterized by features taken from languages other than their parents. This is problematic for phylogenetic tree inference if rates of borrowing are high – imagine if language A borrowed features from a related language B and these features are subsequently inherited by the descendants of language A. The descendants of language A will appear more similar to language B and its descendants because of the borrowed features (see Figure 1 for some terminology). This may have an impact on the placement of languages A and B, along with their descendants, in a phylogenetic tree, causing them to be grouped more closely together (Wang & Minett 2005). A case in point is the position of English in the Indo-European family tree, as exemplified in Figure 1. English has borrowed extensively from Latin, North Germanic, and (Norman) French (Algeo & Pyles 2005: 247–268). Taking into account none of the borrowings but only its history within Germanic, the true genealogical position of English is as a sister language to Frisian (Beekes 2011: 29); however, the phylogenetic analysis of Indo-European languages by Bouckaert et al. (2012), displayed with some changes in Figure 1, places it outside of the West Germanic branch.

The non-tree-like character of language change is reflected not only by borrowing after languages have split, but also by the processes through which new languages come into being. François (2014) and Kalyan & François (2018) make a case against the tree model from the perspective of the wave model of language change, which focuses on dialect networks. They argue that the tree model fails to capture what is perhaps the most common language diversification process, namely 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: 163). This scenario is not well represented by a tree structure, as trees cannot deal with the reticulation introduced by extended (and perhaps continued) language contact and borrowing by emerging language clades. Typically, isoglosses do not form neat linguistic subgroups, but show widely varying overlaps.

This reflects the situation in evolutionary biology, as well, where divergences between gene histories and species histories are well known (see Gray, Greenhill

& Ross 2007: 367–368 and List et al. 2016 for these processes in linguistics). Maddison (1997: 523) describes a phylogenetic tree of species as "a cloud of gene histories." Gene trees may not match with species trees because of horizontal transfer (hybridization), incomplete lineage sorting, and gene duplication and extinction (Maddison 1997; see Koonin, Makarova & Aravind 2001 for an introduction to horizontal gene transfer). Incomplete lineage sorting occurs when there are ancestral polymorphisms, which are by chance inherited only partially by different species, suggesting a different history from the species tree (see List et al. 2016: 16 and Jacques & List, this issue). The problem with gene duplication and extinction is similar: if a gene duplicates in an ancestral species, the two duplicates embark on different evolutionary paths. If some of these copies subsequently go extinct, the true history of the duplicated gene cannot be recovered. Rokas & Carroll (2006: 1903) add long branch attraction to the list of processes affecting the recovery of the species tree: when branches for certain species are very long, subsequent changes will overwhelm the historical signal needed to retrieve the species tree (see Holm 2007: 185 for the same point in linguistics).

Processes that generate reticulation affect both biological and linguistic change. Hence, phylogenetic inference can be problematic not simply because words and other linguistic features can be borrowed across lineages, but also because in a wider sense, words have their own unique etymologies (like "gene trees" in biology) that may or may not match with the family tree ("species tree"). Additionally, words can change their denotatum easily, i.e., can become irrelevant for the study of a particular concept due to semantic shift. For example, the English word *hound* is irrelevant when we study Germanic words for the concept 'dog', while its relatives in German (*Hund*) and Dutch (*hond*) are still relevant. Genes, on the other hand, cannot really change their function, as mutations in a given gene are likely to render the gene defunct or make it more efficient but are not likely to change its function altogether. Hence, we expect that linguistics is more greatly impacted by discrepancies between "gene trees" and "species trees." On the other hand, note that if these problems were severe, phylogenies could not be inferred with any accuracy. This is not true: it is known from several well-studied language families such as Indo-European and Austronesian that it is possible to approximate tree topologies established by conventional techniques (Gray, Drummond & Greenhill 2009; Bouckaert et al. 2012, among others).

Researchers in linguistics, anthropology, and biology have considered several solutions to deal with reticulation, both within and outside of the tree model (please note that no attempt has been made to provide an exhaustive review of all these proposed solutions, as the literature on this topic is extensive). Within the tree model, several studies have simulated data with known rates of borrowing and subsequently tested the behavior of phylogenetic tree inference (Nunn, Borg-

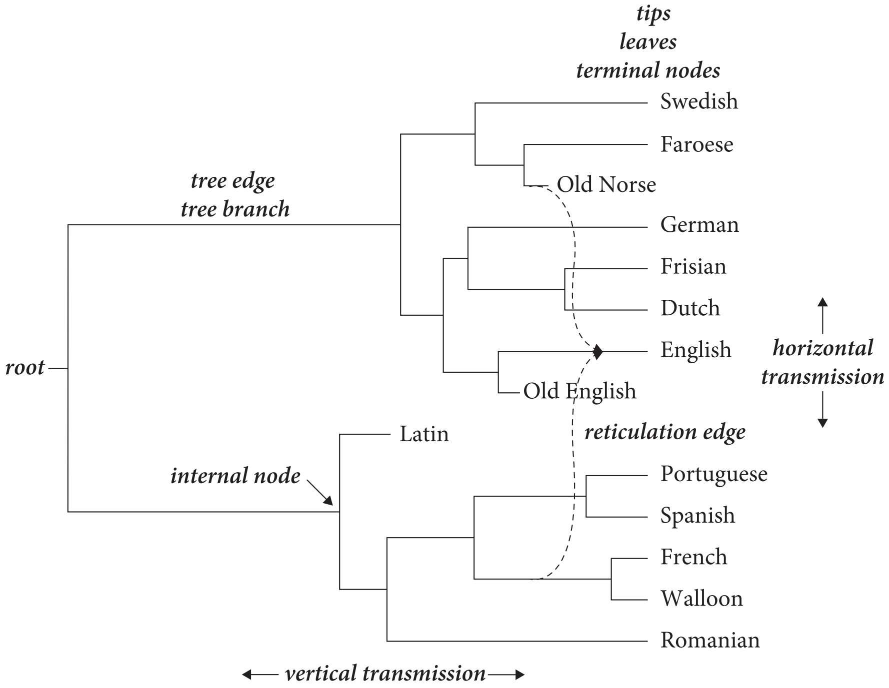


Figure 1. A simplified rooted phylogenetic tree of some Germanic and some Romance languages, illustrating borrowings from the ancestors of the modern Scandinavian languages and French into English during the Old and Middle English period[1]

erhoff Mulder & Langley 2006; Greenhill, Currie & Gray 2009; Nunn et al. 2010). Collard, Shennan & Tehrani (2006) compared biological and cultural datasets using the retention index (RI), a goodness-of-fit measure, and showed that cultural change is neither more nor less tree-like than biological change. Matthews et al. (2011) tested two different models of evolution of Iranian tribal textiles on trees and found that various sets of techniques and designs have different histories of descent. Bowern et al. (2011) showed that hunter-gatherer language lexicons, unlike previously thought, do not have higher rates of borrowing than agriculturalist languages. Tehrani & Collard (2002) demonstrated the importance of tree-like change over borrowing in Turkmen textiles.

Moving beyond the tree model, the most obvious way to incorporate and account for reticulation is through using phylogenetic networks. The case for using them has been made both in biology (Huson & Scornavacca 2010, 2012;

---

[1] The tree topology is based on Bouckaert et al. (2012), a Bayesian phylogeographic analysis of cognate-coded lexical data. The labels indicate terms frequently used in linguistic phylogenetic studies.

Morrison 2016) and linguistics (List et al. 2013). Nakhleh, Ringe & Warnow (2005) presented a method for inferring "perfect phylogenetic networks" by adding a minimal number of lateral edges for characters that do not fit on a set of highly regarded trees. Heggarty, Maguire & McMahon (2010) discussed a network approach based on phonetic divergence. Wichmann et al. (2011) contributed a discussion of measures of reticulation (delta scores and Q-residuals). Nelson-Sathi et al. (2011) analyzed two Indo-European datasets using minimal lateral networks, which add lateral edges to a reference tree for cognate sets that do not fit the tree structure; they found that borrowing is pervasive. List et al. (2013: 147) stated that Nelson-Sathi et al.'s (2011) results are overblown; they applied the same method and found that 369 out of 1,190 cognate sets (31%) were affected by borrowing, considerably less than the 61% found by Nelson-Sathi et al. (2011). Kelly & Nicholls (2017) proposed a stochastic Dollo model that includes lateral transfer in a Bayesian context and tested this model using Eastern Polynesian languages. The most recent application is Willems et al. (2016), who used hybridization networks with an Indo-European dataset; this approach had the advantage of being able to find the directionality of lateral transfer in terms of donor and recipient language. Many of these papers have provided overviews of even more applications of rooted network techniques in linguistics, suggesting that this is a growing area of inquiry.

This article is an exploratory study of a method that may be able to capture reticulation in linguistics. The multiple topologies method infers multiple tree topologies, rather than just one, and is informative with regard to the sites (columns in the dataset; in the current study, cognate sets) that support each topology. In this article, the set of trees inferred always equals two, although more are possible. This technique has the benefit of not requiring a reference tree of the sort used by Nakhleh, Ringe & Warnow (2005), which is required for minimal lateral networks (Nelson-Sathi et al. 2011, List et al. 2013). It was developed in order to account for divergences between gene histories and species histories in evolutionary biology, and therefore it may be suited to account for similar patterns in linguistics. I describe the outcome of applying this tool on four lexical linguistic datasets in Section 4 and speculate on possible non-tree-like processes that could be responsible for the multiple topologies found in Section 5. The four lexical datasets are the following:

i. Austronesian: Gray, Drummond & Greenhill (2009): 400 languages; 210 concepts; 34440 cognate sets. The Austronesian dataset is the largest one considered here, with 400 languages. The authors removed loans from the lexical dataset. Their best-performing model was a covarion model. They produced a dated tree that shows how the expansion of the Austronesian language family is coupled with the expansion of its speakers across the Pacific and that this expansion is probably linked to the development of sea-faring technologies.

ii. Sinitic: Ben Hamed & Wang (2006): 24 languages; 200 concepts. This dataset was downloaded from the Chinese Dialect Database (List 2017), a repository for linguistic information on Chinese dialects, including several lexical datasets. In the database, the dataset used by Ben Hamed & Wang (2006) has been made available as "Wang (2004a)," cited in this article as Wang (2004). In this article I refer to their paper as "Ben Hamed & Wang (2006)" when I compare the results of this study to their phylogenetic results; Wang (2004) is only a reference to the dataset. The only difference between the dataset as used by Ben Hamed & Wang (2006) and the dataset available as "Wang (2004a)" from List (2017) is that the latter does not include Old Chinese. This language was excluded by List, as it should not be part of a contemporary dataset for use in studies of non-tree-like signal. For this reason, the dataset used in the current article contains 23 languages, 200 concepts, and 1511 cognate sets. Ben Hamed & Wang (2006) make clear that a fair amount of reticulation is present in the Sinitic dataset. Their article features an extensive discussion of the applicability of phylogenetic tree inference and network inference methods. Known borrowings were removed from the dataset. The same dataset was subsequently analyzed by List (2015, 2016). Further analysis on the Sinitic languages using a different dataset is performed by List et al. (2014).

iii. Indo-European: Bouckaert et al. (2012): 103 languages; 207 concepts; 5997 cognate sets. The aim of Bouckaert et al. (2012) was to demonstrate an Anatolian origin for the Indo-European language family. In their dataset, known borrowings were removed. Their best performing model was the stochastic Dollo model; their preferred clock was the uncorrelated lognormal relaxed clock. Note that since their model included ancestral inferences of geography, this may have impacted which model and clock was best supported. Since it was not possible to link the columns of this binary dataset to the concepts of the cognate sets, Michael Dunn was kind enough to provide a dataset for which this was possible (Dunn 2018). This dataset can be considered a slightly revised version in terms of cognate coding. It contains 116 languages, 207 concepts, and 6381 cognate sets. The languages that are additional compared to the dataset used by Bouckaert et al. (2012) are listed in the relevant section below.

iv. Japonic: Lee & Hasegawa (2011): 59 languages; 210 concepts; 675 cognate sets. The authors found that the covarion model in combination with a relaxed clock produced the best fit. They produce a dated tree. The main claim of their article is that the arrival of the Japonic languages coincided with the arrival of agriculture in what is now Japan. There is considerable uncertainty in their tree sample, however. While the higher internal nodes are well resolved, many lower internal nodes on the maximum clade credibility tree have a posterior probability of less than .5 (Lee & Hasegawa 2011: 4). See Lee (2018) for the dataset and resulting tree set.

These four datasets were chosen because of their diverging features in terms of size and evolutionary history. Austronesian and Indo-European are large language families, and Indo-European especially is well-studied. Both of these families are of continued interest among phylogeneticists, and new phylogenetic analyses featuring more languages will be presented in the future. The Sinitic and Japonic datasets feature dialects and languages, but for the sake of convenience all language varieties that are differentiated are referred to as "languages." These languages may involve more reticulation (a fact which is definitely true for the Sinitic dataset, as described by Ben Hamed & Wang 2006) and a different type of evolutionary history in comparison to Austronesian and Indo-European.

In these four datasets, known borrowings were removed (Lee & Hasegawa 2011: 3667 do not explicitly state this, but suggest that they have done so in their discussion, in which they point out that knowing whether one has indeed removed all loans is impossible). The goal of this article is to explicitly investigate these datasets, because as both these papers and the results in Section 4 show, evidence for reticulation is still present despite removal of known loans. We know that languages are full of loanwords in general (Tadmor 2009); Nelson-Sathi et al. (2011) and List et al. (2013) have shown undetected borrowings to be present in earlier Indo-European lexical datasets, and I side with Lee & Hasegawa (2011) in agreeing that detecting borrowings can be very difficult (see List et al. 2016: 7 for perspectives on automatic loanword identification). Aside from borrowing, there are other processes, such as incomplete lineage sorting, that result in non-tree-like signal. Therefore, the current article is an application of the multiple topologies method to datasets typically used for phylogenetic inference, which as far as we know typically contain non-tree-like signal despite removal of known loanwords. It would additionally be useful to compare the current results to an analysis of the same dataset with known loans kept in, in order to see whether similar non-tree-like signal is found but more pronounced, or whether instead other signals appear. This is, however, beyond the limits of the present article.

3. Multiple tree topologies

The multiple tree topologies method is implemented in the software *BayesPhylogenies* (Pagel & Meade 2004; n.d.). This software provides a Bayesian MCMC

(Monte Carlo Markov chain) framework for inferring trees and estimating parameters of evolutionary models for a variety of different data types, including binary lexical data. Bayesian MCMC methods for phylogenetic tree inference are increasingly common for linguistic phylogenetic inference and were used in the analysis of the Austronesian, Indo-European, and Japonic datasets presented above. They have proven so popular because they present a practical solution to finding phylogenetic trees that are well supported by the data (Dunn 2014). The reader can refer to Pagel & Meade (n.d.), Huelsenbeck et al. (2001), Ronquist, van der Mark & Huelsenbeck (2009), Dunn (2014), and Drummond & Bouckaert (2015) for introductions to Bayesian MCMC methods and various software packages that implement them.

In the multiple tree topologies approach, regular MCMC phylogenetic inference is extended using a mixture models approach (Pagel & Meade 2004) in order to estimate two topologies at the same time (Pagel & Meade 2006a). This method can identify different evolutionary histories that may be found in subsets of the dataset. A technical description of the method is not given here; rather, I present an example of the workings of the method using simulated data. Figure 2 presents two random trees, both with 50 language taxa, labeled T1 through T50, in which T50 is the outgroup.

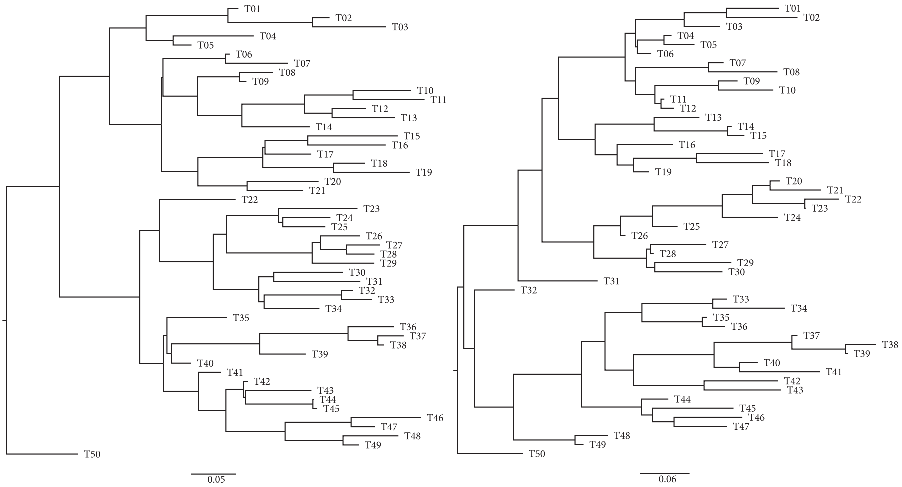


Figure 2. Two random topologies with 50 languages, labeled T1 through T50. Left: tree 1; right: tree 2

The evolution of 10,000 binary traits was simulated across the branches of these two trees using the continuous-time Markov model for discrete trait evolution: 7,500 sites on tree 1 and 2,500 sites on tree 2. After this, phylogenetic tree sets were reconstructed on the basis of these simulated data: once estimating only one topology, and once estimating two topologies. The results are presented in Figures 3 and 4.

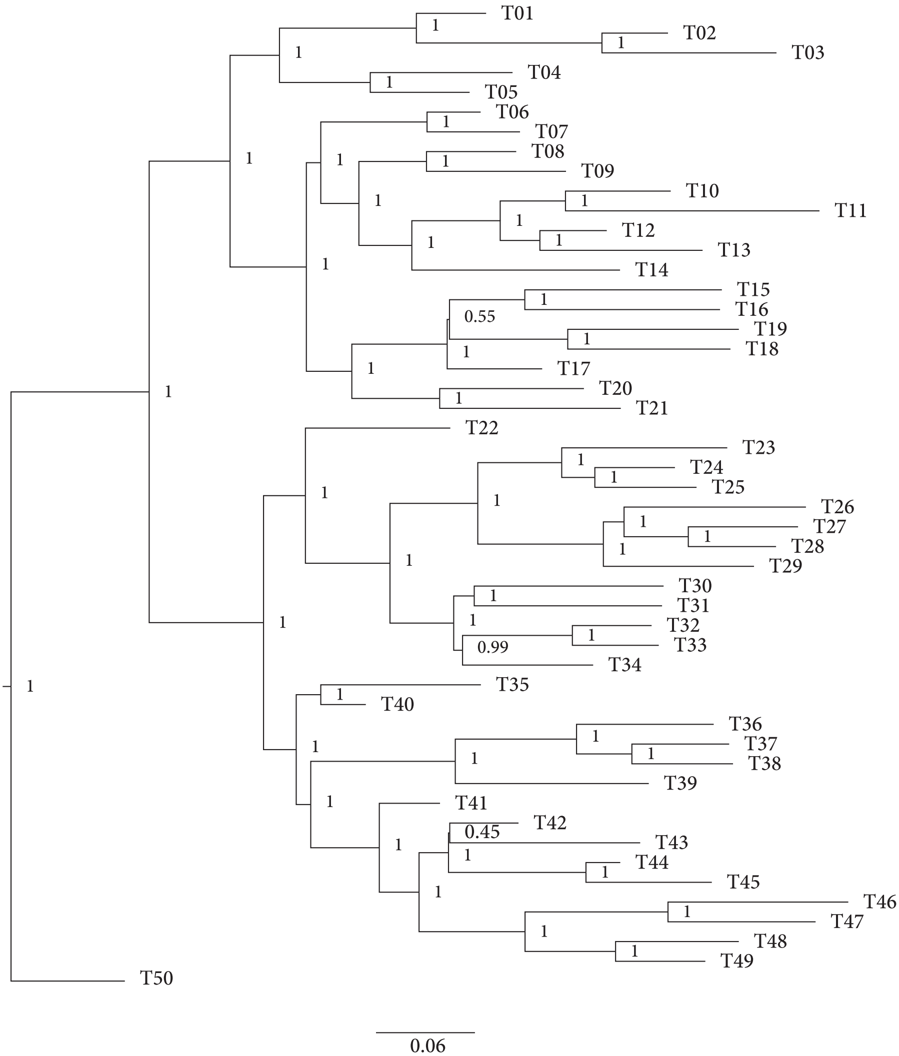


Figure 3. Maximum clade credibility tree of the results of the one topology model on 10,000 simulated binary sites

Figure 3 presents the maximum clade credibility (MCC) tree of the reconstructed tree set when only a single topology is estimated. It looks remarkably similar to the tree used to simulate 7,500 sites and bears no traces of the groupings found in the tree used to simulate the remaining 2,500 sites. There are a few differences between the trees. One difference is the placement of T17, which in the tree used to generate the data is sister to a group of T15 and T16; in the reconstructed tree, it is sister to a larger (poorly supported) group containing T15, T16, T18, and T19. This is not surprising, given the very short branch leading to T17 in the original tree. T40 is sister to a group containing T36, T37, T38, and T39 in the original tree, but in the reconstructed phylogeny, it groups with T35 – again, because it has a very short branch in the original tree. Other than those two differences, support values are high, and no one would from this single topology guess that 25% of the sites used for this reconstruction have an entirely different evolutionary history from that represented by the maximum clade credibility tree in Figure 3.

Figure 4 gives maximum clade credibility trees of the reconstructed tree sets when estimating two topologies. These trees simulate the data with great accuracy: the majority tree has greater support with 76% of the sites, while the minority tree has lower support, 24% of the sites. Not only is the majority tree identical in topology to the original tree that was used for the simulation of the 7,500 sites (T17 and T40 are placed where they were in the original tree), the minority tree captures the topology of the tree used to generate the 2,500 sites with different evolutionary history. Of course, language change will never be as clear-cut as a simulated binary dataset, especially since binary cognate-coded lexical data (the most common type used) is originally multistate and only made binary for the purpose of phylogenetic inference. Likewise, phylogenetic inference models are at best very rough approximations of reality, and would be so even if our data were as noise-free and well-sampled as simulated data. A better simulation would perhaps be to let multistate characters evolve up the branches of the tree, then convert the simulated data to binary before using it for reconstructions (see Pagel & Meade 2006b: 175 and Chousou-Polydouri et al. 2016 on multistate versus binary phylogenetic inference). However, one would then have to specify aspects of multistate character evolution (such as the probability of emergence of new states) that would go beyond the purpose of this article. Therefore, the intention of this simulation is simply to show that the multiple topologies method can in principle recover evolutionary histories that would otherwise be overwhelmed by the patterns found in the majority of the dataset. More specific simulations regarding lexical data should be conducted in the future.

There are different patterns that could be identified by the multiple topologies methods. It could be the case that no alternative histories are present, in which case the minority tree will have no highly supported groupings, basically returning "noise." Another possibility is that there might be an alternative grouping of only a few languages, so only a few highly supported groupings would be found in the minority tree while the rest of the tree would have low support, once again

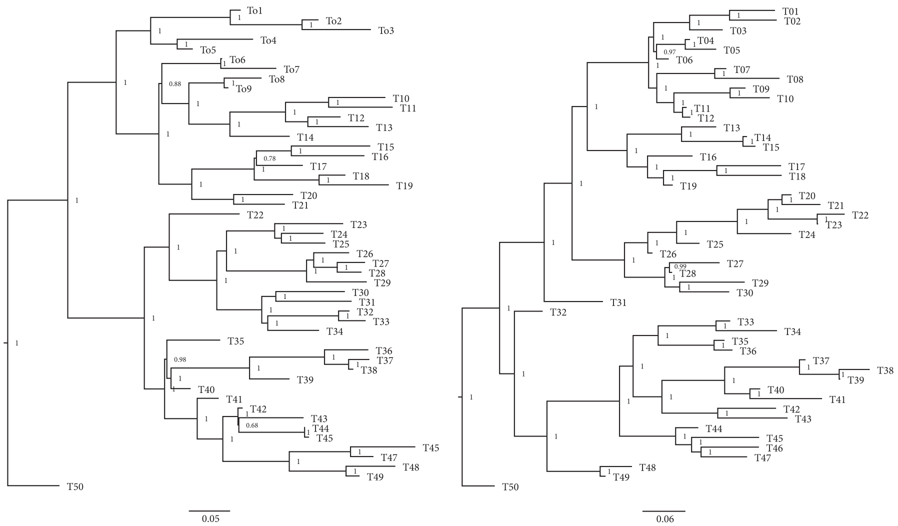


Figure 4. Maximum clade credibility trees of the results of the two-topology model on 10,000 simulated binary sites. Left: The majority tree, supported by 76% of the sites; right: The minority tree, supported by 24% of the sites

representing noise. One might find that the minority topology picks up on differences in rate, in which case the two tree topologies would be similar but have different branch lengths. All of these different outcomes can be caused by different evolutionary processes, making interpretation potentially difficult. Given that the current study is the first application of the multiple topologies method on linguistic data, all findings must be considered tentative; indeed, in many cases, it is not possible to pinpoint the cause underlying the non-tree-like signal that this method recovers.

It is possible to use the multiple topologies method to assess the log-likelihood of each site (cognate set) for each tree in the posterior sample of both the majority and the minority tree set. What it means to be the majority tree is to have higher log-likelihoods than the minority tree has for at least half of the sites. This functionality is used below to identify cognate sets that are associated specifically with the minority and majority tree and to see how they support particular language groupings.

Before describing the behavior of the multiple tree topologies method in analyzing the four lexical datasets examined in this article, a few technical details need to be mentioned. Eight models of evolution were tested (for further information see the BayesPhylogenies manual, Pagel & Meade n.d.; for an explanation of covarion and gamma rate heterogeneity see Ronquist, van der Mark & Huelsenbeck 2009):

1. standard two-state model (standard)
2. two topologies (top2)
3. covarion (cov)
4. covarion + two topologies (cov+top2)
5. gamma rate heterogeneity (gam)
6. gamma rate heterogeneity + two topologies (gam+top2)
7. covarion + gamma rate heterogeneity (cov+gam)
8. covarion + gamma rate heterogeneity + two topologies (cov+gam+top2)

Perhaps comparing these eight models is not necessary – gamma rate heterogeneity is a model well-suited for genetic data, but it normally does not improve the fit of phylogenetic analysis of linguistic data. Phylogenetic analysis of linguistic data typically performs well with the covarion model (see Gray, Drummond & Greenhill 2009; Lee & Hasegawa 2011; Bowern & Atkinson 2012, and Grollemund et al. 2015 for examples of phylogenetic analyses of lexical datasets in which a covarion model outperforms a gamma rate heterogeneity model; see Chang et al. 2015: 216 for general comments). However, in order to see the interaction between these three different components, all combinations of yes/no covarion, yes/no gamma rate heterogeneity, and yes/no two topologies were investigated. Each model was run for 120 million iterations with a 20 million burn-in; every 100,000th iteration was sampled, creating posterior tree samples of 1,000 trees. Five duplicate analyses were carried out for each model to ensure convergence. No clock, calibration points, or outgroups were provided in order to make the analyses comparable between language families. Nor were trees forced to be ultrametric, i.e. to have the same root-to-tip distance for all tips, as this would be inappropriate for the Indo-European and Japonic datasets, which include non-synchronous language varieties. The result, then, is tree samples in which branch length reflects amount of lexical change rather than chronological time. Appropriate sampling was assessed by measuring the Pearson correlation between iteration and log-likelihood, which should be close to zero. Convergence of each chain was assessed using a linear regression of iteration and log-likelihood, where the β coefficient should be as small as possible and not show a significant downward or upward trend. In Section 4, the converged run of each model with the highest marginal log-likelihood as estimated by stepping stone analysis (Xie et al. 2010) that was conducted after each run is given. The input files for each analysis are shared in the supplementary information files, available on Zenodo (zenodo.org/record/2653209).

4. Results

All result files have been made available as supplementary information, found on Zenodo (zenodo.org/record/2653209).

4.1 Austronesian

Table 1 gives an overview of the results of the different models of evolution for Gray, Drummond & Greenhill's (2009) lexical dataset drawn from 400 Austronesian languages. The models are ordered by rank in the second column; the first row details the best-supported model, the last the worst-performing model. Model performance was assessed using marginal log-likelihoods as estimated by stepping stone analysis (Xie et al. 2010). The third column gives the difference between models in log-likelihood units. The fourth and fifth columns contain the weight associated with each topology for those models in which two topologies were estimated. This represents the number of cognate sets associated with each topology.

Table 1 shows that the cov + gam + top2 model is the best-performing model in terms of marginal log-likelihood. However, it is not altogether clear why this is the case: the weight associated with the minority tree, i.e. the least supported tree, is very low (0.005), so it is unclear what having two topologies adds to the cov + gam model, which scores third. The majority topologies have several well-supported clades and capture the well-known structure of the Austronesian family – these are not considered here in detail due to lack of space, as the current interest is finding evidence for non-tree-like signal. The minority topology is assessed by looking at clades supported at a score of over .5; these clades represent language groupings that are present in at least 50% of the 1000-tree sample. The only analysis for which internal nodes in the minority tree have a posterior support larger than 1% is for the model with two topologies (no covarion, no gamma rate heterogeneity), in which some well-supported internal nodes of pairs of languages are found in the minority topology. This is probably due to the minority topology capturing some aspect of lexical change that is normally captured by the covarion or gamma rate heterogeneity models.

The low weightings for the minority trees across all models, plus the lack of well-supported clades in the minority trees for the top-ranking models, make it clear that Gray, Drummond & Greenhill's (2009) Austronesian dataset does not show any significant support for a second topology. This is highly surprising, as it is well-known that the Polynesian languages likely originated from a dialect chain break-up (Gray, Bryant & Greenhill 2010 and references therein; Kelly & Nicholls 2017). Taking Eastern Polynesian as an example, a language group which was also

Table 1. Model comparison for Gray, Drummond & Greenhill's (2009) lexical dataset on 400 Austronesian languages

Model	Marginal l-lh [a]	Marginal l-lh difference [b]	Weight majority tree [c]	Weight minority tree
cov + gam + top2	−280975		0.995	0.005
gam + top2	−281462	487	0.995	0.005
cov + gam	−281544	82	–	–
gam	−281949	405	–	–
cov + top2	−282265	316	0.994	0.006
cov	−282840	575	–	–
top2	−285127	2122	0.98	0.02
standard	−286478	1516	–	–

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs
b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row
c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

examined by Gray, Bryant & Greenhill (2010: 3926) and Kelly & Nicholls (2017: 12), we might expect to find Marquesic and Tahitic groupings in one of the two trees, with the other tree reflecting the affinity between Hawaiian or Mangareva with Tahitic; we might also expect to find Rapanui with Marquesan, Mangareva, and Penrhyn. So while we would not see a reconstruction of the dialect chain itself, the multiple topologies method would show divergent signals in the different trees.

4.2 Indo-European

As stated in Section 2, I used a different, updated version of Bouckaert et al.'s (2012) Indo-European lexical dataset kindly provided by Michael Dunn. This dataset has updated cognate coding, and it incorporates 116 rather than 103 languages. The languages present in addition to those in the original dataset are Proto-Indo-European, Albanian_T, Ossetic, Greek_K, Greek_D, Greek_Md, Khaskura, Panjabi_ST, Lithuanian_O, Afrikaans, Pennsylvania_Dutch, Brazilian, and Irish_B. These mostly represent variants of languages otherwise included, many of which have their origin ultimately in the database by Dyen, Kruskal & Black (1992). I will refer to this dataset as "Bouckaert et al.'s (2012) updated lexical dataset."[2]

---

[2] Some names in the updated dataset are different. They are as follows: Persian_List = Persian; Nepali_List = Nepali; Romani = Gypsy_Gk; Riksmal = Norwegian; English_ST = English; German_ST = German; Romanian_List = Rumanian_List.

Because it proved difficult to get some of the models to converge, we performed sets of 50+ shorter runs for some models. For the cov + gam + top2 and the cov + top2 models, we used 30 million iterations and a 10 million iteration burn-in, sampling every 20,000th iteration. All other models were run (as stated above) with 120 million iterations and a 20 million iteration burn-in, sampling every 100,000th iteration.

Table 2 gives an overview of the results of the different models of evolution for Bouckaert et al.'s (2012) updated lexical dataset of 116 Indo-European languages. The results for Indo-European look very different from the Austronesian results. The cov + gam + top2 model performs best. In all models that estimate two topologies, the minority topology is quite well supported, with support from 23% to 40% of the sites (cognate sets). Bouckaert et al.'s (2012) maximum clade credibility tree (S1) is reproduced as Figure 5. To first compare Bouckaert et al.'s (2012) results to a model without two topologies, the maximum clade credibility tree of the cov + gam model (the best-supported model without two topologies) is included in Figure 6. What follows are the maximum clade credibility trees of both the minority and majority tree sets for the best-performing cov + gam + top2 model in Figures 7 and 8.

Table 2. Model comparison for Bouckaert et al.'s (2012) updated lexical dataset on 116 Indo-European languages

Model	Marginal l-lh [a]	Marginal l-lh difference [b]	Weight majority tree [c]	Weight minority tree
cov + gam + top2	−51167		0.60	0.40
cov + top2	−51182	15	0.65	0.35
gam + top2	−51375	193	0.77	0.23
cov + gam	−51450	75	–	–
cov	−51524	74	–	–
gam	−51791	267	–	–
top2	−51832	41	0.64	0.36
standard	−52812	980	–	–

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs
b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row
c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

To compare an analysis without two topologies to the maximum clade credibility tree presented in Bouckaert et al. (2012), we can compare Figure 5 and Figure 6. In this comparison, we see the following:

1. Higher order subgrouping. Starting from the top, both trees have a Celtic-Romance clade, which is sister to Germanic. The Bouckaert et al. (2012) tree topology then features Balto-Slavic, Indo-Iranian, and the remaining clades in a stepwise fashion. The cov + gam MCC tree has a main split between Celtic-Romance-Germanic and the other subgroups, but this lower major clade has low support for its main splits, e.g. Balto-Slavic and Indo-Iranian (0.32), Balto-Slavic+Indo-Iranian and Albanian-Greek (0.24).
2. Indo-Iranian. The position of Vedic Sanskrit, Old Persian, and Avestan in the cov + gam model deviates from the position of these languages in the Bouckaert et al. (2012) tree. Old Persian and Avestan are sisters and together are a clade sister to Vedic Sanskrit. Wakhi (an Eastern Iranian language) is not in the right place. The Indo-Aryan clades in both trees are similar with minor differences the positions of smaller groups. Interesting is the position of Romani, called Gypsy_Gk in the updated dataset, as an outlier in Bouckaert et al. (2012) but as sister to Kashmiri in the cov + gam tree.
3. Slavic. Polish is not placed correctly (with other West Slavic languages Czech and Slovak) in either tree. Old Church Slavonic is pulled outside of the contemporary Slavic group in the cov + gam tree, while it is correctly placed in the South Slavic group in Bouckaert et al. (2012).
4. Germanic. In the cov + gam tree, Old High German is placed as a sister to West Germanic rather than as a sister to German and Luxembourgish.
5. Romance. The two analyses have the same smaller subgroups within this clade, but their grouping is different. In the cov + gam tree, Umbrian and Oscan are sisters to Latin rather than being placed outside of the clade containing Latin and the modern Romance languages. Sardinian is not a sister to the Italian subgroup.
6. Remaining smaller groups. Tocharian is sister to the Anatolian languages in the cov + gam tree, rather than to the Armenian languages in Bouckaert et al. (2012). The Armenian languages show a rather well-supported affiliation with the Hellenic languages instead.

Note that some of these changes may have been affected by aspects of the phylogenetic analysis by Bouckaert et al. (2012) that were not included in the current analyses, including calibration points on language divergence times and the inference of ancestral geographical locations.

The following is a list of points regarding the minority and majority trees of the cov + gam + top2 model, starting at the top (Celtic) and ending at the bottom (Anatolian), comparing them to Bouckaert et al.'s (2012) maximum clade credibility tree in Figure 5.

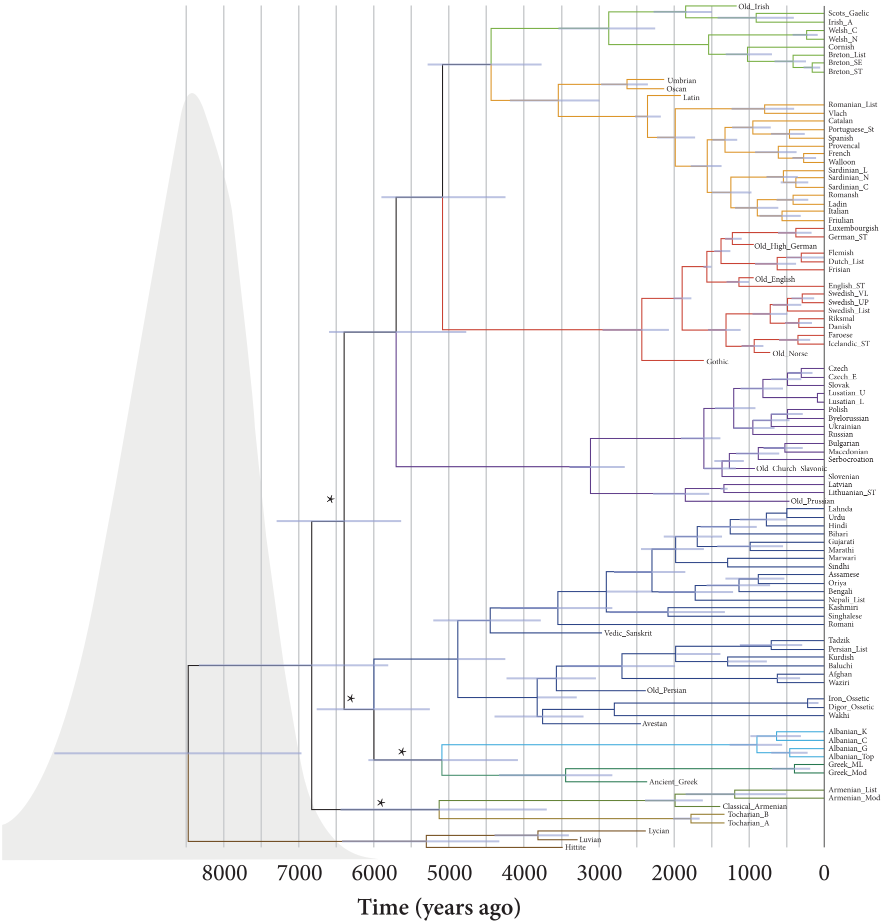


Figure 5. Bouckaert et al.'s (2012) dated maximum clade credibility tree for 103 Indo-European languages. From Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337.957–960. Reprinted with permission from AAAS

1. Higher order subgrouping. Neither the minority or the majority tree has the same subgrouping of major branches as the Bouckaert et al. (2012) maximum clade credibility tree has; this is especially true regarding the placement of non-contemporary languages. In the minority tree, Luvian, Lycian, Umbrian, and Oscan form a clade that is sister to Celtic. This tree also deviates by pulling Germanic outside of the Celtic-Romance-Germanic group, albeit with very low posterior probability (0.20). Albanian is placed as sister to Romance. In the majority tree, Oscan and Umbrian are also sister to Celtic, again with

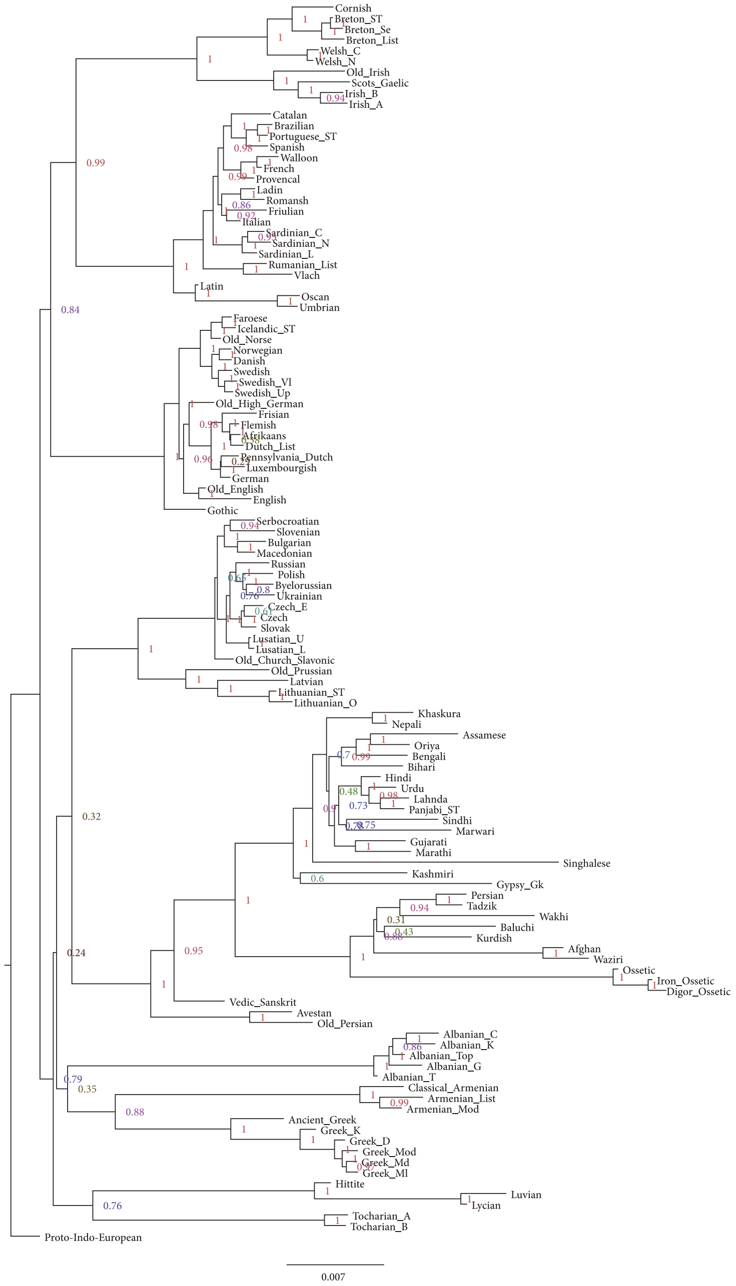


Figure 6. The maximum clade credibility tree for the cov + gam model with the highest marginal log-likelihood from Bouckaert et al. (2012)

very low posterior probability (0.20); this may reflect uncertainty regarding their placement within Celtic-Romance.

2. Celtic. The structure is mostly similar in both the majority and the minority tree and similar to the structure in the Bouckaert et al. (2012) tree. In the majority tree, Cornish and Welsh are sisters, while the minority tree captures the received view that Cornish and Breton are sisters.
3. Romance. In the minority tree, Lycian and Luvian (Anatolian) form a group with Umbrian and Oscan, an odd result with relatively high support: a cursory inspection of the dataset suggests that this is probably caused by a substantial amount of missing data for the same concepts in these four languages. In the minority tree, Catalan is pulled away from Portuguese and Spanish. Other than that, its structure is very similar to that of the cov + gam MCC tree in Figure 6. The majority tree does not recover this structure and has very short branch lengths in part.
4. Germanic. The minority tree has the same topology as the cov + gam MCC tree in Figure 6, with the sole exceptions of the position of Old Norse, which is pulled outside of Northern Germanic, and the slightly more outward position of Frisian. The majority tree does not capture the division between Northern and Western Germanic and looks quite messy. Gothic is pulled inside the clade as sister to Old High German.
5. Slavic. In the minority tree, Old Church Slavonic is placed correctly in South Slavic (albeit with very low posterior probability, 0.23), an improvement upon the cov + gam MCC tree in Figure 6. Lusatian is sister to Russian (again with low posterior probability, 0.23). Polish is again not placed correctly (as it should be placed with other West Slavic languages Czech and Slovak). As was also true for Romance and Germanic, the majority tree does not capture the internal structure of the Slavic subgroup well. Russian is pulled outside as a sister to the entire Slavic group, and South Slavic is not a clade. What it captures rightly is Lusatian as sister to Polish.
6. Iranian. Avestan is not placed correctly (as sister to Ossetic) in either tree; it is attracted to Old Persian in the majority tree and outside of the Indo-Iranian clade in the minority tree. The placement of Wakhi is problematic in both trees (most likely because there are no languages that are closely related to Wakhi included in the dataset). Old Persian is sister to (Modern) Persian in the minority tree.
7. Indo-Aryan. The structure of this group in the minority tree is much like a stepladder, with several languages being pulled outwards from their position in Bouckaert et al.'s (2012) MCC tree, including Sindhi and Singhalese. While the position of Vedic Sanskrit is fine in the minority tree, it is pulled to a position as ancestor to the entire Indo-Iranian clade in the majority tree. The

grouping structure of the majority tree is closer to Bouckaert et al.'s (2012) MCC tree, although the placement of Nepali and Khaskura seems off.

8. The backbone of the tree. Bouckaert et al.'s (2012) maximum clade credibility tree has a stepladder structure, with Anatolian splitting off first, Armenian-Tocharian second, Indo-Iranian+Albanian-Greek third, Balto-Slavic fourth, Germanic fifth, and Romance-Celtic last. While most of these internal nodes have high posterior probabilities (0.80 or 1), two nodes have lower posterior probabilities: the node connecting Indo-Iranian and Albanian-Greek (0.46) and the node connecting this group with the Balto-Slavic-Germanic-Romance-Celtic group (0.49). Neither the majority nor the minority have this same structure, although the majority tree is closer to it. The posterior probabilities of the backbone nodes differ from those of Bouckaert et al. (2012), with those of the minority tree being particularly low.

The impression from studying these two maximum clade credibility trees is that: (i) while the minority tree represents the structure of the Romance, Germanic, and Balto-Slavic groups better than the majority tree, it is the other way around for the Indo-Iranian group; (ii) the majority tree captures the higher-order subgrouping better than the minority tree (with the exception of the placement of Oscan and Umbrian), while the minority tree capitalizes on uncertainty with regard to a small set of languages, including Lycian, Luvian, Oscan, Umbrian, Albanian, and Tocharian; and (iii) the majority tree has markedly short branch lengths within Celtic, Romance, Germanic, and Balto-Slavic, suggesting that the minority and majority tree capture different cognate sets changing at different rates.

Now that the differences in tree topologies of the minority and majority tree have been discussed, I can attempt to explain the high support for two topologies in the Bouckaert et al. (2012) updated dataset. This can be helped by an assessment of the cognate sets that support each tree topology. Using the command "sitelh" in *BayesPhylogenies*, the log-likelihood that each cognate set evolved on the majority and the minority tree can be calculated. An additional analysis using the highest-ranking cov + gam + top2 model was conducted adding the "sitelh" command. The results are very similar to those reported in Table 2 and Figures 7 and 8; they have been included in the Appendix for reasons of space.

In total, Bouckaert et al.'s (2012) updated dataset of 116 Indo-European languages includes 6381 cognate sets. Given 1001 trees in the posterior sample, the log-likelihood of, in total, $6381 \times 1001 = 6387381$ cognate sets is assessed during the site log-likelihood analysis. The log-likelihood difference between the minority and majority tree is assessed by subtracting the log-likelihood value for the majority tree from the log-likelihood value for the minority tree, with any positive values implying greater support for the minority tree. In calculating the differences,

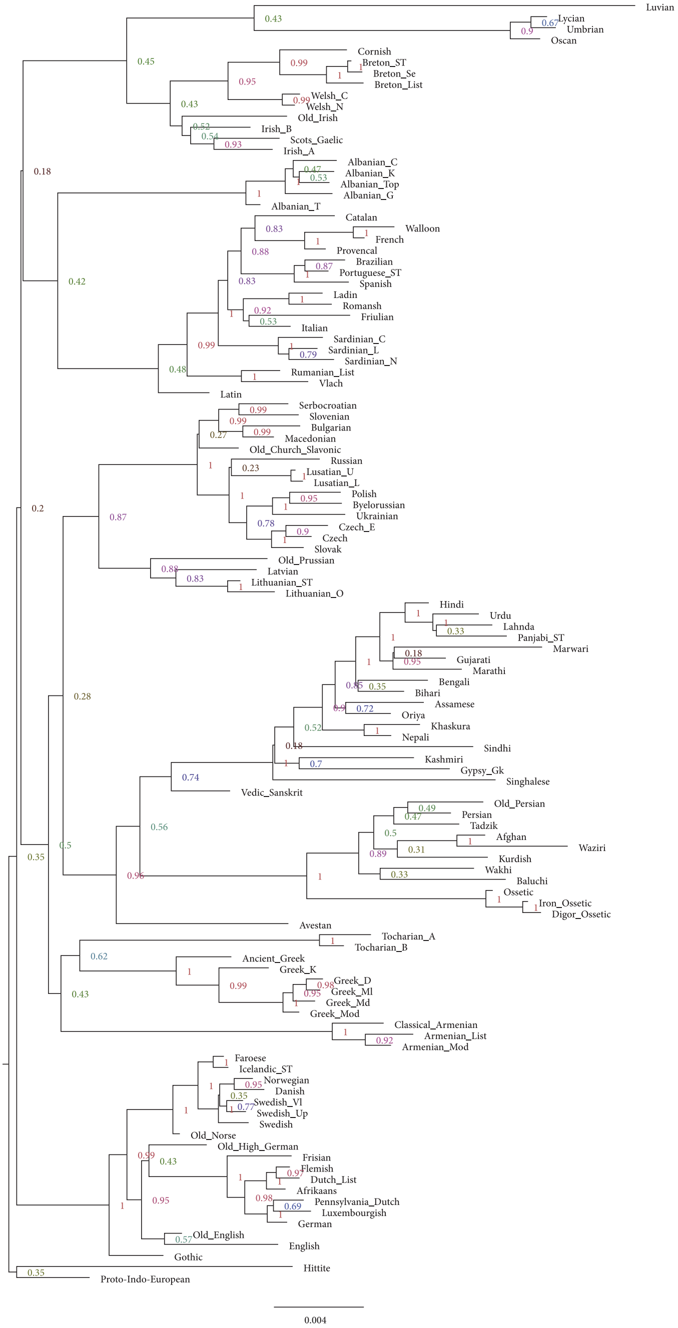


Figure 7. The minority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Bouckaert et al. 2012)

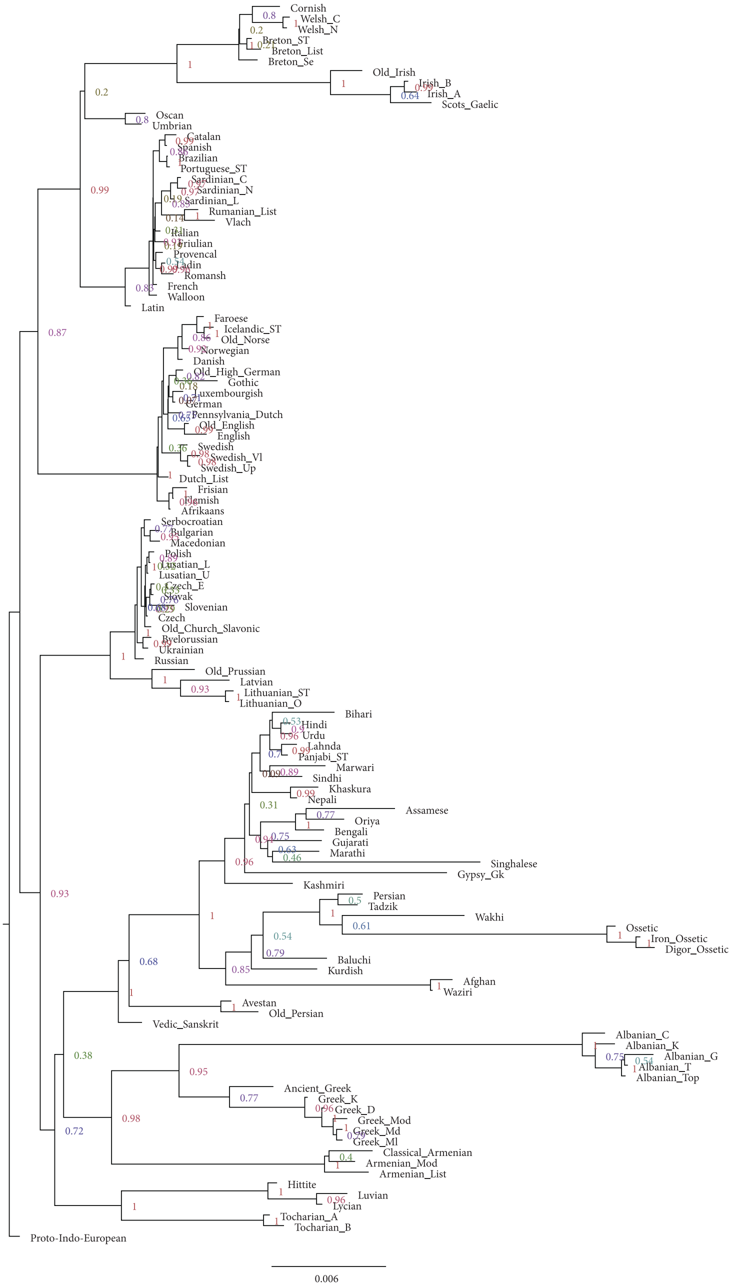


Figure 8. The majority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Bouckaert et al. 2012)

we find that almost all of the individual cognate sets support the minority tree at times; at other times, they support the majority tree. This variation occurs with different trees of the posterior sample (support for minority tree: 6045 cognate sets; support for majority tree: 5709 cognate sets). The log-likelihood differences range from 0 to 40.0 (median minority tree: 0.51; median majority tree: 0.39). So most of the support log-likelihood differences are small and inconsequential, a fact that is also indicated by the fact that almost all cognate sets support different tree topologies in different trees of the sample. When we look at log-likelihood differences of support of 5 or larger, we find 586 cognates sets that strongly support the minority tree and 97 cognates sets that strongly support the majority tree. These are almost entirely mutually exclusive, i.e. these are the cognate sets that really fit only with the minority tree or only with the majority tree, not with both. It may seem strange that there are more cognates sets strongly supporting the minority tree while it is weighted less strongly overall than the majority tree (see Appendix, support majority tree: 0.55; support minority tree: 0.45): this is because we are looking at the log-likelihood differences of 5 or larger taken by cognate set, not at all log-likelihood scores. So where the minority tree has more cognate sets with strong support, the majority tree is associated with more cognate sets overall, which favor it with small or medium support.

To see what kind of topologies emerge from the 586 cognate sets that support the minority tree and the 97 cognate sets that support the majority tree, two additional nexus files were created and two Neighbor Net analyses (Bryant & Moulton 2004, conducted in *SplitsTree*; Huson & Bryant 2006) were conducted. These have been placed in the Appendix for reasons of space (Figures A.3 and A.4). The Neighbor Net analyses show a pattern similar to that of the two topologies analysis, a conclusion we can draw now that we know which cognate sets are responsible for a large part of the two separate signals. Some aspects of the minority and majority tree topology that were remarked upon above are reduplicated:

1. Luvian, Lycian, Umbrian, and Oscan form a clade in the minority Neighbor Net, although this clade is inside Indo-Iranian and not sister to Celtic, as in the minority tree in Figure 7.
2. Catalan seems to be pulled away from Spanish, Portuguese, and Brazilian (Portuguese) in the minority Neighbor Net, just like in the minority tree.
3. The four old Germanic languages (Gothic, Old English, Old High German, and Old Norse) form a clade in the majority Neighbor Net; their attraction is evident in the majority tree from Old High German and Gothic forming a (badly supported) group. In the minority Neighbor Net, Gothic, Old English, and Old High German form a group.

4. The placement of Old Church Slavonic is strange in both the minority Neighbor Net (in a group with Classical Armenian, Pennsylvania Dutch, and Waziri) and the majority Neighbor Net (in a group with Classical Armenian and Pennsylvania Dutch).
5. Four other old or ancient languages – Proto-Indo-European, Vedic Sanskrit, Avestan, and Old Persian – form a clade in the majority Neighbor Net; their attraction is evident in the majority tree based upon the fact that Avestan and Old Persian form a group with Vedic Sanskrit just outside on the next level of a ladder-like structure. Proto-Indo-European, Vedic Sanskrit, and Avestan also form a group in the minority Neighbor Net.

The cognate sets associated most strongly with the minority and majority trees generate Neighbor Nets in which some of the old and ancient languages mentioned above cluster together. The reason for this could potentially be methodological. Cognate coding is usually done with the help of etymological dictionaries and other materials that discuss the history of words. Etymological dictionaries of ancient languages often focus on relationships of words with other ancient languages: i.e., in an etymological dictionary of Gothic, we would expect to find whether a given word has a reflex in Greek, Latin, Old English, Old Norse, or Old Irish, for instance. Contemporary languages, especially non-Germanic ones, would not be featured in this dictionary (except for aberrations as early borrowings into Romance, etc.). By using material from different etymological dictionaries, it is usually possible to find the relevant word histories for both ancient and contemporary languages; nevertheless, the focus on non-contemporary languages in such dictionaries may produce a greater number of cognates between old or ancient languages.

The Neighbor Nets also reduplicate the difference in branch length found in the minority and majority trees. The minority Neighbor Net has long branch length for Germanic, Romance, and Slavic, while the majority Neighbor Net has much shorter branches for these groups, more in line with typical branch length found in the rest of the network. The branch length differences are very striking when comparing the minority and majority trees and can be related to differences in the rate of evolution of the cognate sets most strongly associated with them: the longer branches of the minority tree relate a greater amount of lexical change. Bouckaert et al.'s (2012) updated dataset includes 207 concepts. The 97 cognate sets that show high support for the majority tree feature 79 unique concepts, while the 586 cognate sets with high support for the minority tree feature 163. The concepts associated with the minority tree have significantly higher rates of lexical replacement (mean 3.49) than those associated with the majority tree (mean 2.85) (t-test, $t = -2.65$, $p > 0.01$; rates of lexical change taken from Pagel, Atkinson & Meade

2007). This suggests that the minority topology, in addition to showing differences in language grouping as discussed above, also captures faster-changing concepts in Germanic, Romance, and Slavic.

The reason for finding support for two topologies that are similar in terms of major subfamily grouping but have differences in branch length and topology in Bouckaert et al.'s (2012) updated lexical dataset must, then, be primarily that there is a significantly large group of cognate sets in Germanic, Romance, and Slavic that change at markedly higher rates of evolution. The differences in topology are at least partly indirectly caused by that – the minority topology captures more detailed information on language relationships within these three groups because the cognate sets associated with it display more change, while the majority tree captures the higher-order subgrouping of the subfamilies better. Close associations between old and ancient languages emerge in both the minority and majority trees and may point to methodological aspects of cognate-coding. The Luvian-Lycian-Umbrian-Oscan clade found in the minority tree is probably caused by missing data for a highly overlapping set of concepts in these four languages, but it serves as a good illustration of the multiple topologies method, i.e. that such a distinctive pattern is picked up on by only one of the trees. For the Indo-European analysis, no clear indication of the role of reticulation is found (despite earlier findings such as Nelson-Sathi et al. 2011 and List et al. 2013). This is not the case for Japonic and Sinitic, as is discussed next.

4.3 Japonic

Table 3 gives an overview of the results of the different models of evolution for Lee & Hasegawa's (2011) lexical dataset for 59 Japonic languages.

The best-scoring model for the Japonic dataset is the cov + gam + top2 model. That model has limited support for two topologies, with the minority tree only getting a 0.04 weight. The other models, however, show more support for the minority topology. The cov + gam + top2 model results are discussed first, then compared to the cov + top2 model, which is the second-best scoring model. The reason for this is that the cov + top2 model has much larger support for the minority tree than the cov + gam + top2 model; it is interesting to see what causes this difference.

A maximum clade credibility tree of both the minority and majority trees of the highest scoring cov + gam + top2 model is included in Figures 10 and 11, respectively. First, I carry out a comparison of the majority tree with Lee & Hasegawa's (2011) tree (their Figure 2, here reproduced as Figure 9). The major split in the Japonic languages is one between the Ryukyuan languages and the mainland Japanese languages. This split is likewise found in the majority tree.

Table 3. Model comparison for Lee & Hasegawa's (2011) lexical dataset for 59 Japonic languages

Model	Marginal l-lh [a]	Marginal l-lh difference [b]	Weight majority tree [c]	Weight minority tree
cov + gam + top2	−6759		0.96	0.04
cov + top2	−6783	24	0.81	0.19
cov + gam	−6855	72	–	–
gam + top2	−6886	31	0.76	0.24
cov	−6900	14	–	–
top2	−6999	99	0.83	0.17
gam	−7201	202	–	–
standard	−7583	382	–	–

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs

b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row

c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

Aside from that, it is difficult to make a comparison, because several internal nodes in Lee & Hasegawa's (2011) MCC tree are poorly resolved and the same is true of the current majority tree. The major groupings, however, are similar (from top to bottom): Hokkaido-Yamanashi (although Tokyo is not included in the majority tree), Gifu-Aichi, Aomori-Miyagi-Akita-Yamagata-Iwate-Fukushima-Ibaragi-Tochigi, Shiga-Nara-Kyoto-Osaka-Wakayama (although in Lee & Hasegawa's (2011) MCC tree, Shiga is not included in this group), Hyogo-Tokushima-Kagawa-Ehime-Kochi-Okayama (although in Lee & Hasegawa's (2011) MCC tree, this group includes Hiroshima and Shiga), and Fukuoka-Saga-Nagasaki-Kumamoto-Miyazaki-Oita-Kagoshima. The grouping of Chiba-Gunma-Saitama-Kanagawa as found in the majority tree is not found in Lee & Hasegawa's (2011) MCC tree, and neither is the Nigata-Toyama-Ishikawa-Fukui-Nagano grouping.

In the minority tree, only a few groupings are supported at a reasonably high level: Miyagi-Fukushima-Miyazaki (0.75), Saitama-Tokushima-Nagano (0.43), Toyama-Okayama (0.60), Aomori-Osaka-Shiga-Wakayama (0.38), and Irafu-Ishigaki (0.49). None of these groupings are found in either Lee & Hasegawa's (2011) MCC tree or in the majority tree.

As was done for Indo-European, additional analyses using the "sitelh" command in *BayesPhylogenies* were conducted to calculate the log-likelihood that each cognate set evolved on the majority and the minority tree and thus to assess the signal retrieved by the minority tree. In Figure 12, an overview is presented of

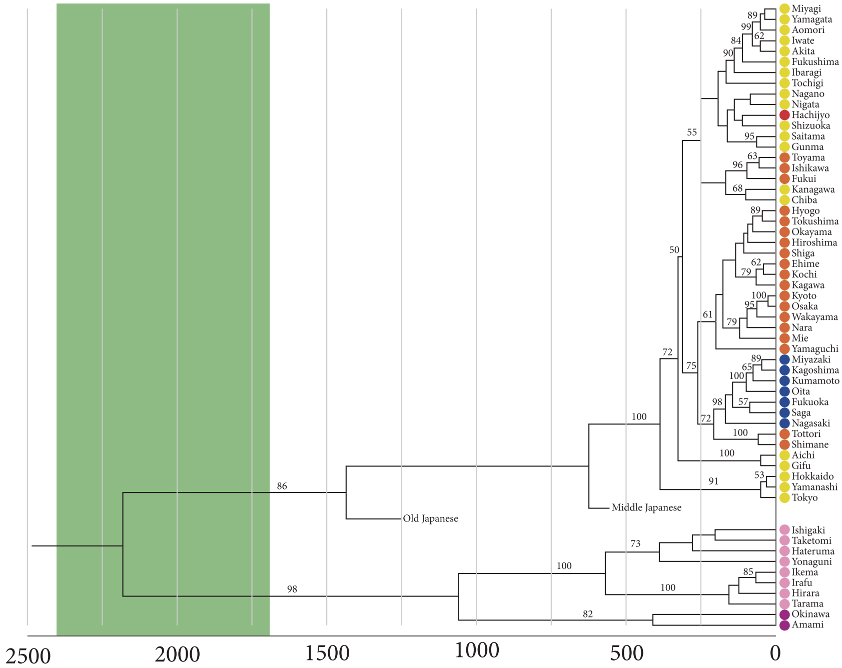


Figure 9. Lee & Hasegawa's (2011) dated maximum clade credibility tree of the Japonic languages. From Lee, Sean, & Toshikazu Hasegawa. 2011. Bayesian Phylogenetic Analysis Supports an Agricultural Origin of Japonic Languages. *Proceedings of the Royal Society B* 278:1725.3662–3669. By permission of the Royal Society.

the 83 cognate sets that support the minority tree for the cov + gam + top2 model. For most of these 83 cognate sets, support is rather low: a line has been drawn at 2, a point which is rather arbitrary but which helps the reader see which cognate sets really stand out. The meanings for cognate sets that support the minority tree with a log-likelihood difference of 2 or more are given on top of the bars.

Three pairs of cognate sets for three concepts stand out: cognate sets for 'tongue', 'to open, unlock', and 'night'. These sets support the minority tree with a large log-likelihood difference – over 15 log-likelihood units. It could be the case that these three pairs of cognate sets reflect a subgrouping of languages that is well-supported by the minority tree topology. In these three cases, there are only two cognate sets for each concept; therefore, both cognate sets make the same language grouping twice, which is why both cognate sets for each of these three concepts pattern similarly. Figure 12 will be discussed further below.

The trees in Figures 10 and 11, which give the majority and minority MCC trees for the cov + gam + top2 model and analyze which cognate sets support the minority tree in Figure 12, can be compared to the results from the cov

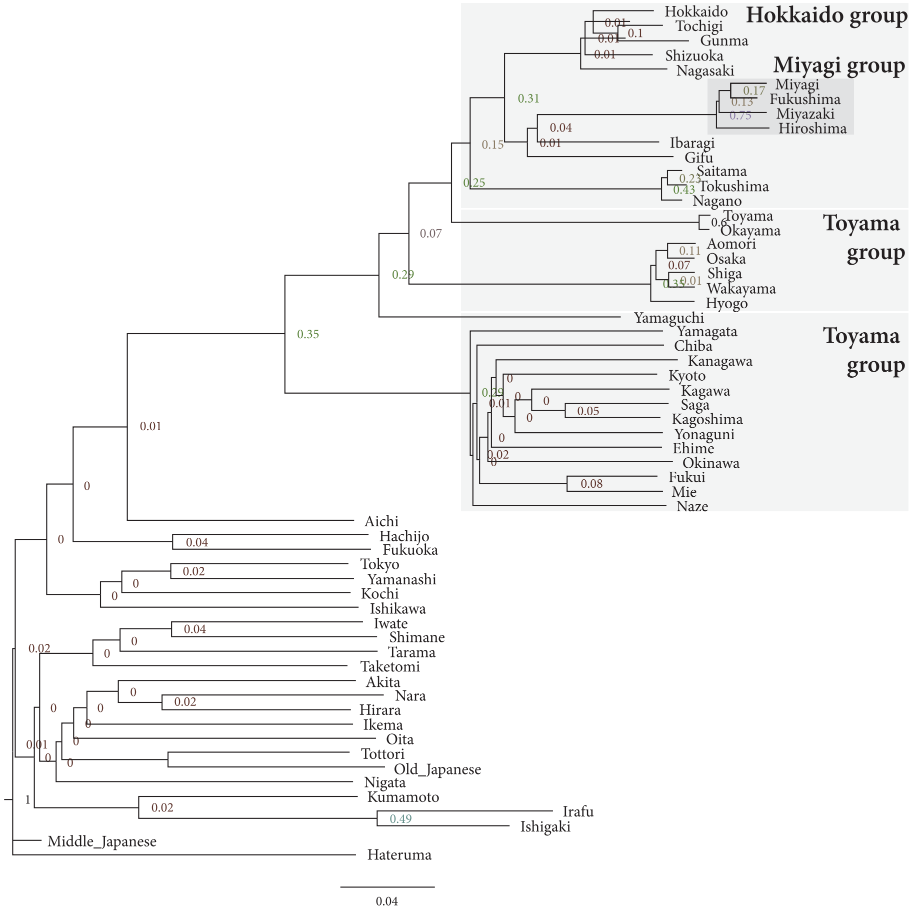


Figure 10. The minority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011). Gray boxes represent subgroups with relatively high posterior support that are discussed in text

+ top2 model, which has far greater support for the minority topology. The trees are included in Figures 13 and 14. The majority topology of this model looks very similar to the majority tree of the cov + gam + top2 model, except for the placement of Nigata and Nagasaki. The support values are similar as well. When comparing the minority tree topologies (Figures 10 and 13), these are likewise similar, but the reasonably supported groups – for instance, the Miyagi group – have higher posterior support values (0.75 in the cov + gam + top2 model for Miyagi-Fukushima-Miyazaki; 0.90 in the cov + top2 model for Miyagi-Hiroshima-Miyazaki-Fukushima). Thus, the minority topology captures the alternative signal better without gamma rate heterogeneity than it does with gamma rate heterogeneity.

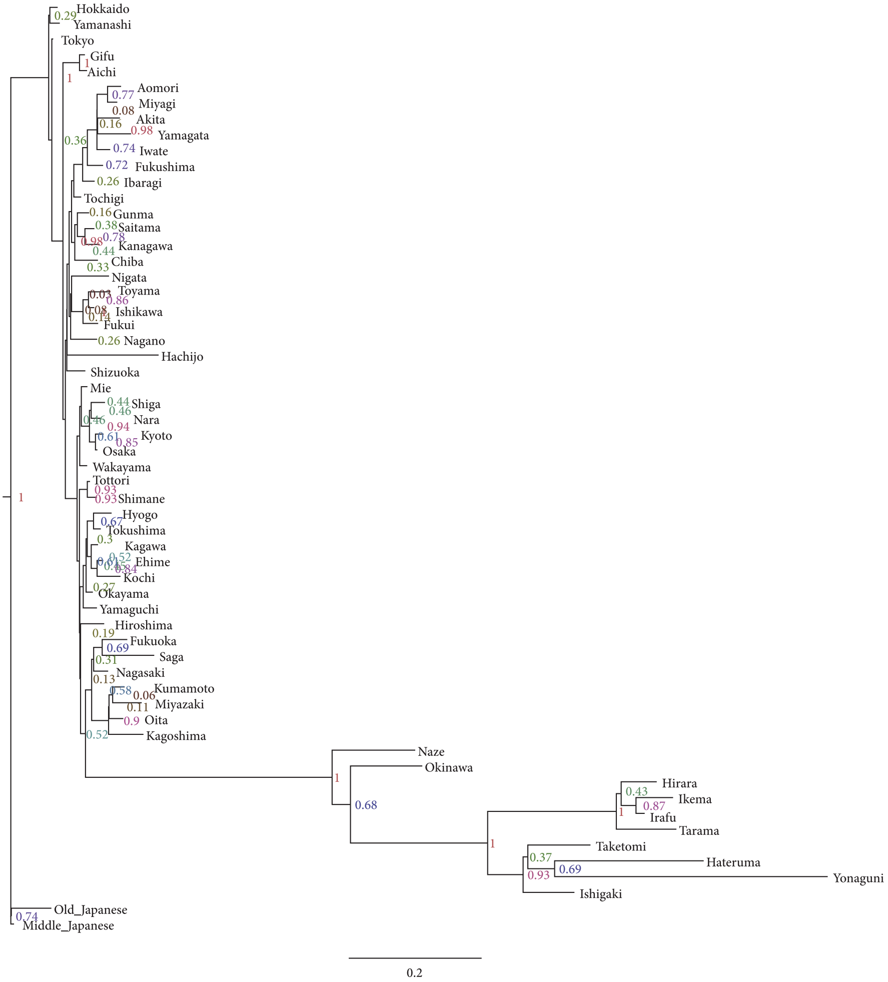


Figure 11. The majority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

The increased support for the minority tree is reflected by Figure 15, which plots the 192 cognate sets that support the minority tree of the cov + top2 model. This is more than double the number of cognate sets associated with the minority tree in the cov + gam + top2 model. Many of the same cognate sets are encountered here: 'to live, be alive', 'painful, sick', 'to shoot', 'thunder', 'to split', 'cold', 'tongue', 'ten', 'to vomit', 'to open, uncover', 'stick, wood', 'dust', 'night'.

It is easy to see how the three highest scoring cognate sets ('tongue', 'to open, uncover', 'night') inform the structure of the minority topology. In Figure 10, which represents the minority maximum clade credibility tree for the cov + gam + top2 model, the top right-most clade designated as "Hokkaido group" (with posterior probability 0.15) captures the division between the languages made by

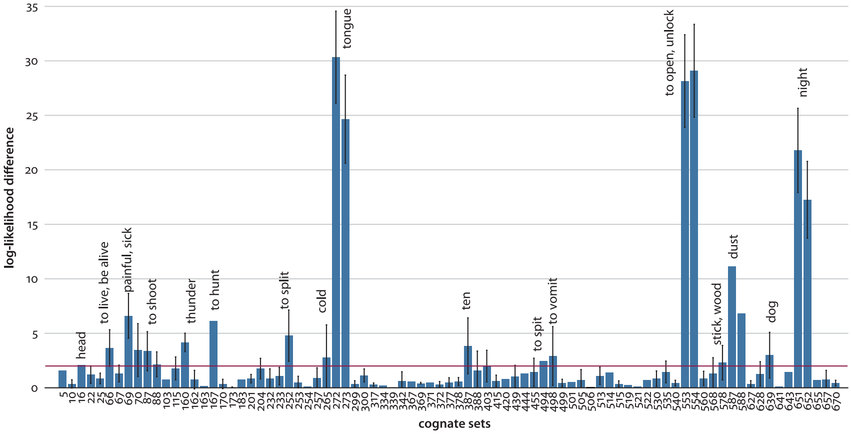


Figure 12. 83 cognate sets (out of 675) that support the minority tree for the cov + gam + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

the cognate sets for ‘tongue’ (the labels “Hokkaido group” and others in Figure 10 and 13 are arbitrary and just for convenience). The ‘to open, uncover’ cognate sets group together Miyagi, Fukushima, Miyazaki, and Hiroshima, a small clade with 0.75 posterior indicated by “Miyagi group,” and a larger group designated as “Toyama group.” Lastly, the ‘night’ cognate sets group together Saitama, Tokushima, Nagano, the upper part of the Toyama group, and Yamaguchi. In Figure 13, which represents the minority maximum clade credibility tree for the cov + top2 model, we find in part the same groupings. There is also a “Hokkaido group,” although Saitama, Tokushima, and Nagano are missing from it; a “Miyagi group,”; and what I called in Figure 13 the “Aomori group,” which again groups together languages on the basis of the ‘night’ cognate sets. Undoubtedly, the other cognate sets with reasonable support for the minority topologies in Figure 10 and 13 have similar (but partial) groupings, leading to the structure observed in the minority tree.

4.4 Sinitic

The Sinitic dataset used in this study comes from Wang (2004), a set which excludes Old Chinese. This set was obtained from the Chinese Dialect Database (List 2017) (see above). Table 4 gives an overview of the results of the different models of evolution for Ben Hamed & Wang’s (2006) lexical dataset for 23 Sinitic languages. The best-supported model is the cov + top2 model. However, the marginal log-likelihood difference with the second-best performing model, cov + gam + top2, is not very large (10 log-likelihood units). All models that estimate a second topology find a large amount of support for it. Figure 16 features the major-

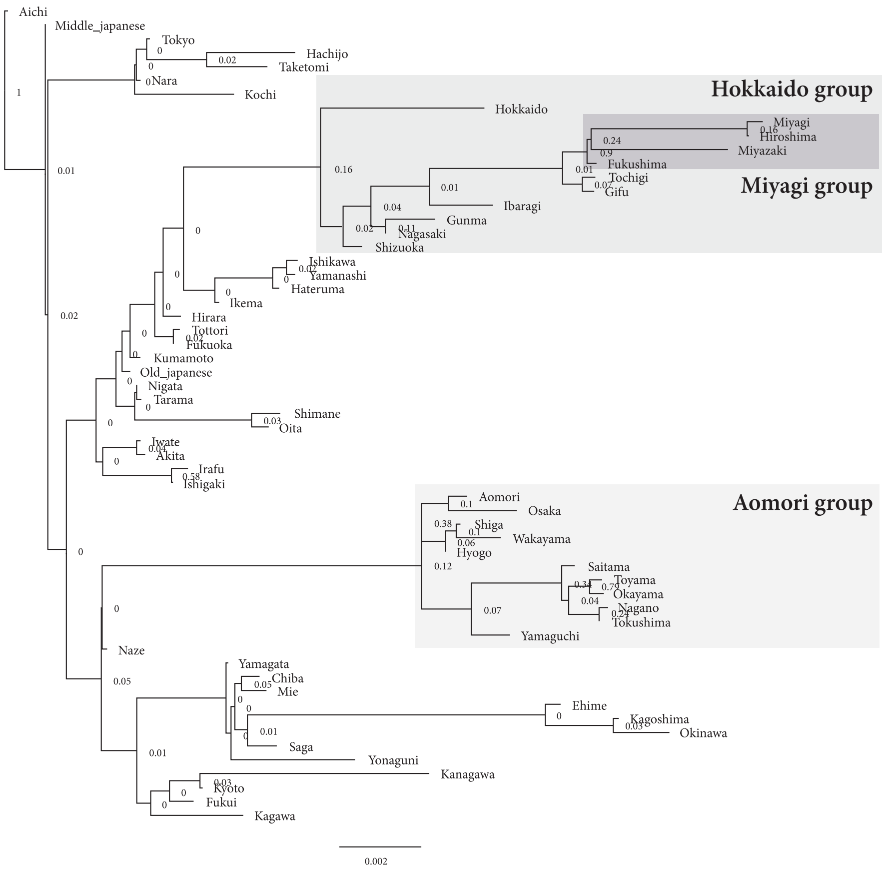


Figure 13. The minority maximum clade credibility tree for the cov + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011). Gray boxes represent subgroups with relatively high posterior support that are discussed in text

ity tree and the minority tree of the best-performing cov + top2 model, while Figure 17 features the same for the second-best model, cov + gam + top2. Comparing the cov + top2 results in Figure 16 to the cov + gam + top2 results in Figure 17, it is striking to see that there are no large differences. The majority trees have the same topology, but the cov + gam + top2 majority tree is better supported (higher posterior values on the internal nodes). Strangely enough, the cov + gam + top2 minority tree is also much better supported than the cov + top2 minority tree. They have (almost) the same structure, but the cov + top2 minority tree has very low posterior values for the internal nodes. The difference between the two models is that the cov + gam + top2 model has an almost-equal division of weighting between the majority and minority tree (0.55 and 0.45, respectively), while the cov

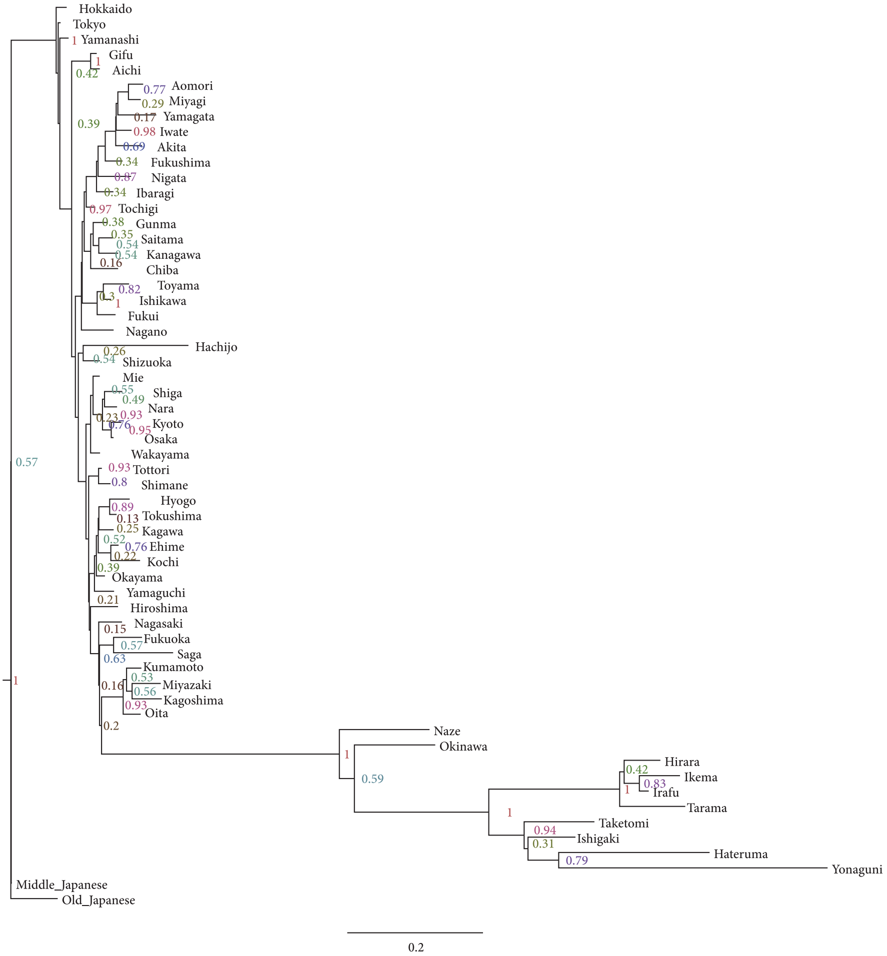


Figure 14. The majority maximum clade credibility tree for the cov + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

+ top2 model has more unequal weighting favoring the majority tree (majority tree 0.62, minority tree 0.38).

The majority and minority tree structures for both the cov + top2 model and the cov + gam + top2 model can be compared to network analyses in Ben Hamed & Wang (2006) and List (2015). Ben Hamed & Wang (2006) report the results of a Neighbor Net analysis, which was redone by the current author for the sake of better legibility and is presented in Figure 18. The Neighbor Net analysis retrieves the traditional groupings established for Sinitic languages (Norman 1988: 181ff):

- Mandarin: Wuhan, Yingshan, Chengdu, Ningxia, Beijing, Taiyuan, Yuci
- Wu: Ningbo, Shanghai, Shanghai B, Suzhou, Wenzhou

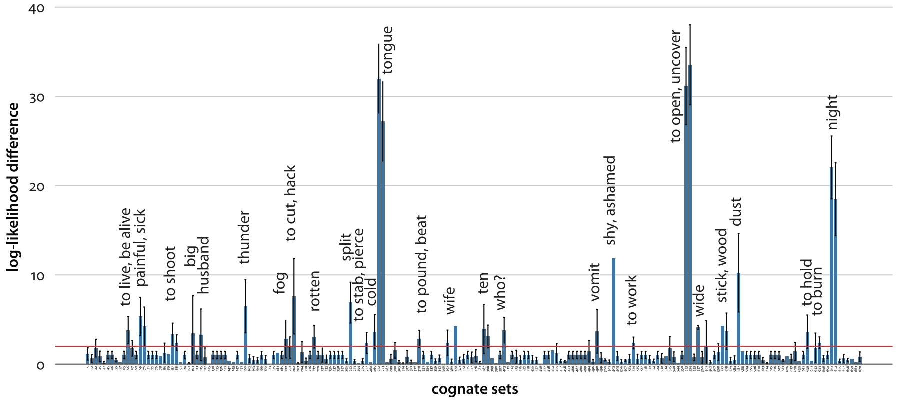


Figure 15. 192 cognate sets (out of 675) that support the minority tree for the cov + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

Table 4. Model comparison for Ben Hamed & Wang's (2006) lexical dataset on 23 Sinitic languages

Model	Marginal l-lh [a]	Marginal l-lh difference [b]	Weight majority tree [c]	Weight minority tree
cov + top2	−9007		0.62	0.38
cov + gam + top2	−9017	10	0.55	0.45
gam + top2	−9046	29	0.62	0.38
cov + gam	−9052	6	–	–
cov	−9060	8	–	–
gam	−9106	46	–	–
top2	−9138	32	0.58	0.42
standard	−9392	254	–	–

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs
b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row
c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

- Gan: Nanchang, Anyi
- Xiang: Changsha, Shuangfeng
- Hakka: Liancheng, Meixian
- Yue: Guangzhou
- Min: Fuzhou, Taibei (Taiwan), Zhangping, Xiamen

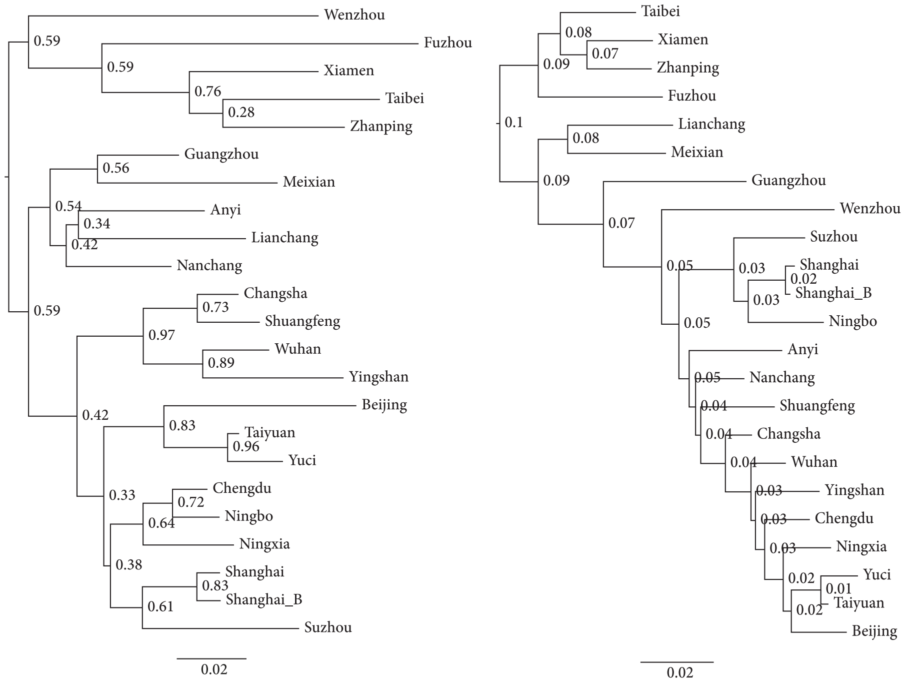


Figure 16. Maximum clade credibility trees for the cov + top2 model for Ben Hamed & Wang's (2006) Sinitic dataset. On the left is the majority topology (support 0.62); on the right is the minority topology (support 0.38)

Norman (1988: 181–183) distinguishes three higher-order subgroups: the Northern/Mandarin group, the Central group consisting of Wu, Gan, and Xiang languages, and the Southern group of Hakka, Yue, and Min languages. Ben Hamed & Wang's (2006) Neighbor Net reflects the Southern group, but we cannot recognize the Central group as such; rather, the main division is between Northern+Central and Southern.

Since Ben Hamed & Wang's (2006) Neighbor Net in Figure 18 captures traditional groupings as well as reticulate signals quite well, it is possible to use it as a benchmark for the behavior of the multiple topologies results. To start at the top of the majority topology in Figures 16 and 17, Wenzhou is included in the Min group of Fuzhou-Taibei-Zhangping-Xiamen, rather than with the Wu languages. Next, Guangzhou, the only Yue language, is sister to Meixian, one of the two Hakka languages. The other Hakka language, Liancheng, forms a group with Anyi and Nanchang, the two Gan languages. Further down, the two Xiang languages, Changsha and Shuangfeng, form a group with two of the Mandarin languages, Yingshan and Wuhan. This group splits up the Mandarin languages above and below it: first Beijing, Taiyuan, and Yuci; and lower down, a group of the remaining two Mandarin

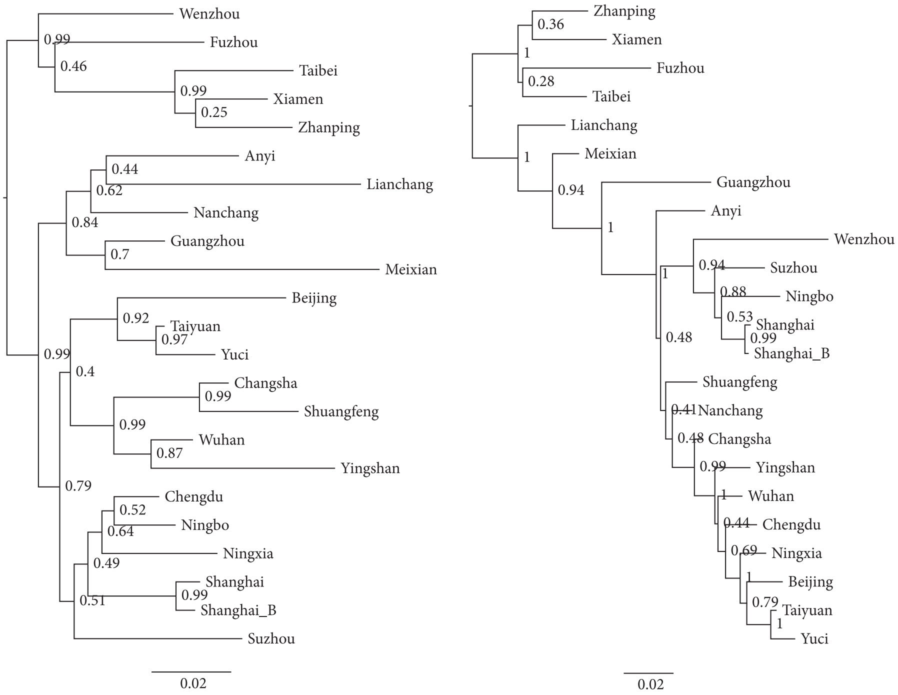


Figure 17. Two topologies for the cov + gam + top2 model for the Sinitic dataset (Ben Hamed & Wang 2006). On the left is the majority topology (support 0.55); on the right is the minority topology (support 0.45)

languages, Ningxia and Chengdu, with Ningbo (Wu). At the bottom, the remaining three Wu languages – Shanghai, Shanghai B, and Suzhou – form a group. All in all, the majority tree does not match Ben Hamed & Wang's (2006) Neighbor Net very closely at all: most clades contain languages from 2 different language groups.

Next, let us consider List (2015). List (2015) constructs minimal lateral networks on the basis of three different reference trees: (i) Laurent Sagart's *Arbre des Dialectes Chinois*, (ii) Jerry Norman's *Southern Chinese Hypothesis*, and (iii) Yóu Rǔjié's *Hànyǔ Fāngyán Shùxíngtú* 'Tree chart of Chinese dialects." These trees are abbreviated by List (2015) as *Arbre*, *Southern Chinese*, and *Shùxíngtú*, respectively, a naming system which will be followed here for the sake of convenience. List (2015: 36–37) additionally includes three reference trees that were reconstructed using distance-based methods and maximum parsimony, but since List (2015) emphasizes the need for the "correct" reference tree for minimal lateral networks to perform optimally, these additional trees will not be considered here.

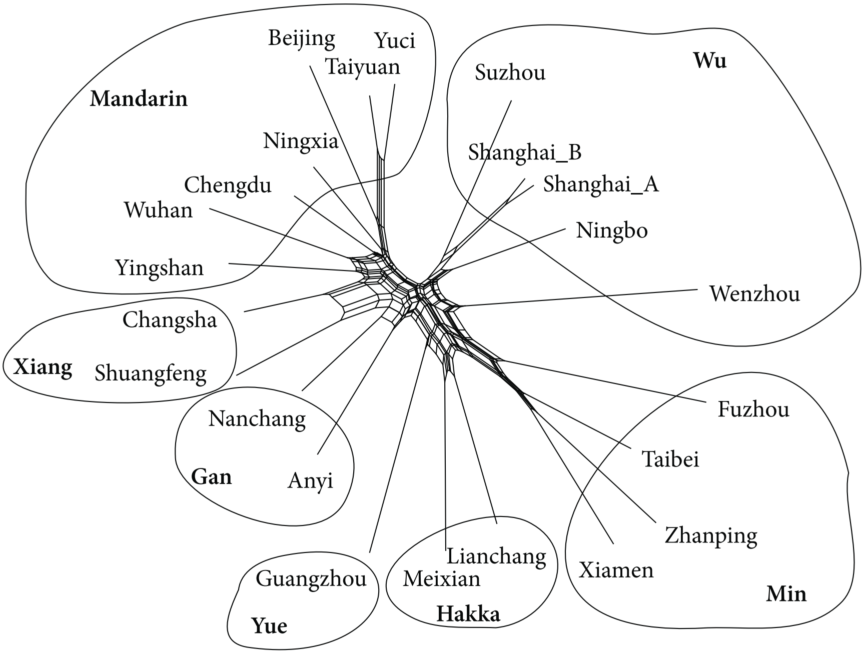


Figure 18. A Neighbor Net produced with Ben Hamed & Wang's (2006) dataset. It is identical to their Neighbor Net (their Figure 10, Ben Hamed & Wang 2006: 54), except for the inclusion of Old Chinese

The minimal lateral network based on the *Arbre* (reproduced in Figure 19) and *Southern Chinese* reference trees distinguishes the following clades (List 2015, Supplementary Information IV, going from right to left):

1. Zhangping, Xiamen, Taibei, Fuzhou (Norman's 1988 Min group)
2. Liancheng, Meixian, Guangzhou (Norman's 1988 Hakka + Yue groups)
3. Shanghai, Shanghai B, Suzhou, Ningbo, Wenzhou (Norman's 1988 Wu group)
4. Nanchang, Anyi (Norman's 1988 Gan group)
5. Changsha, Shuangfeng (Norman's 1988 Xiang group)
6. Yingshan, Wuhan, Ningxia, Chengdu, Beijing, Taiyuan, Yuci (Norman's 1988 Mandarin group)

The minimal lateral network based on the *Shùxíngtú* reference tree only differs from this network in that Guangzhou does not form a group with any other languages, while Meixian, Liancheng, Nanchang, and Anyi form a group. Given that List's (2015) minimal lateral networks have an almost identical structure to Ben Hamed & Wang's (2006) Neighbor Net, it is clear that the current majority tree deviates from these earlier results for the same dataset.

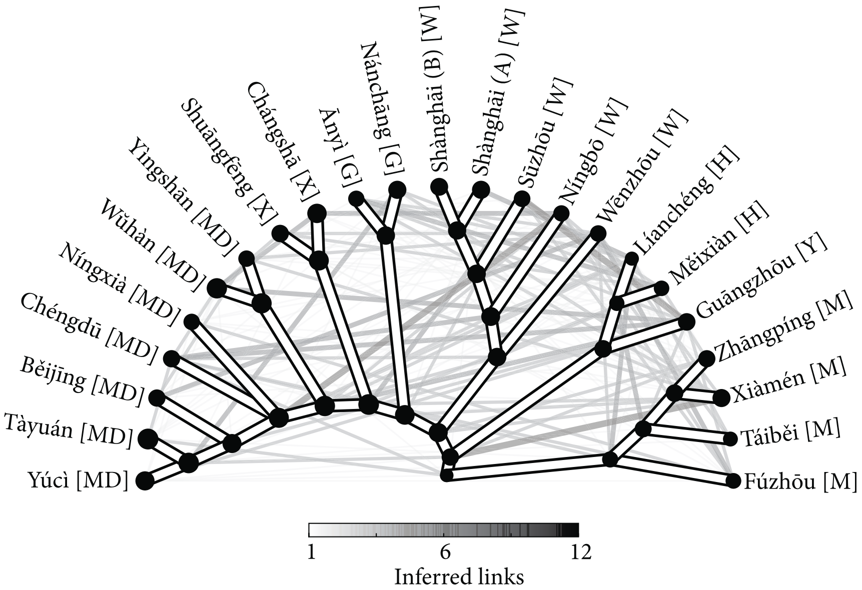


Figure 19. List's (2015) minimal lateral network using the *Arbre* reference tree. Reproduced with permission from the author. To allow for better legibility, the original figure has been changed so that the language labels are bigger and the legend has been moved

Given that the majority trees of the cov + top2 model and the cov + gam + top2 model are so different from earlier traditional and quantitative analyses, does this suggest that the majority trees pick up on affiliations between the languages that are non-tree-like? List (2015:40) presents the strongest lateral edges, connecting languages in different groupings. Let us compare the majority tree groupings to this table:

1. Wenzhou: Groups with the Min languages rather than the Wu languages. The reticulation drawing it towards the Min languages is clearly visible in Ben Hamed & Wang's (2006) Neighbor Net. List (2015: 40) also notes a heavy lateral edge between Wenzhou and the Min, Hakka, and Yue languages.
2. Guangzhou: Groups with Meixian, one of two Hakka languages, rather than being by itself. Again, reticulation between these two languages is visible in Ben Hamed & Wang's (2006: 54) Neighbor Net. Three of List's (2015: 40) strongest lateral edges involve Guangzhou: those involving Wuhan, Suzhou, and Meixian.
3. Liancheng: Groups with Anyi and Nanchang, the two Gan languages, rather than with its Hakka sister Meixian. This connection is not evident from Ben

Hamed & Wang (2006) or from List (2015), but may perhaps be an indirect consequence of the grouping of Guangzhou and Meixian.

4. Changsha and Shuangfeng (Xiang): Form a group with two of the Mandarin languages, Yingshan and Wuhan. These four languages are indeed close with considerable reticulation in Ben Hamed & Wang's (2006: 54) Neighbor Net. The attraction between Changsha and the Mandarin languages is noted by List (2015: 40) as well.
5. Ningbo: Groups with Ningxia and Chengdu, two of the Mandarin languages, rather than with the Wu languages. This reticulation is also present in Ben Hamed & Wang's (2006: 54) Neighbor Net, and List (2015: 40) notes a strong lateral edge between Ningbo and Mandarin, excluding Yingshan and Wuhan.

This discussion indicates that indeed, the groupings observed in the majority tree are caused by a signal in the data that is identified as reticulation in Ben Hamed & Wang's (2006: 54) Neighbor Net and as lateral edges in some of List's (2015) minimal lateral networks. What does the minority tree capture regarding the placement of these six languages?

1. Wenzhou: Groups with the Min languages rather than with the Wu languages in the majority tree. In the minority tree, it groups with the Wu languages Suzhou, Ningbo, Shanghai, and Shanghai B; this is the traditional genealogical grouping retrieved by Ben Hamed & Wang's (2006: 54) Neighbor Net and List's (2015) minimal lateral networks.
2. Guangzhou: Groups with Meixian, one of two Hakka languages in the majority tree. In the minority tree, it is by itself, but still close to Meixian.
3. Liancheng, the other Hakka language: Groups with Anyi and Nanchang, the two Gan languages in the majority tree. In the minority tree, it is much closer to Meixian, but the two are not sisters as they are in Ben Hamed & Wang's (2006) Neighbor Net.
4. Changsha and Shuangfeng (Xiang): Form a group with two of the Mandarin languages, Yingshan and Wuhan, in the majority tree. In the minority tree, Changsha is located just outside of the Mandarin group, followed by Nanchang and then by Shuangfeng. This indicates that Changsha has a closer affinity to the Mandarin languages than Shuangfeng does.
5. Ningbo: Groups with Ningxia and Chengdu, two of the Mandarin languages, in the majority tree. In the minority tree, it groups with the Wu languages Suzhou, Wenzhou, Shanghai, and Shanghai B; this is the traditional genealogical grouping retrieved by Ben Hamed & Wang's (2006: 54) Neighbor Net and List's (2015) minimal lateral networks.

The minority tree picks up on the signals that are not present in the majority tree, especially with regard to the structure of the Wu language group, i.e. the placement of Wenzhou and Ningbo. However, where Ben Hamed & Wang (2006) and List (2015) mark the Wu group as a valid vertical clade with Shanghai, Shanghai B, Suzhou, Ningbo, and Wenzhou, the multiple topologies method only recognizes this in the minority tree, not in the majority tree.

To further assess the contribution of different cognate sets to the weightings of the majority and minority trees, additional analyses using the "sitelh" command in *BayesPhylogenies* were conducted, as was done for Indo-European and Japonic. It is possible in this analysis to check this data against the meanings identified by List (2015: 40) that heavily contribute to the lateral edges in his minimal lateral networks. For the cov + gam + top2 analysis, it was found that almost all cognate sets offer support both for the majority tree and for the minority tree across all iterations of the MCMC chain: 1460 sites (out of 1511) offer support for the majority tree, and 1482 sites offer support for the minority topology (similar to what was found for Indo-European). There are some sites that are only associated with one of the two trees: 51 sites are associated only with the majority topology, and 29 sites are associated only with the minority topology. The language groupings made by these 51 and 29 cognate sets, unfortunately, do not present any discernible pattern. Comparing these cognate sets to cognate sets associated only with the majority and minority trees for the cov + top2 model, only one site is found to exclusively support the minority topology, while 25 sites support only the majority topology. These sites are a subset of the sites identified as majority-supporting only for the cov + gam + top2 model. They do not overlap with List's (2015: 40) overview of meanings that contribute heavily to the lateral edges found in his minimal lateral networks. It is unfortunate that a clear explanation does not arise; however, it is interesting that the two models find such a narrowly-defined set of cognate sets that support only the minority tree.

5. Discussion

In this article, the behavior of the multiple topologies method implemented in *BayesPhylogenies* has been described for four lexical datasets, Austronesian (Gray, Drummond & Greenhill 2009), Indo-European (Bouckaert et al. 2012), Japonic (Lee & Hasegawa 2011), and Sinitic (Ben Hamed & Wang 2006). These datasets have all displayed rather different behavior; in all cases, it is impossible to point to the exact reasons for the non-tree-like signal that was found, given the large number of processes that are responsible for change in lexical datasets. The results so far indicate that there has been no evidence whatsoever for a second topology for the Austronesian dataset.

There has been support for alternative histories in the Indo-European dataset. The two topologies found for the highest ranking cov + gam + top2 model look quite similar: in both the majority and minority MCC trees (Figures 7 and 8), the main subfamilies of Indo-European are correctly identified, and there are no languages with differing subgroup affiliation between the trees, as was observed for Japonic and Chinese (with the exception of Oscan, Umbrian, Luvian, and Lycian, which is due to large amounts of missing data). Hence no clear evidence for reticulation driving support for a minority topology was found. The main driver of the well-supported minority tree set seems to be a large group of cognate sets with larger rates of change that have reflexes in Germanic, Romance, and Slavic languages. Greater branch lengths both leading to and within these subfamilies are evident from the minority MCC tree, but this is especially clear from the Neighbor Net conducted on only those cognate sets that have a support for the minority topology of 5 log-likelihood units or larger (Figure A.4). As a result of this split, the branches of the Germanic, Romance, and Slavic subfamilies in the majority tree seem rather stunted, and so this tree does a worse job than the minority tree does in identifying subfamily groupings such as Northwest Germanic and South Slavic.

Indo-European is a unique language family because of its well-attested older and ancient languages, of which 22 (including Cornish) are included in the current analyses (see Chang et al. 2015: 205, 214–215 on special characteristics of ancient languages in phylogenetic analyses). Their placement within the minority and majority tree is not always correct, with the majority MCC tree especially showing a tendency to place non-contemporary languages close to each other. This trend is seen in the groupings of Avestan, Old Persian, and Vedic Sanskrit, as well as the grouping of Old High German and Gothic; similarly, Old Church Slavonic is pulled out of South Slavic. It may be the case that during cognate-coding, it is more straightforward to find cognates between ancient languages than it is to find cognates with modern reflexes due to the way etymological materials are organized. But this is not a problem in normal phylogenetic analyses, where older and ancient languages are constrained through calibration points. Chang et al. (2015: 215) conclude that the ancient languages Hittite, Vedic Sanskrit, Avestan, Ancient Greek, and Latin do not have more word forms per concept than modern languages do (their dataset is very similar to the current one). If anything, then, the emphasis on links between older or ancient languages does not result in them having significantly more synonyms for a given concept.

The Japonic dataset, likewise, was found to support more than one topology. It consists of data drawn from highly related languages, and as noted earlier, Lee

& Hasegawa's (2011) maximum clade credibility tree is not well-supported for several internal nodes. Nor does it retrieve the clear division between east and west Japonic languages established in traditional Japanese linguistics (Shibatani 1990: 196ff). The same was true for the current analysis, with the majority trees capturing some groupings not in Lee & Hasegawa's (2011) maximum clade credibility tree and the minority trees model capturing some language groupings that were not evident from the majority topology. Looking at the differences between Lee & Hasegawa's (2011) maximum clade credibility tree and the majority trees of the cov + gam + top2 and the cov + top2 models, it is helpful to consider the geographic location of the different languages given in Figure 20. Interesting differences between Lee & Hasegawa's (2011) analysis and the majority trees included the following:

1. The position of Shiga (number 25 in Figure 20), which is included in a clade with Nara (29), Kyoto (26), Osaka (28), and Wakayama (30) in the majority trees, similar to in Lee & Hasegawa's (2011: 6) Neighbor Net. Shiga is geographically very close to these languages. Shiga's position in Lee & Hasegawa (2011) is not well supported, but close to Hyogo (32), Tokushima (38), Okayama (33), and Hiroshima (35). It is unclear which of these groupings captures vertical or horizontal signal. In the minority trees, Shiga is positioned close to Osaka too.
2. The position of Hiroshima (number 35 in Figure 20), which is included in the Hyogo-Tokushima-Kagawa-Ehime-Kochi-Okayama group in Lee & Hasegawa's (2011) analysis but is placed just outside of that group in the majority trees. It seems that this latter position is caused by Hiroshima being attracted to the Southern Kyushu languages (in blue in Figure 20) – this is also clear from Lee & Hasegawa's (2011: 6) Neighbor Net.
3. The group Chiba (number 15 in Figure 20)-Gunma (11)-Saitama (12)-Kanagawa (14), which is found in the majority trees but not in Lee & Hasegawa's (2011) maximum clade credibility tree. These are all very close geographically (see Figure 20). In Lee & Hasegawa's (2011) analysis, Saitama (12) and Gunma (11) group with Nagano (16), Nigata (7), Hachijyo (21), and Shizuoka, but as in the grouping in the majority trees, this is not supported in more than 50% of the tree set. Saitama patterns with Tokushima (38) and Nagano (16) in the minority trees.
4. Nigata (number 7 in Figure 20)-Toyama (22)-Ishikawa (23)-Fukui (24)-Nagano (16), a grouping in the majority trees not found in Lee & Hasegawa (2011). It groups together the Western and Eastern languages that are spoken on the northern coast, so these languages are geographically close. In Lee & Hasegawa's (2011) consensus tree, Nigata (7) and Nagano (16) form

a group together with Hachijyo (21), Shizuoka (18), Saitama (12), and Gunma (11). In the majority tree, Hachijyo (21) and Shizuoka (18) are pulled outside of the Nigata-Toyama-Ishikawa-Fukui-Nagano group, and Saitama (12) and Gunma (11) group with Chiba (15) and Kanagawa (14). None of these groups are well-supported, either in the majority trees or in Lee & Hasegawa (2011), suggesting a large amount of uncertainty.

What can we speculate about the processes underlying the reticulation found in the Japonic dataset? These are all highly related language varieties that started to diverge only about 2,400 years ago, which makes a process like long branch attraction unlikely to have any effect. Japonic languages spread from Kyushu, the southernmost of Japan's large islands, and from there split up into what are called the mainland Japanese languages on Kyushu, Shikoku, Honshu, and Hokkaido; and the Ryukyuan languages on the smaller islands located to the south (Robbeets 2015: 27ff). The absence of a clear east-west split and instead the grouping of languages that are in close proximity, such as the Nigata-Toyama-Ishikawa-Fukui-Nagano group in all analyses, suggests that borrowing or a slow splitting up of dialect chains may be the cause (Lee & Hasegawa 2011: 6). Lee & Hasegawa (2011: 4) note that the number of isoglosses separating languages is small and there is not much overlap; additionally, they state that mainland Japanese languages remained in contact through the road network. Further evidence for contact within the Japanese islands and diversification across islands is provided by Lee & Hasegawa (2014).

This is, of course, speculation, but it seems that there is some evidence for incomplete lineage sorting as well. A few patterns in the data suggest processes other than dialect chain break up and/or (continued) borrowing. Incomplete lineage sorting occurs when there are ancestral polymorphisms that are by chance inherited only partially by different species or languages, possibly in a scattered manner. A possible case could be the well-supported connection between Toyama and Okayama in the minority trees, as well as the connection within the set of languages that share a reflex for one of the cognate sets for 'tongue', illustrated by the red circles in Figure 20. The cognate sets for 'tongue' serve to group together languages spoken far apart in non-adjacent geographical areas, making diffusion or borrowing unlikely. An alternative explanation could be influence from the political and religious capital between 700 and 1868, which was located first in Nara (number 29 in Figure 20) and then in Kyoto (26) – both in Central Japan – although neither of these languages share this particular reflex for 'tongue'. While both the Kyoto and Tokyo languages, as languages of the capital, served as lingua francas, and although the latter served as the basis for the formation of the standard language (Shibatani 1990: 185–187), neither seems involved in much reticu-

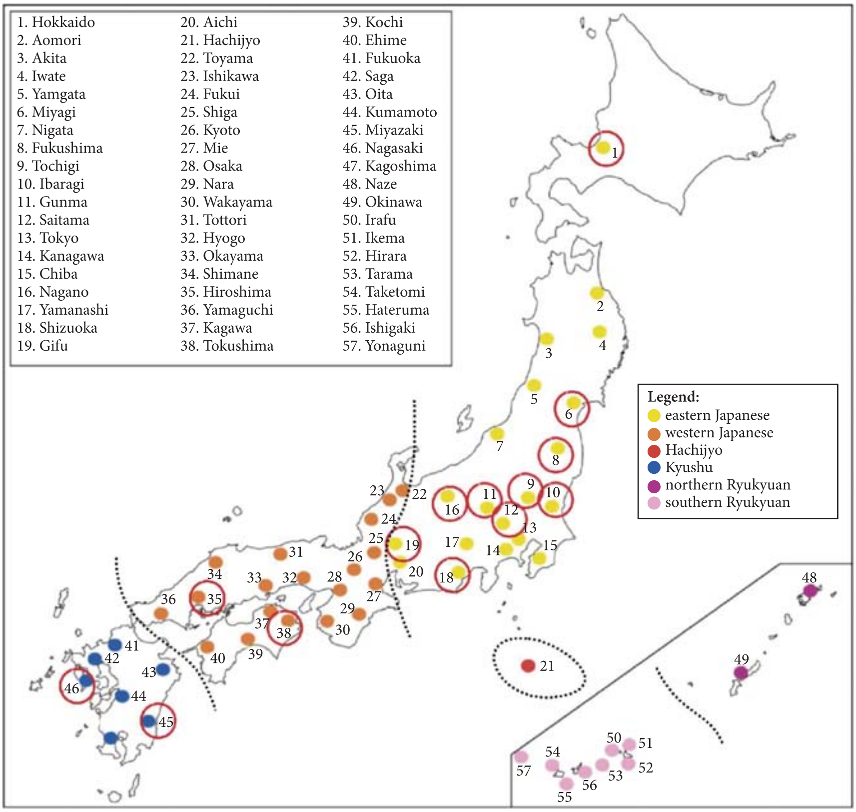


Figure 20. Map detailing the locations[3] of the contemporary languages taken from Lee & Hasegawa (2011) (their Figure s1). From Lee, Sean, & Toshikazu Hasegawa. 2011. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B*, 278:1725.3662–3669. By permission of the Royal Society.

lation for the current dataset. Lee & Hasegawa's (2011) Neighbor Net does not suggest this, and neither do the Delta scores and Q-Residuals calculated from the Neighbor Net analysis (as performed by the current author). Nevertheless, the influence of the old capitals as centers of innovation, pushing western characteristics outwards, should not be disregarded as a potential alternative explanation for these scattered patterns (this process is well attested in traditional Japanese linguistics; see Shibatani 1990: 200ff).

The Sinitic dataset shows the most dramatic division in majority and minority tree weightings, with an almost equal split between sites. Analyses of which sites

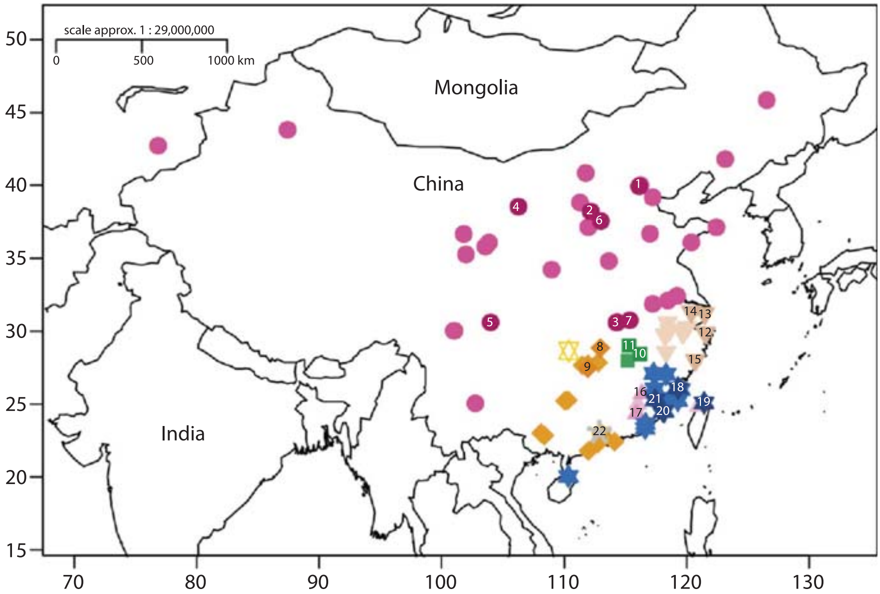
---

[3] The red circles were Legend also added by the current author and mark languages that have reflexes of one of two cognate sets for the meaning 'tongue'; see text.

support which tree revealed that across iterations, almost all sites could support both the majority and the minority topology. It was already clear from earlier studies (Ben Hamed & Wang 2006; List 2015) that a large amount of uncertainty and reticulation is present in this dataset. Ben Hamed & Wang (2006) find that parsimony trees might find appropriate subgroups but have poor bootstrap support, while Neighbor Nets capture both this reticulation and the major subdivisions identified by the trees. This is due to two opposing forces in Sinitic language history: one of differentiation, in which people spreading throughout China established new communities and languages; and one of homogenization, in which wide-spread multilingualism and borrowing created linkages between the languages (Ben Hamed & Wang 2006: 53). Norman (1988: 185) describes the first process as a "centrifugal force," and the second as a "centripetal counterbalance." The standard languages of the Chinese dynasties have influenced all Sinitic languages throughout most of their histories, but at the same time, they have had different impacts on different languages (Norman 1988: 185ff, Wang 1997: 57).

It is helpful to look at the geographic position of the languages, and so a map giving the position of Ben Hamed & Wang's (2006) sampled languages is presented in Figure 21. Several of the points raised earlier regarding the majority trees and how they deviate from Ben Hamed & Wang (2006) and List (2015) might be relevant to issues of borrowing between closely situated languages or to major migrations (see Zhou 1991 on migrations). Wenzhou (number 15 in Figure 21), which is affiliated in the majority trees with the Min languages rather than with the Wu languages proposed by Ben Hamed & Wang (2006) and List (2015), is closest to the Min languages geographically, as it borders with Fuzhou (18). Changsha (8) and Shuangfeng (9) (Xiang) form a group with two of the Mandarin languages, Yingshan (7) and Wuhan (3), in the majority trees. These four languages are indeed situated quite close together. Clearly, not all divergences away from the topologies found in Ben Hamed & Wang (2006) and List (2015) can be explained by geographical proximity. A case in point is Ningbo (12) (Wu), whose closest affiliation in the majority trees is with Ningxia (4) and Chengdu (5) (Mandarin): these are far apart on the longitudinal axis. Zhou (1991) describes in detail three major migrations that have contributed to language formation. Most of these migrations took place from north to south and/or from west to east. Zhou (1991: 41) mentions one particular migration towards the coast of what is now Jiangsu and Zhejiang from the north and specifically mentions northern migrants settling in Ningbo.

The analyses of the four language families show a two-way split: on the one hand, we have Austronesian and Indo-European, which have no evidence for multiple topologies or evidence that does not point towards language contact; on the other hand, we have Japonic and Sinitic, which have closely related language varieties and where continuous opposing forces of language diversification and



Legend:
Mandarin (or **Northern**) (●): (1) Beijing, (2) Taiyuan, (3) Wuhan, (4) Ningxia, (5) Chengdu, (6) Yuci, (7) Yingshan
Xiang (◆): (8) Changsha, (9) Shuangfeng
Gan (■): (10) Anyi, (11) Nanchang
Wu (or **Wu-Hui**) (▼): (12) Ningbo, (13) Shanghai, (14) Suzhou, (15) Wenzhou
Hakka (▲): (16) Liancheng, (17) Meixian
Min (★): (18) Fuzhou, (19) Taiwan, (20) Xiamen, (21) Zhangping
Yue (or **Cantonese**) (✳): (22) Guangzhou
Waxianghua (✡)

Figure 21. A map of the locations[4,5] of the Sinitic languages, taken from Glottolog (Hammarström et al. 2017)

homogenization through language contact and standard language use seem to have created reticulate patterns. The analyses show that it is worthwhile to consider using methods that aim to capture this reticulation even for datasets where known borrowings have been removed, because (i) we are very unlikely to remove all borrowings, especially more ancient ones (see Chang et al. 2015: 205), and (ii) borrowing is not the only process through which reticulate patterns may arise.

Ben Hamed & Wang (2006: 55) make an interesting and valid point regarding the impact of data coding. Given that their coding (and this applies to the coding of lexical data in almost all phylogenetic endeavors) does not distinguish between

---

[4] Light and dark hues of the same color indicate subfamily and numbered points in darker color indicate varieties present in Ben Hamed & Wang (2006).

[5] In order to increase legibility, the distances between the points representing Taiyuan and Yuci, Anyi and Nanchang, and Liancheng and Zhangping were increased slightly.

change in a suffix and a change in the root, does not contain transitional states such as variant roots, and carries no information about the likelihood of the direction of change, it could be the case that the tree-like signal is "hidden" by the "neutral" coding of lexical data. This point has also been made by List (2016), who studies the effects of coding partial cognacy in different ways and pushes for the adoption of multistate models over the binary gain-lose models most commonly used now (see also List 2015: 35 for an example of this problem in Ben Hamed & Wang's 2006 dataset). This is a more general problem with the application of phylogenetic methods in linguistics. Furthermore, this fact would not have an equal effect on all four datasets considered in this study, due to the different morphological make-ups of languages in different language families.

Another issue regarding data collection and coding can be detected by dividing the number of cognate sets by the number of meanings and subsequently by the number of languages in the sample. For Austronesian, Sinitic, and Indo-European, the resulting number of cognate sets per language and meaning is similar: 0.41, 0.31, and 0.27, respectively. For Japonic, this number is radically lower: 0.05 cognate sets per language per meaning. This may be one of the reasons for Lee & Hasegawa's (2011) poorly-supported clades: it might be the case that not enough variation is present in the dataset. In Indo-European, the inclusion of ancient languages may lead to a proliferation of cognate sets, as it may be unclear which of several attested near-synonyms is the "Swadesh term" (this term refers to the single word, mapping onto a given concept, that is most common, non-specialized and non-bound; see Swadesh 1952: 457). The Bouckaert et al. (2012) dataset may suffer from cognate over-population: although this has not been quantified, the dataset may include non-Swadesh terms for some concepts. Looking at Dutch, it includes *waar*, *recht*, *rechtmatig* and *echt* for 'true', and *breed*, *wijd*, and *ruim* for 'wide'. In both cases, only one term is the true Swadesh term, respectively *waar* 'true' and *breed* 'wide'. Including three or four synonym words for a concept, each with their own cognate history, may potentially introduce a source of reticulate signal as well as inflate the time depth of certain groups (see Chang et al. 2015: 211 on "overloaded" concept slots). Note that this is just speculation: it is unclear whether this played a significant role in the Bouckaert et al. (2012) analysis; note also that the number of cognate sets per language and meaning (0.27) is lower than that of the Austronesian (0.41) and Sinitic (0.31) datasets. Future advances in our knowledge of the impact of coding decisions on synonyms, in the partial cognate coding proposed by List (2016), and in other models of language evolution, including phoneme level analysis (Hruschka et al. 2015), may possibly help efforts to find phylogenetic signal in the future.

The current article has shown that the multiple topologies method has recovered, at least in part, the reticulation in lexical datasets that we knew about on the basis of distance-based phylogenetic inference, mostly Neighbor Net analyses, as reported in the four original articles and earlier follow-up studies. It also recovered more than one signal in the Indo-European analysis, which was speculated to be mostly due to differences in rate of lexical change. The Japonic and Sinitic results are not surprising: (i) the reticulate nature of the Sinitic dataset has been discussed previously, and (ii) the original analysis of the Japonic dataset results in a consensus tree that is not always well supported, suggesting the presence of reticulation. After describing the results, I speculated on the potential sources of reticulation, describing differential break-up of dialect chains and borrowing, both shortly after language/dialect break-up and long after in Japonic and Sinitic. There may be evidence for incomplete lineage sorting in Japonic too. The Indo-European results show that even when borrowing is not available as an explanation, the multiple topologies method can pick up on other processes that generate multiple signals. However, as the current study is not a strict test of the multiple topologies method, it is not possible to generalize in any way the current results and provide guidelines on how and when to use the multiple topologies method, nor on the interpretation of the results if multiple topologies are found.

How does the multiple topologies method compare with the method proposed by Nakhleh, Ringe & Warnow (2005) and the minimal lateral network method (for instance, List 2015)? The big difference is that the latter informs the user on what is vertical and what is horizontal evolution (as reference phylogenies are used), while the multiple topologies method cannot do this. This makes the multiple topologies method less prone to biases introduced by the reference phylogeny, but some users might prefer to "know" which aspects of the two evolutionary histories that are recovered constitute genealogical descent, and which can be attributed to borrowing and other horizontal processes. For the Sinitic dataset, the minority trees were found to pick up the genealogical grouping of the Wu languages, while the majority trees group Wenzhou and Ningbo together with other languages from other subgroups. The majority trees for Sinitic thus seemingly capture an amalgamate of vertical and horizontal signal, while only the minority trees capture the true vertical history of the Wu group. Of course, this is not a surprising result either. After all, the multiple topologies method does not know anything about the different processes of language evolution that are distinguished here; it only knows that some sites support one tree topology and other sites support another. In this respect, it is similar to other tools that are used to assess non-tree-like signal, such as Neighbor Nets and minimal lateral networks. However, the multiple topologies method outperforms these in two ways. First of all, it does not require a reference topology, thus taking one source of possible bias or error away from the analysis. Second, it is a character-based rather than a distance-based method, thus providing the user with the different paths along which different sites have changed, rather than summarizing average distances between language pairs, as Neighbor Net does.

6. Conclusion

In this article, I have described the first results of the use of the multiple topologies method (Pagel & Meade 2006a) implemented in *BayesPhylogenies* (Pagel & Meade 2004) with linguistic datasets. Four different language families were involved: Austronesian (data from Gray, Drummond & Greenhill 2009), Sinitic (Ben Hamed & Wang 2006), Indo-European (Bouckaert et al. 2012), and Japonic (Lee & Hasegawa 2011). Evidence for non-tree-like signal was found in all except Austronesian. Speculations on the origins of the non-tree-like signal were made in terms of a number of processes, including dialect chain break-up, borrowing (both shortly after language splits and later on), incomplete lineage sorting, and characteristics of lexical datasets. The Indo-European results suggest that rate of change differences between cognate sets can also be picked up as differentiating signals. The multiple topologies method is a useful tool for studying the dynamics of language evolution and may be used to help identify non-tree-like signal. However, as this article constitutes only an exploratory experiment, more formal tests of its behavior and applicability to linguistic datasets are needed.

Acknowledgements

First and foremost, I would like to thank Mark Pagel and Andrew Meade for their assistance on this project, especially Andrew, who helped to do analysis far beyond the call of duty. I would also like to thank Michael Dunn for preparing a new version of his Indo-European dataset to use in this paper, Sean Lee for making his Japonic dataset available, and Mattis List for help with the Sinitic dataset. Comments from three anonymous reviewers helped to make this a better paper, for which thanks are given. Thanks to Marcel Sommer for technical support. I thank the audience of ICHL22, the organizers of the workshop "Non-cladistic approaches to language genealogy," Siva Kalyan and Alexandre François, and additional guest editor Harald Hammarström for their commitment. This research was supported by ERC grant 'The evolution of human languages' (reference number 268744).

References

Algeo, John & Thomas Pyles. 2005. *The Origins and Development of the English Language (Sixth Edition)*. Boston: Wadsworth.

Atkinson, Quentin D. & Russell D. Gray. 2005. Curious Parallels and Curious Connections: Phylogenetic Thinking in Biology and Historical Linguistics. *Systematic Biology* 54:4.513–526. https://doi.org/10.1080/10635150590950317

Beekes, Robert S. P. 2011. *Comparative Indo-European Linguistics: An Introduction (Second Edition)*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.172

Ben Hamed, Mahé & Feng Wang. 2006. Stuck in the Forest: Trees, Networks, and Chinese Dialects. *Diachronica* 23:1.29–60. https://doi.org/10.1075/dia.23.1.04ham

Blench, Roger. 2015. "New Mathematical Methods" in Linguistics Constitute the Greatest Intellectual Fraud in the Discipline Since Chomsky. Talk presented at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Bouchard-Côté, Alexandre, David Hall, Thomas L. Griffiths & Dan Klein. 2013. Automated Reconstruction of Ancient Languages Using Probabilistic Models of Sound Change. *Proceedings of the National Academy of Sciences* 110:11.4224–4229. https://doi.org/10.1073/pnas.1204678110

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337:6097.957–960. https://doi.org/10.1126/science.1219669

Bowern, Claire & Quentin D. Atkinson. 2012. Computational Phylogenetics and the Internal Structure of Pama-Nyungan. *Language* 88:4.817–845. https://doi.org/10.1353/lan.2012.0081

Bowern, Claire, Patience Epps, Russell D. Gray, Jane Hill, Keith Hunley, Patrick McConvell & Jason Zentz. 2011. Does Lateral Transmission Obscure Inheritance in Hunter-Gatherer Languages? *PLOS ONE* 6:9.e25195. https://doi.org/10.1371/journal.pone.0025195

Bowern, Claire & Bethwyn Evans, eds. 2014. *The Routledge Handbook of Historical Linguistics*. London: Routledge.

Bryant, David & Vincent Moulton. 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution* 21:2.255–265. https://doi.org/10.1093/molbev/msh018

Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis. *Language* 91:1.194–244. https://doi.org/10.1353/lan.2015.0005

Chousou-Polydouri, Natalia, Joshua Birchall, Sérgio Meira, Zachary O'Hagan & Lev Michael. 2016. A Test of Coding Procedures for Lexical Data with Tupí-Guaraní and Chapacuran Languages. *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics* ed. by Christian Bentz, Gerhard Jäger & Igor Yanovich. Tübingen: University of Tübingen. Available at https://publikationen.uni-tuebingen.de/xmlui/handle/10900/68558

Collard, Mark, Stephen J. Shennan & Jamshid J. Tehrani. 2006. Branching, Blending, and the Evolution of Cultural Similarities and Differences Among Human Populations. *Evolution and Human Behavior* 27:3.169–184. https://doi.org/10.1016/j.evolhumbehav.2005.07.003

Croft, William. 2008. Evolutionary Linguistics. *Annual Review of Anthropology* 37.219–234. https://doi.org/10.1146/annurev.anthro.37.081407.085156

Drummond, Alexei J. & Remco R. Bouckaert. 2015. *Bayesian Evolutionary Analysis with BEAST*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139095112

Dunn, Michael. 2014. Language Phylogenies. In Claire Bowern & Bethwyn Evans, eds., 190–211. London: Routledge.

Dunn, Michael. 2014. *Indo-European Lexical Cognate Data*. Zenodo. Available at https://doi.org/10.5281/zenodo.1284236

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals. *Nature* 473.79–82. https://doi.org/10.1038/nature09923

Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indo-European Classification: A Lexicostatistical Experiment. *Transactions of the American Philosophical Society* 82.1–132. https://doi.org/10.2307/1006517

Edwards, Anthony W. F. & Luigi Luca Cavalli-Sforza. 1964. Reconstruction of Evolutionary Trees. *Phenetic and Phylogenetic Classification* ed. by V. H. Heywood & J. McNeill, 67–76. London: Systematics Association.

Eska, Joseph F. & Don Ringe. 2004. Recent Work in Computational Linguistic Phylogeny. *Language* 80:3.569–582. https://doi.org/10.1353/lan.2004.0123

Forster, Peter & Colin Renfrew, eds. 2006. *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute Monographs.

François, Alexandre. 2014. Trees, Waves, and Linkages: Models of Language Diversification. In Claire Bowern & Bethwyn Evans, eds., 161–189.

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds.), *Let's talk about trees: Genetic Relationships of Languages and Their Phylogenic Representation* (Senri Ethnological Studies, 98). Ōsaka: National Museum of Ethnology. 59–89.

Galucio, Ana Vilacy, Sérgio Meira, Joshua Birchall, Denny Moore, Nilson Gabas Júnior, Sebastian Drude, Luciana Storto, Gessiane Picançio & Carmen Reis Rodrigues. 2015. Genealogical Relations and Lexical Distances Within the Tupian Linguistic Family. *Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas* 10:2.229–274. https://doi.org/10.1590/1981-81222015000200004

Geisler, Hans & Johann-Mattis List. 2013. Do Languages Grow on Trees? The Tree Metaphor in the History of Linguistics. *Classification and Evolution in Biology, Linguistics and the History of Science: Concepts – Methods – Visualization* ed. by Heiner Fangerau, Hans Geisler, Thorsten Halling & William Martin, 111–124. Stuttgart: Franz Steiner Verlag.

Gray, Russell D., David Bryant & Simon J. Greenhill. 2010. On the Shape and Fabric of Human History. *Philosophical Transactions of the Royal Society B* 365:1559.3923–3933. https://doi.org/10.1098/rstb.2010.0162

Gray, Russell D., Simon J. Greenhill & Robert M. Ross. 2007. The Pleasures and Perils of Darwinizing Culture (With Phylogenies). *Biological Theory* 2:4.360–375. https://doi.org/10.1162/biot.2007.2.4.360

Gray, Russell D. & Fiona M. Jordan. 2000. Language Trees Support the Express-Train Sequence of Austronesian Expansion. *Nature* 405.1052–1055. https://doi.org/10.1038/35016575

Gray, Russell D., Alexei J. Drummond & Simon J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* 323:5913.479–483. https://doi.org/10.1126/science.1166858

Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. 2009. Does Horizontal Transmission Invalidate Cultural Phylogenies? *Proceedings of the Royal Society B* 276.2299–2306. https://doi.org/10.1098/rspb.2008.1944

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu Expansion Shows That Habitat Alters the Route and Pace of Human Dispersals. *Proceedings of the National Academy of Sciences* 112:43.13296–13301. https://doi.org/10.1073/pnas.1503793112

Hammarström, Harald, Sebastian Bank, Robert Forkel & Martin Haspelmath. 2017. *Glottolog 3.1*. Jena: Max Planck Institute for the Science of Human History.

Heggarty, Paul. 2006. Interdisciplinary Indiscipline? Can Phylogenetic Methods Meaningfully be Applied to Language Data – And to Dating Language? In Peter Forster & Colin Renfrew, eds., 183–194.

Heggarty, Paul, Warren Maguire & April McMahon. 2010. Splits or Waves? Trees or Webs? How Divergence Measures and Network Analysis Can Unravel Language Histories. *Philosophical Transactions of the Royal Society B* 365:1559.3829–3843. https://doi.org/10.1098/rstb.2010.0099

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* 14:2–3.167–214. https://doi.org/10.1080/09296170701378916

Hruschka, Daniel J., Simon Branford, Eric D. Smith, Jon Wilkins, Andrew Meade, Mark Pagel & Tanmoy Bhattacharya. 2015. Detecting Regular Sound Changes in Linguistics as Events of Concerted Evolution. *Current Biology* 25:1.1–9. https://doi.org/10.1016/j.cub.2014.10.064

Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science* 294:5550.2310–2314. https://doi.org/10.1126/science.1065889

Huson, Daniel H. & David Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23:2.254–267. https://doi.org/10.1093/molbev/msj030

Huson, Daniel H. & Celine Scornavacca. 2010. A Survey of Combinatorial Methods for Phylogenetic Networks. *Genome Biology and Evolution* 3.23–35. https://doi.org/10.1093/gbe/evq077

Huson, Daniel H. & Celine Scornavacca. 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* 61:6.1061–1067. https://doi.org/10.1093/sysbio/sys062

Jacques, Guillaume & Johann-Mattis List. This issue. Save the Trees: Why We Need Tree Models in Linguistic Reconstruction. *Journal of Historical Linguistics* 9:1. 128–167.

Jaeger, Gerhard & Søren Wichmann. 2016. Inferring the World Tree of Languages From Word Lists. *The Evolution of Language: Proceedings of the 11th International Conference on the Evolution of Language (EvoLang XI)* ed. by Séan G. Roberts, Christine Cuskley, Luke McCrohon, Lluis Barceló-Coblijn, Olga Feher & Tessa Verhoef. Available at http://evolang.org/neworleans/papers/147.html

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the tree model: A framework for Historical Glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds), *Let's talk about trees: Genetic Relationships of Languages and Their Phylogenic Representation (Senri Ethnological Studies, 98)*. Ōsaka: National Museum of Ethnology. 59–89.

Kelly, Luke J. & Geoff K. Nicholls. 2017. Lateral Transfer in Stochastic Dollo Models. *The Annals of Applied Statistics* 11:2.1146–1168. https://doi.org/10.1214/17-AOAS1040

Koonin, Eugene V., Kira S. Makarova & L. Aravind. 2001. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annual Review of Microbiology* 55:1.709–742. https://doi.org/10.1146/annurev.micro.55.1.709

Lee, Sean. 2018. *The Best Fitting Tree Model for Japonic Languages*. Zenodo. Available at https://doi.org/10.5281/zenodo.1220896

Lee, Sean & Toshikazu Hasegawa. 2011. Bayesian Phylogenetic Analysis Supports an Agricultural Origin of Japonic Languages. *Proceedings of the Royal Society B*, 278.3662–3669. https://doi.org/10.1098/rspb.2011.0518

Lee, Sean & Toshikazu Hasegawa. 2014. Oceanic Barriers Promote Language Diversification in the Japanese Islands. *Journal of Evolutionary Biology* 27:9.1905–1912. https://doi.org/10.1111/jeb.12442

Levinson, Stephen C. & Russell D. Gray. 2012. Tools from Evolutionary Biology Shed New Light on the Diversification of Languages. *Trends in Cognitive Sciences* 16:3.167–173. https://doi.org/10.1016/j.tics.2012.01.007

List, Johann-Mattis. 2015. Network Perspectives on Chinese Dialect History: Chances and Challenges. *Bulletin of Chinese Linguistics* 8:1.27–47. https://doi.org/10.1163/2405478X-00801002

List, Johann-Mattis. 2016. Beyond Cognacy: Historical Relations Between Words and Their Implication for Phylogenetic Reconstruction. *Journal of Language Evolution* 1:2.119–136. https://doi.org/10.1093/jole/lzw006

List, Johann-Mattis. 2017. *Chinese Dialect Database*. Jena: Max Planck Institute for the Science of Human History. Available at https://github.com/digling/cddb

List, Johann-Mattis, Shijulal Nelson-Sathi, Hans Geisler & William Martin. 2013. Networks of Lexical Borrowing and Lateral Gene Transfer in Language and Genome Evolution. *Bioessays* 36:2.141–150. https://doi.org/10.1002/bies.201300096

List, Johann-Mattis, Shijulal Nelson-Sathi, William Martin & Hans Geisler. 2014. Using Phylogenetic Networks to Model Chinese Dialect History. *Language Dynamics and Change*, 4:2.222–252. https://doi.org/10.1163/22105832-00402008

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* 11.360–375. https://doi.org/10.1186/s13062-016-0145-2

Macklin-Cordes, Jayden, & Erich Round. 2015. High-Definition Phonotactics Reflect Linguistic Pasts. *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics* ed. by Johannes Wahle, Marisa Kollner, Harald Baayen, Gerhard Jäger & Tineke Baayen-Oudshoorn, 1–5. Tübingen: University of Tübingen.

Maddison, Wayne P. 1997. Gene Trees in Species Trees. *Systematic Biology* 46:3.523–536. https://doi.org/10.1093/sysbio/46.3.523

Matthews, Luke J., Jamie J. Tehrani, Fiona M. Jordan, Mark Collard & Charles L. Nunn. 2011. Testing for Divergent Transmission Histories Among Cultural Characters: A Study Using Bayesian Phylogenetic Methods and Iranian Tribal Textile Data. *PLOS ONE* 6.e14810. https://doi.org/10.1371/journal.pone.0014810

Meira, Sérgio, Joshua Birchall & Natalia Chousou-Polydouri. 2015. A Character-Based Internal Classification of the Cariban Language Family. Paper presented at the 48th Annual Meeting of the Societas Linguistica Europaea, Leiden, The Netherlands, September 2–5, 2015.

Morrison, David A. 2016. Genealogies: Pedigrees and Phylogenies Are Reticulating Networks Not Just Divergent Trees. *Evolutionary Biology* 43:4.456–473. https://doi.org/10.1007/s11692-016-9376-5

Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81:2.382–420. https://doi.org/10.1353/lan.2005.0078

Nelson-Sathi, Shijulal, Johann-Mattis List, Hans Geisler, Heiner Fangerau, Russell D. Gray, William Martin & Tal Dagan. 2011. Networks Uncover Hidden Lexical Borrowing in Indo-European Language Evolution. *Proceedings of the Royal Society B* 278:1713.1794–1803. https://doi.org/10.1098/rspb.2010.1917

Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.

Nunn, Charles L., Christian Arnold, Luke Matthews & Monique Borgerhoff Mulder. 2010. Simulating Trait Evolution for Cross-Cultural Comparison. *Philosophical Transactions of the Royal Society B* 365:1559.3807–3819. https://doi.org/10.1098/rstb.2010.0009

Nunn, Charles L., Monique Borgerhoff Mulder & Sasha Langley. 2006. Comparative Methods for Studying Cultural Trait Evolution: A Simulation Study. *Cross-Cultural Research* 40:2.177–209. https://doi.org/10.1177/1069397105283401

Pagel, Mark. 2009. Human Language as a Culturally Transmitted Replicator. *Nature Reviews Genetics* 10.405–415. https://doi.org/10.1038/nrg2560

Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of Word-Use Predicts Rates of Lexical Evolution Throughout Indo-European History. *Nature* 449.717–721. https://doi.org/10.1038/nature06176

Pagel, Mark & Andrew Meade. 2004. A Phylogenetic Mixture Model for Detecting Pattern-Heterogeneity in Gene Sequence or Character-State Data. *Systematic Biology* 53:4.571–581. https://doi.org/10.1080/10635150490468675

Pagel, Mark & Andrew Meade. 2006a. Detecting Conflicting Phylogenetic Signal: A Mixture Model Approach with Multiple Tree Topologies. Paper presented at the New Zealand Phylogenetics Meeting, Kaikoura, New Zealand, February 12–17, 2006. Available at https://zenodo.org/record/1117335#.WjeEzSOZPFR

Pagel, Mark & Andrew Meade. 2006b. Estimating Rates of Lexical Replacement on Phylogenetic Trees of Languages. In Peter Forster & Colin Renfrew, eds., 173–182.

Pagel, Mark & Andrew Meade. N.d. Draft Manual for *BayesPhylogenies*. Available at http://www.evolution.reading.ac.uk/BayesPhy.html

Pereltsvaig, Asya & Martin W. Lewis. 2015. *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107294332

Robbeets, Martine. 2015. *Diachrony of Verb Morphology: Japanese and the Transeurasian Languages*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110399943

Rokas, Antonis & Sean B. Carroll. 2006. Bushes in the Tree of Life. *PLOS Biology* 4:11.e352. https://doi.org/10.1371/journal.pbio.0040352

Ronquist, Fredrik, Paul van der Mark & John P. Huelsenbeck. 2009. Bayesian Phylogenetic Analysis Using MrBayes. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* ed. by Marco Salemi, Philippe Lemey & Anne-Mieke Vandamme, 210–266. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819049.009

Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.

Swadesh, Morris. 1952. Lexico-statistic Dating of Prehistoric Ethnic Contacts: With Special Reference for 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. https://doi.org/10.1086/464321

Tadmor, Uri. 2009. Loanwords in the World's Languages: Findings and Results. *Loanwords in the World's Languages: A Comparative Handbook* ed. by Martin Haspelmath & Uri Tadmor, 55–75. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110218442.55

Tehrani, Jamshid & Mark Collard. 2002. Investigating Cultural Evolution Through Biological Phylogenetic Analyses of Turkmen Textiles. *Journal of Anthropological Archaeology* 21:4.443–463. https://doi.org/10.1016/S0278-4165(02)00002-8

Verkerk, Annemarie. 2014. The Evolutionary Dynamics of Motion Event Encoding. Radboud University Nijmegen PhD dissertation.

Wang, Feng. 2004. BCD: Basic Words of Chinese Dialects. Available at https://github.com/digling/cddb/tree/master/datasets/Wang2004a

Wang, William S.-Y. 1997. Languages or Dialects? *The CUHK Journal of Humanities* 1.54–62.

Wang, William S.-Y. & James W. Minett. 2005. Vertical and Horizontal Transmission in Language Evolution. *Transactions of the Philological Society* 103:2.121–146. https://doi.org/10.1111/j.1467-968X.2005.00147.x

Wichmann, Søren, Robert S. Walker, Taraka Rama & Eric W. Holman. 2011. Correlates of Reticulation in Linguistic Phylogenies. *Language Dynamics and Change* 1:2.205–240. https://doi.org/10.1163/221058212X648072

Widmer, Manuel, Sandra Auderset, Johanna Nichols, Paul Widmer & Balthasar Bickel. 2017. NP Recursion Over Time: Evidence from Indo-European. *Language* 93:4.799–826. https://doi.org/10.1353/lan.2017.0058

Willems, Matthieu, Etienne Lord, Louise Laforest, Gilbert Labelle, François-Joseph Lapointe, Anna Maria Di Sciullo & Vladimir Makarenkov. 2016. Using Hybridization Networks to Retrace the Evolution of Indo-European Languages. *BMC Evolutionary Biology* 16:1.180. https://doi.org/10.1186/s12862-016-0745-6

Xie, Wangang, Paul O. Lewis, Yu Fan, Lynn Kuo & Ming-Hui Chen. 2010. Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection. *Systematic Biology* 60:2.150–160. https://doi.org/10.1093/sysbio/syq085

Zhou, Kevin & Claire Bowern. 2015. Quantifying Uncertainty in the Phylogenetics of Australian Numeral Systems. *Proceedings of the Royal Society B* 282:1815.20151278. https://doi.org/10.1098/rspb.2015.1278

Zhou, Zhenhe. 1991. Migrations in Chinese History and Their Legacy on Chinese Dialects. *Journal of Chinese Linguistics Monograph Series* 3.29–49.

Appendix

Table 5. Model comparison for Bouckaert et al.'s (2012) updated lexical dataset on 116 Indo-European languages, adding the additional analysis including site log-likelihoods (in bold)

Model	**Marginal l-lh** [a]	**Marginal l-lh difference** [b]	**Weight majority tree** [c]	**Weight minority tree**
cov + gam + top2	−51167		0.60	0.40
cov + gam + top2 sitelh	**−51179**	**12**	**0.55**	**0.45**
cov + top2	−51182	3	0.65	0.35
gam + top2	−51375	193	0.77	0.23
cov + gam	−51450	75	–	–
cov	−51524	74	–	–
gam	−51791	267	–	–
top2	−51832	41	0.64	0.36
standard	−52812	980	–	–

a. marginal l-lh = highest marginal log-likelihood out of 5 runs

b. marginal l-lh difference = difference in log-likelihood units with better performing model in above row

c. weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

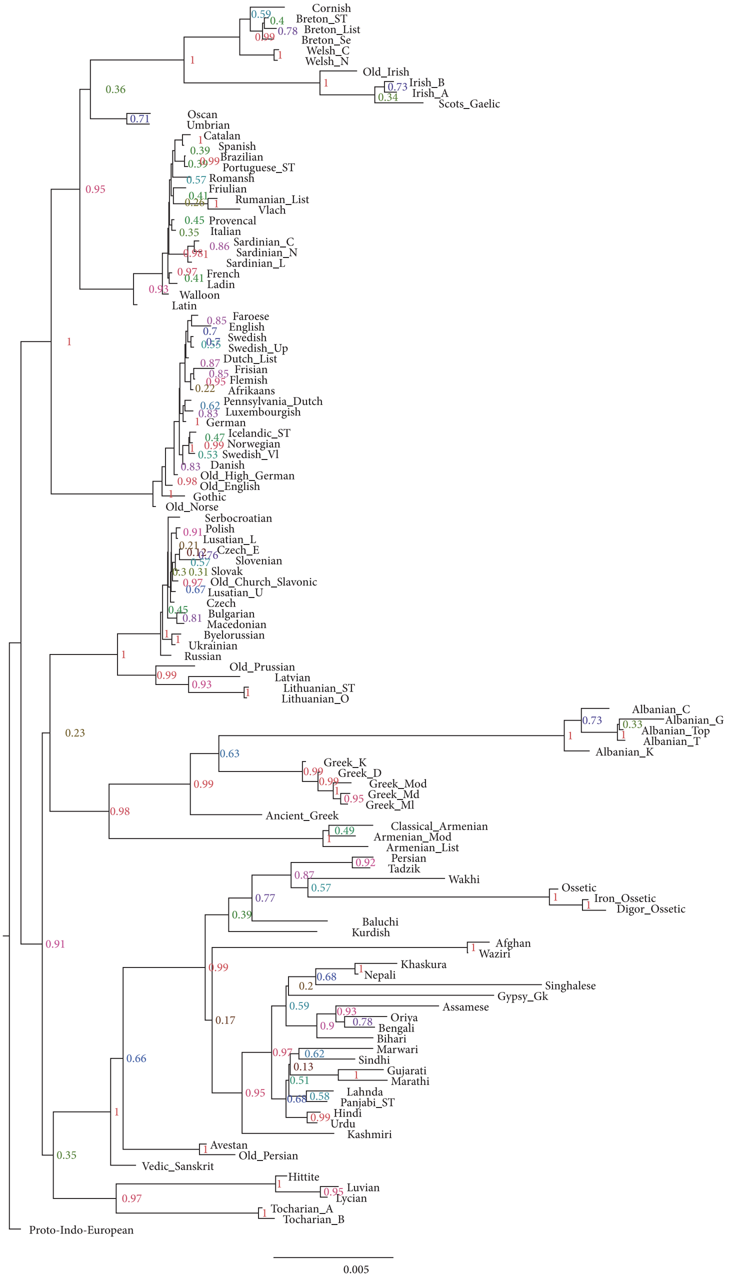


Figure 22. The majority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood for the additional site log-likelihood analysis (Bouckaert et al. 2012)

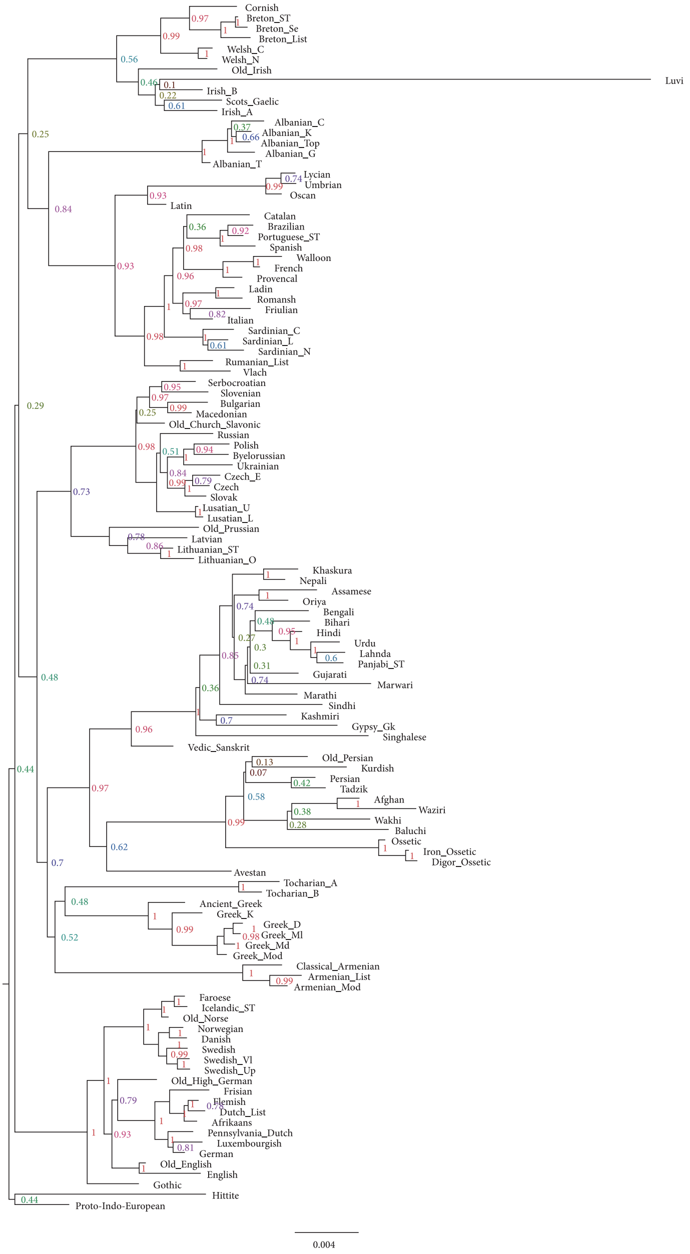


Figure 23. The minority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood for the additional site log-likelihood analysis (Bouckaert et al. 2012)

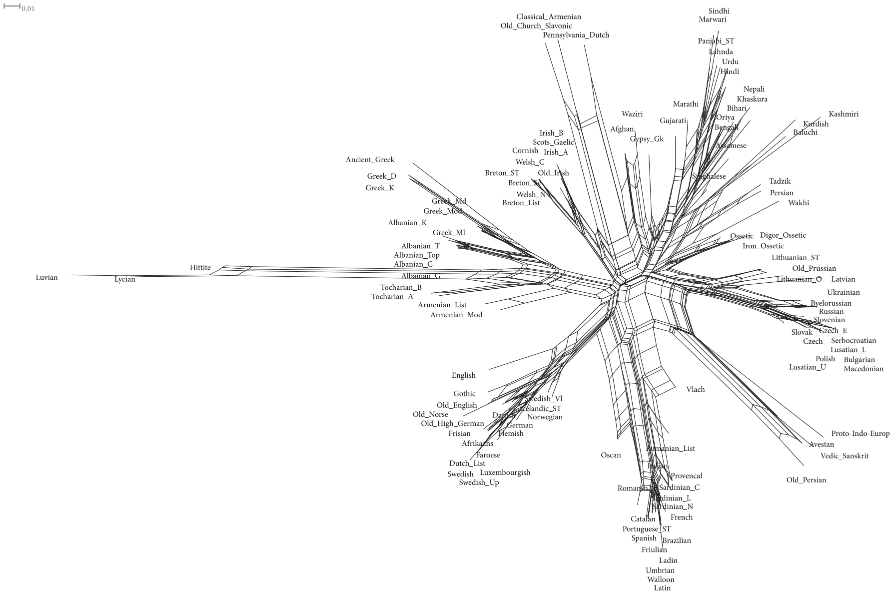


Figure 24. Neighbor Net including only the 97 cognate sets that show a supporting log-likelihood difference of 5 or more for the majority tree in analysis of Indo-European (Bouckaert et al. 2012)

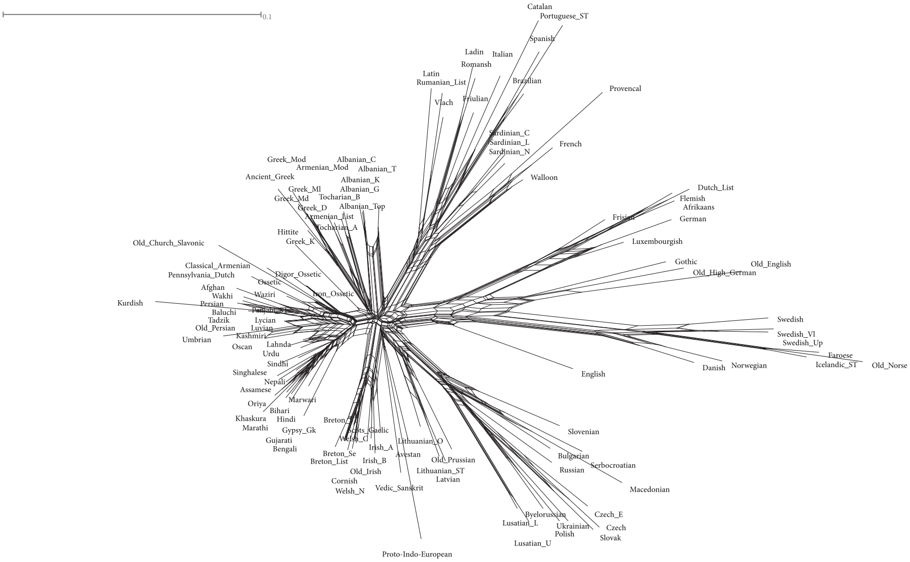


Figure 25. Neighbor Net including only the 586 cognate sets that show a supporting log-likelihood difference of 5 or more for the minority tree in analysis of Indo-European (Bouckaert et al. 2012)

Address for correspondence

Annemarie Verkerk
Max Planck Institute for the Science of Human History
Kahlaische Strasse 10
07745 Jena
Germany
verkerk@shh.mpg.de

Visualizing the Boni dialects with Historical Glottometry

Alexander Elias
Leiden University

This paper deals with the historical relations between dialects of Boni, a Cushitic language of Kenya and Somalia. Boni forms the subject of Volume 10 of the *Language and Dialect Atlas of Kenya* (Heine & Möhlig 1982). Heine presents evidence for three subgroups within Boni, as well as several areas of convergence between dialects belonging to different proposed subgroups. In reviewing his evidence, I find that two of the three splits are not supported by the data, and therefore his conclusions on convergence must also be reinterpreted. Given the presence of numerous intersecting isoglosses, the tree diagram is an inappropriate model for describing the relations between Boni dialects, and I turn to Historical Glottometry (Kalyan & François 2018) to provide a visualization of the data.

Keywords: Boni dialectology, Historical Glottometry, wave model

1. Introduction

1.1 The Boni dialect group

Boni is a dialectally diverse Cushitic (Afroasiatic) language spoken mainly in Lamu District of northern Kenya, with some speakers of the northernmost dialect Kilii over the border in southern Somalia. The 2009 Kenya census counted 7,600 Boni speakers (Kenya Census 2009; reported in Simons & Fennig 2017).

The anthropological information cited below is taken from Stiles (1988). Boni speakers are traditionally hunter-gatherers inhabiting lowland forest areas slightly inland from the coast. The potentially derogatory name "Boni" is proposed to derive from a southern Somali word *bon*, referring to a lower caste of hunter-gatherers in an unequal economic relationship with pastoralist groups. This type of social structure is attested elsewhere in the area (the Midgaan of northern Somalia and the Ribe along the middle Juba River), but the Boni are the only ones with their own language. The Boni of Kenya refer to themselves as "Aweer," but

https://doi.org/10.1075/jhl.18009.eli
Journal of Historical Linguistics 9:1 (2019), pp. 70–91. issn 2210-2116 | e-issn 2210-2124

the Aweer autonym excludes the northernmost Kilii. Thus, following the authors of the *Language and Dialect Atlas of Kenya* (LDAK; Heine & Möhlig 1982), I will use "Boni" to refer to the group as a whole, and "Aweer" to refer to the subset of dialects which excludes the northernmost Kilii dialect.

Boni is most closely related to Somali and Rendille, together forming the Sam subgroup of Eastern Cushitic. Heine places the pastoralist Proto-Sam around 0–300 AD, and proposes that they migrated southwest to the coast from a homeland in inland northern Kenya. They spread along the coast, and the group that became the Boni gave up animal husbandry as an adaptation to the forest environment they found there (Heine 1978 passim). In more recent times, the Boni transitioned from a hunter-gatherer lifestyle to become settled agriculturalists (Heine & Möhlig 1982: 12).

The most divergent dialect of Boni is Kilii, which Heine states is not "immediately intelligible" with the others (Heine & Möhlig 1982: 12). Therefore, it may be justifiable to refer to Kilii as a separate language, but in this paper I will refer to it as a dialect of Boni. The remaining dialects, on the other hand, are not strongly differentiated and remain mutually intelligible. Stiles (1988: 43) dates the split between Kilii and the others to the 16th century, writing the following on the arrival of the Oromo at that time:

> The arrival of the Oromo changed the situation. Whatever organization that previously existed was broken down and people began to migrate south and to the off-shore islands … The Oromo must have been in the area as my informants claim that at this time the Boni would set upon Bajunis and sell them and their property to the Oromo. Before this time the Boni say that the Oromo were not in Shungwaya. The Boni who remained with the Bajunis when they fled to the islands are today known as the Wakatwa; the Boni who stayed in the southern Somalia forests are called Kilii, and the Boni who later associated with the Somalis are called Garra … The dialect differentiation between Kilii and the other Boni dialects which Heine [& Möhlig] (1982) notes probably occurred at this time, i.e. in the 16th century.

Thus, the time depth of differentiation of the Boni dialects is not very great, and this is reflected in their many similarities. Importantly, Stiles (1988) proposes a concrete historical event, the invasion of the Oromo, which led to a physical split in the Boni community, with the Kilii dialects remaining in southern Somalia and the remaining dialects migrating south to their present locations in northern coastal Kenya. The migrating group ("Aweer" in this paper) remained relatively cohesive as they moved south to their current area in Lamu District and entered into relations with the Oromo people there. The Kijee are said to be made up of Oromo people who joined the Boni, and the word *Kijee* means "Oromo" in Boni (Stiles 1988: 43–44).

1.2 Data source

The data in this paper are drawn entirely from Volume 10 of the *LDAK*, which deals specifically with the Boni dialects. Prior to this publication, it was known that Boni was dialectally diverse, but work had focused only on the Bireeri dialect (Heine 1977; Sasse 1979). This volume of *LDAK* presents the results of fieldwork carried out by Heine on the dialects Jara and Kilii, with comparative data from the other dialects. The first section is a compact overview of the phonology and morphology of Jara and Kilii, and the second section deals with relations between dialects and reconstruction of Proto-Boni. The map in Figure 1 shows the layout of the traditional Boni territorial divisions along the coast of Kenya and Somalia. Data from six of these ten divisions are presented in the *LDAK* – Bireeri (1), Safaree (4), Kijee (6), Jara (7), Baddey (8), and Kilii (10).

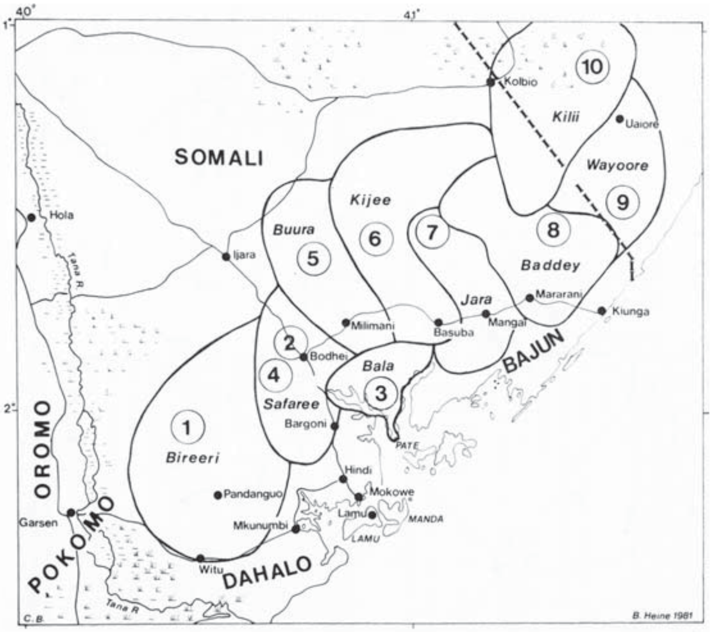


Figure 1. Map of traditional Boni territorial divisions (reproduced from Heine & Möhlig 1982: 14). The dotted line in the top right corner is the border between Kenya and Somalia

Heine & Möhlig (1982: 17–18) listed 13 people as their main informants. I list them in Table 1 with their initials, sex, age, place of birth/residence, and the name of their native dialect, when that information was included. No further details were provided on how the data were collected.

Table 1. List of informants in Heine & Möhlig (1982: 17–18)

Initials	Sex	Age	Residence	Dialect
AD	male	25	Basuba	Bala
KB	male	32	Mangai	Kijee
BB	male	35	Badada	Kilii
FH	female	18	Mangai	not listed
AM	female	19	Mangai	not listed
JC	male	18	Mangai	not listed
OJ	male	16	Mangai	Jara
OM	male	18	Mangai	Kijee
MB	male	11	Bargoni	Safaree
SA	male	13	Bargoni	Bala
HB	male	32	Bargoni	Safaree
AK	male	33	Bargoni	Buura
GW	male	35	Bargoni	Kijee

The Boni dialects are typologically almost identical to each other. The basic word order is SOV and modifiers usually follow their head. They have large consonant inventories, which include ejective voiceless and implosive voiced stops (though the latter are not present in all dialects). All Boni dialects show vowel harmony between a set of tense and lax vowels, and they have a two-tone system where each syllable takes either a high or low tone. They show complex interactions between gender and number (termed "gender polarity," following Meinhof 1912) often seen in Cushitic languages.

Boni speakers are a minority in both Kenya and Somalia, and as a result they are often competent in a local lingua franca. A majority of Boni speakers of both genders are competent in Swahili (Bantu), the lingua franca of East Africa. This has introduced loanwords from Bantu into all Boni dialects to varying degrees. There are also two Cushitic languages commonly spoken by Boni speakers: Oromo and Somali. Most male Boni speakers are competent in one of these two languages, and the influence of Oromo is clearest in the peripheral dialects (Bireeri and Kilii). Although English is the national language of Kenya, less than 3% of Boni men were competent in English at the time of writing of the LDAK (Heine & Möhlig 1982: 12).

1.3 Theoretical background

Not all language groups are equally well represented by a tree diagram. The tree diagram is most appropriate when the language group has undergone a series of splits, whereby one group of speakers loses contact with the rest and both groups undergo a period of exclusively shared innovation in isolation from each other. These splits correspond to nodes on a tree diagram and are understood to reflect real events of societal splitting in the ancestral speech community. In this idealized scenario, isoglosses will never intersect, and distinguishing between vertical and horizontal transmission is usually not problematic. However, this pattern of strictly non-intersecting innovations will only hold in the ideal case. In cases where there were no abrupt splits in the speech community, there can indeed be intersecting isoglosses which complicate construction of a tree diagram. Furthermore, since nodes on a tree diagram correspond to actual splitting events, the tree diagram has nothing to represent if the community did not undergo any splits. Overreliance on the tree model leads to the proposal of unrealistic splits, and often requires the linguist to sweep certain problematic data under the rug in order to arrive at a strictly nested set of isoglosses.

The problematic reasoning which the tree model can lead to is found in Heine & Möhlig (1982). For instance, Heine proposed the subgroup "Central Aweer" consisting of Safaree, Jara, and Kijee based on the fact that the Proto-Boni distal demonstrative /*-óóhe/ changes to /*-óóho/ in those three dialects (Heine & Möhlig 1982: 81). Heine then noted that a number of innovations group the dialects of Bireeri, Safaree, and Kijee together, to the exclusion of Jara. These innovations include the replacement of the Proto-Boni negative perfect prefix /*mə-/ with /húú-/ (< /*hááb-/) and the change of /*b/ > /w/ word-internally, at least in some words (Heine & Möhlig 1982: 85). I have dubbed the group of languages affected by these innovations the "Southern Convergence Area" in this paper, whereas Heine simply referred to the "North-Central isogloss." Heine was agnostic as to whether this set of innovations reflected a genetic or areal effect, saying its "historical significance remains to be investigated" (Heine & Möhlig 1982: 85). However, it is implicitly assumed by the tree model that any non-genetic group must arise by later convergence, hence the choice of name. The evidence provided in favor of Central Aweer (one irregular sound change in a demonstrative) is in fact weaker than that adduced for the Southern Convergence Area (one morphological innovation and one semi-regular sound change), but in the process of constructing a nested set of innovations, Heine was forced to discard the evidence in favor of the Southern Convergence Area and treat it as an unexplained phenomenon.

This paradox arises mainly from the impossibility of distinguishing vertical and horizontal transmission in such a closely-knit group of dialects. The tree diagram fails to accurately capture the linguistic history of groups which differentiate while maintaining contact and mutual intelligibility with closely related speech varieties. Such a situation will lead to the formation of a dialect chain at first, and if the process of differentiation continues, a linkage (as defined in Ross 1988). In a dialect chain or linkage, intersecting isoglosses are an empirical fact which simply cannot be accounted for under the tree model. The wave model provides an alternative which can account for these intersecting isoglosses and provide a principled way to avoid Heine's paradox in the first place: in the wave model, linguistic change is viewed as a set of ripples spreading in a pond (an image put forth by the first proponent of the wave model, Schmidt 1872). A change begins at an innovative center and spreads to neighboring speech varieties. The change can continue spreading as long as there is contact and some level of intelligibility or at least bilingualism between neighboring populations. Each change can spread over a different portion of the whole speech area, resulting in the pattern of intersecting isoglosses commonly observed in dialect chains. This is a much more natural and realistic account of differentiation in cases where there is no reason to suspect that any major splitting events have separated speakers from one another. In addition, the wave model has the advantage over the tree model in that tree-like language groups are simply a special case in which isoglosses happen not to intersect.

The reason that the tree model remains so dominant over the wave model is that the tree model provides an intuitive visual representation in the form of a tree diagram. The wave model has seen numerous proposals for visualizations, including trees modified with double lines (Ross 1988), isogloss maps (Anttila 1989), and NeighborNets (Bryant, Filimon & Gray 2005; Bryant & Moulton 2003). These suffer from various drawbacks of either informativity or readability. One recent attempt to improve on these methods is the glottometric diagram in the framework of Historical Glottometry (HG). This is an ongoing project led by Alexandre François and Siva Kalyan, laid out in a series of recent publications (François 2014; François 2017; Kalyan & François 2018). I have chosen to recast the data contained in the *LDAK* in the framework of HG and to present the findings in the form of a glottometric diagram. Boni makes a good case study for developing methods of visualizing non-tree-like language groups because the conditions for the overlapping spreading of innovations are present: mutual intelligibility and continued contact.

2. Review of Heine's evidence

In this section I will present the reconstructed phonology of Proto-Boni and the evidence presented by Heine in support of the various groupings within Boni that he identifies. He first presents the divergent changes that he takes as evidence for subgroups within Boni, followed by convergent changes which intersect the subgroup boundaries. Based on the evidence provided, I conclude that only one of the three proposed subgroups is strongly supported by linguistic (and historical) evidence. I assign numbers to each of the 25 innovations he identifies, and this body of evidence will then form the input for the creation of a glottometric diagram.

2.1 Proto-Boni phonology

According to Heine, Proto-Boni phonology is unproblematic to reconstruct because the modern dialects agree quite closely with each other (Heine & Möhlig 1982: 71). He reconstructs the inventories of consonant and vowel phonemes found in Tables 2 and 3.

Table 2. Reconstructed Proto-Boni consonant inventory (Heine & Möhlig 1982: 71)

	Labial	Dental/Alveolar	(Alveo)palatal	Velar	Uvular	Glottal
Stop	*p *b	*t *d	*c *ɟ	*k *g	*q	*ʔ
Ejective		*t’	*c’	*k’		
Implosive		*ɗ	*ʄ	*ɠ		
Fricative	*f	*s	*ʃ			*h
Trill		*r *ɽ				
Lateral		*l				
Nasal	*m	*n	*ɲ	*ŋ		
Glide	*w		*j			

Table 3. Reconstructed Proto-Boni vowel inventory (Heine & Möhlig 1982: 71)

	Front		Central		Back	
	Lax	**Tense**	**Lax**	**Tense**	**Lax**	**Tense**
Close	*i	*i̱			*u	*u̱
Mid	*e	*e̱	*ə	*ə̱	*o	*o̱
Open			*a	*a̱		

Each vowel comes in a lax-tense pair, and there is a system of vowel harmony where lax vowels can assimilate to tense ones across morpheme boundaries. Tense vowels are underlined in this paper (/i̱ e̱ a̱ o̱ u̱/). Phonetically, the tense vowels are close to the cardinal vowels while the lax vowels are more centralized. Vowels can appear long or short, and the long vowels are indicated by doubled spelling. It is also possible to reconstruct a two-tone system where each syllable can take a high or low tone. In this paper, the high tone is indicated by an acute accent (/v́/) while the low tone is left unmarked.

2.2 Proposed splits

2.2.1 *Aweer*

The first proposed split is between Kilii and Aweer (consisting of the other five dialects: Bireeri, Safaree, Jara, Kijee, and Baddey) (Heine & Möhlig 1982:78). This evidence is strong enough to justify Aweer as a valid subgroup, created by a genuine split between Kilii and Proto-Aweer. The linguistic evidence of a split is corroborated by historical evidence of an Oromo invasion in the 16th century (see Section 1.1, and Stiles 1988) which caused the Proto-Aweer speakers to flee south, while the Kilii speakers stayed in place. The nine innovations seen in Tables 4–12 support this split. After each innovation, the page number where it is introduced is listed, as well as the page numbers of each example in the wordlist at the end.

Table 4. Illustration of innovation #1 (Heine & Möhlig 1982:78, 135)

Innovation #1: Proto-Boni *nt > Proto-Aweer *t, Kilii /n/

Gloss	Proto-Boni	Proto-Aweer	Kilii
‘to vomit’	*məntah-	*mətah-	*manah-*

Table 5. Illustration of innovation #2 (Heine & Möhlig 1982:78, 106, 115)

Innovation #2: Proto-Boni *dt > Proto-Aweer *dd, Kilii /tt/

Gloss	Proto-Boni	Proto-Aweer	Kilii
‘fingernail, claw’	*idtíʔi	*iddíʔi	*ittíi*
‘leg’	*ídtə	*íddə	*íttə*

Table 6. Illustration of innovation #3 (Heine & Möhlig 1982:79, 90, 102, 99)

Innovation #3: Proto-Boni *m > Proto-Aweer *w, *j between unlike vowels, deleted between like vowels

Gloss	Proto-Boni	Proto-Aweer	Kilii
‘to destroy, spoil’	*humeej-	*huwej-	*humeej-*
‘to come (impf.)’	*-imaad-	*-ijaad-	*-imaad-*
‘to be angry’	*umuʃooɗ-	*uuʃooɗ-	*umusooʔ-*

Table 7. Illustration of innovation #4 (Heine & Möhlig 1982:79, 104, 107)

Innovation #4: Proto-Boni *m > Proto-Aweer *ŋ morpheme-finally

Gloss	Proto-Boni	Proto-Aweer	Kilii
‘to eat’	*-aham-	*-ahaŋ-	*-aham-*
‘food made of flour’	*məsím-ə	*məsíŋə	*məsímə*

Table 8. Illustration of innovation #5 (Heine & Möhlig 1982: 79, 108, 116)

Innovation #5: Proto-Boni *t’ > Proto-Aweer *d, Kilii /t/

Gloss	**Proto-Boni**	**Proto-Aweer**	**Kilii**
‘forest’	*t’úur	*dúur	*túur*
‘maggot’	*t’iʔírə	*diʔírə	*tihírə*

Table 9. Illustration of innovation #6 (Heine & Möhlig 1982: 79, 134)

Innovation #6: Proto-Boni *kʃ > Proto-Aweer *ʃ, Kilii /ks/

Gloss	**Proto-Boni**	**Proto-Aweer**	**Kilii**
‘topi’	*ha̱wkʃé̱é̱	*ha̱wʃé̱e̱	*hoksé̱*

Table 10. Illustration of innovation #7 (Heine & Möhlig 1982: 79, 117)

Innovation #7: Proto-Boni *mb > Proto-Aweer *b

Gloss	**Proto-Boni**	**Proto-Aweer**	**Kilii**
‘marabou stork’	*bá̱á̱mbo̱	*bá̱á̱bo̱	*bá̱á̱mbo̱*

Table 11. Illustration of innovation #8 (Heine & Möhlig 1982: 79, 90, 91, 107)

Innovation #8: Low vowels assimilate to a following non-low vowel in some words

Gloss	**Proto-Boni**	**Proto-Aweer**	**Kilii**
‘ant species’	*ʄaríira	*ʄiríira	*ʄaríira*
‘lower arm’	*ta̱gó̱g	*tó̱gó̱g	*ta̱gó̱g*
‘flower, blossom’	*barúúrə	*burúúrə	*barúúrə*

Table 12. Illustration of innovation #9 (Heine & Möhlig 1982: 79, 97, 99)

Innovation #9: Vowels are rounded before *b or *w (in some words)

Gloss	**Proto-Boni**	**Proto-Aweer**	**Kilii**
‘to catch, seize’	*qabəɗ-	*ɠobəɗ-	*k’abaʔ-*
‘civet cat’	*símáad	*suwáad	*simá̱a̱d*

2.2.2 *South Aweer*

The second proposed split is between Baddey and South Aweer (consisting of Bireeri, Safaree, Jara, and Kijee) (Heine & Möhlig 1982: 79). The evidence for this as a genuine split is weak, and therefore the existence of a South Aweer subgroup is not strongly supported. Only one piece of unproblematic evidence is adduced, displayed in Table 13.

Table 13. Illustration of innovation #10 (Heine & Möhlig 1982: 81, 90)

Innovation #10: Proto-Boni *w is deleted between two /a/'s in Proto-South Aweer

Gloss	Proto-Boni	Proto-South Aweer	Baddey
'harmless edible animal'	*buʃáwa	*buʃáa	*buʃáwa*

Two more pieces of evidence are adduced by Heine in support of the subgroup South Aweer, but one of them is in fact a shared retention (Proto-Boni *ɟ > /j/ in Baddey, but is retained in South Aweer; Heine & Möhlig 1982: 81). Furthermore, the regular shift of Proto-Boni *ɠ > *ʔ in South Aweer which he cites as evidence (Heine & Möhlig 1982: 79) is questionable because the same change also affects Kilii. Kilii may well have innovated this independently, since there is a parallel change of Proto-Boni *ɗ > /ʔ/ in Kilii as well (Heine & Möhlig 1982: 78). If it is the case that these two innovations are the same by coincidence, then this is a second shared innovation in support of South Aweer. However, I have left it out of the final analysis, given the uncertainty surrounding it.

2.2.3 *Central Aweer*

The third proposed split is between Bireeri and Central Aweer (Safaree, Jara, and Kijee) (Heine & Möhlig 1982: 81). The evidence for this as a genuine split is also weak, and therefore the existence of a Central Aweer subgroup is not strongly supported. It consists of a single piece of unproblematic evidence, seen in Table 14.

Table 14. Illustration of innovation #11 (Heine & Möhlig 1982: 81, 75)

Innovation #11: Vowel assimilation in distal marker

Gloss	Proto-Boni	Proto-Central Aweer	Bireeri
'that, distal'	*-óóhe	*-óóho	(Missing)

The other evidence offered is a set of consonant cluster simplifications which occur only in Bireeri and thus in fact constitute a shared retention in Proto-Central Aweer.

2.3 Proposed convergence areas

In the next section, Heine presents evidence for changes which cut across the boundaries of the genetic groups previously established. The implicit hypothesis of the tree model is that these must have arisen by areal convergence, and this is indeed how Heine presents the peripheral convergence with Oromo (Section 2.3.4) and the Central Convergence Area (Section 2.3.3). As for Southern Convergence Area (Section 2.3.1) and South-Central Convergence Area (Section 2.3.2), these are newly coined names; Heine refers to them as the "South-Central isogloss" and "North-Central isogloss" respectively (Heine & Möhlig 1982: 85). He is agnostic as to whether they are the result of genetic splitting or areal convergence, but I have listed them here as convergence areas because that is the implicit position taken, assuming they are non-genetic groups.

2.3.1 *Southern Convergence Area*

The neighboring pair of southern dialects Bireeri and Safaree form the Southern Convergence Area (bounded by the "South-Central isogloss" in Heine & Möhlig 1982: 85). They share a common morphological innovation, whereby the negative imperative plural suffix is replaced by the affirmative imperative suffix. They also have four lexical innovations in common (see Table 15).

Table 15. Lexical innovations #13–16 affecting the Southern Convergence Area (Heine & Möhlig 1982: 82, 132, 126, 134, 137)

Innovation #12: Proto-Boni *-náha 'neg. imperative pl.' is lost in favor of affirmative *-níja (Heine & Möhlig 1982: 85, 75–76)

Gloss	Proto-Boni	Jara, Baddey	Kilii	Kijee	Bireeri	Safaree
#13 'tamarind'	*múkaj	*múkaj*	*múkaj*	(Missing)	*róɠa*	*róka*
#14 'sheep'	*aléeŋ	*aléeŋ*	*aléeŋ*	(Missing)	*hoolá*	*hoola*
#15 'to touch'	*taar-	*taar-*	*taab-aʔ-*	(Missing)	*k'ol-, taar-*	*k'ol-, ɠol-*
#16 '2pl'	*ado	*ado*	*ado*	*ado*	*isanə*	*is(i)no*

2.3.2 *South-Central Convergence Area*

These dialects are roughly those which lie south of the Dodori river: Bireeri, Safaree, and Kijee. They are bounded by Heine's "North-Central isogloss" (Heine & Möhlig 1982: 85). There are at least two innovations which they share, displayed in Table 16.

Table 16. Illustration of innovation #18 (Heine & Möhlig 1982: 85, 97, 100)

Innovation #17: Proto-Boni *mə- 'neg. perfect' > scca /húú-/ (Heine & Möhlig 1982: 85, 76)
Innovation #18: Proto-Boni *b > scca /w/ word internally

Gloss	Proto-Boni	scca	Jara	Baddey	Kilii
'to catch, seize'	*qabəɗ-	*owɗ-*	*obəɗ-*	*ɠobɗ-*	*k'abaʔ-*
'to curse'	*habaaɽ-	*hawaar-*	*habaar-*	(Missing)	*habaaɽ-*

2.3.3 *Central Convergence Area*

This area covers the dialects Safaree, Kijee, Jara, and Baddey (all except the peripheral dialects Bireeri and Kilii) (Heine & Möhlig 1982: 82). There are five shared innovations cited for this group of dialects, including two instances of shared loans from Bantu, seen in Tables 17–20.

Table 17. Illustration of innovation #19 (Heine & Möhlig 1982: 84, 131)

Innovation #19: Proto-Boni *b > CCA /f/ word-finally (in some words)				
Gloss	**Proto-Boni**	**CCA**	**Bireeri**	**Kilii**
‘small sword’	*gólhob	*gólof*	*gólob*	*gólhob*

Table 18. Illustration of innovation #20 (Heine & Möhlig 1982: 84, 134)

Innovation #20: Proto-Boni *lk > CCA /rk/				
Gloss	**Proto-Boni**	**CCA**	**Bireeri**	**Kilii**
‘teeth’	*ilkə	*irki, irkə*	*ilkə*	*ekə́*

Table 19. Illustration of innovation #21 (Heine & Möhlig 1982: 84, 101, 128)

Innovation #21: Proto-Boni *ɠt > CCA /gt/, Bireeri, Kilii /tt/				
Gloss	**Proto-Boni**	**CCA**	**Bireeri**	**Kilii**
‘darkness’	*múɠtə	*múgti, múgtə*	*múttə*	*múʔuttə*
‘sisal-like plant’	*ɗeɠté	*ɗégtə*	(Missing)	*etté*

Table 20. Lexical innovations #22–23 affecting the Central Convergence Area (Heine & Möhlig 1982: 84, 95, 111)

Two lexical innovations (shared loans from Bantu)				
Gloss	**Proto-Boni**	**CCA**	**Bireeri**	**Kilii**
#22 ‘blacksmith’	*tumal	*fuundə, fúúndi*	*túma*	*tumáal*
#23 ‘hare’	*hileesə	*kituŋgúe*	*híleesə*	*hiléésə*

2.3.4 *Peripheral convergence with Oromo*

This is a fundamentally different type of convergence which affects the peripheral dialects Bireeri and Kilii most. They are both under particularly strong influence from Oromo, according to Heine & Möhlig (1982: 84–85). They are not directly converging with each other, but rather converging on the same target, leading to shared innovations between them. Heine lists two Oromo loanwords shared exclusively by the peripheral dialects as evidence for this process, shown in Table 21.

Table 21. Lexical innovations #24–25 affecting the peripheral dialects Bireeri and Kilii (Heine & Möhlig 1982: 84, 99, 118)

Two lexical innovations (shared loans from Oromo)				
Gloss	**Oromo**	**Bireeri**	**Kilii**	**Other**
#24 'cloud'	*duumans, dúmes*	*duumánsə*	*duumə́ssə*	(Missing)
#25 'vervet monkey'	*k'amaleʔ*	*k'amála*	*kamá̱le̱*	tabáábu, (Missing)

2.4 Assessment of evidence

Heine proposes three nested subgroups within Boni: Aweer, South Aweer, and Central Aweer. He then proposes the existence of convergence zones to explain the fact that many isoglosses cut across the boundaries of the proposed subgroups. Heine's view of the splits within Boni is summed up by the tree below. Heine does not summarize his proposal on the genetic splits in a tree diagram, so I have created a tree which represents his subgrouping proposal, displayed in Figure 2.

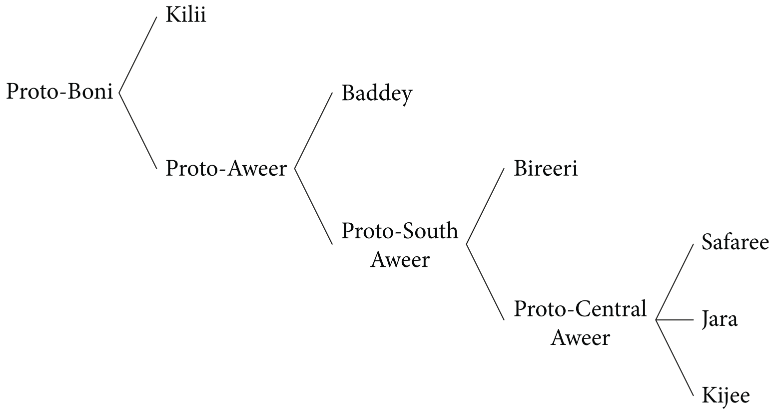


Figure 2. Heine's view of the splits within Boni (my interpretation)

However, given the strength of the evidence presented, only the existence of the first subgroup (Aweer) is well-supported. This is based on both linguistic and historical evidence of a split in the Proto-Boni community due to the invasion of the Oromo, dated to approximately the 16th century (Stiles 1988). The evidence for the proposed South Aweer and Central Aweer groups consists of a single innovation in both cases (possibly two in the case of South Aweer), and neither is strong enough to confidently identify a genuine splitting event in the history of Boni. Furthermore, no major historical event like the arrival of the Oromo can be identified which could potentially have caused such splits. Therefore, Heine's analysis that Proto-Boni split into Kilii and Proto-Aweer is well-supported, but there is not strong enough evidence to posit tree-like subgroups below the level of Proto-Aweer.

3. Glottometric analysis

I conclude two things from reviewing Heine's evidence. Firstly, I conclude that there was indeed a genuine split between speakers of Kilii and Proto-Aweer. Secondly, I conclude that below the level of Proto-Aweer there is no evidence for genuine splits in the Boni-speaking community. Instead, the differences between modern dialects are best understood as the result of wave-like processes of language differentiation, enabled by the continued mutual intelligibility and contact between these dialects. Given that there is only one well-supported split, a tree diagram would contain only a single node at Proto-Aweer and would not be the most informative method of representing the history of the Boni dialects. It is desirable to represent information not only about the splits, but also about the overlapping waves of change which have targeted different sets of dialects. I will now explain the methodology behind creating a glottometric diagram using an example calculation.

3.1 Methodology

The database of innovations which formed the input for the diagram in this case looks like the spreadsheet displayed in Table 22, in which cells are shaded to represent their participation in an innovation.

From this database, it is possible to calculate the following values for any set of languages or dialects (as defined in Kalyan & François 2018):

- ε = number of exclusively shared innovations
- p = number of supporting innovations
- q = number of conflicting innovations

The concept of an exclusively-shared innovation is already familiar. A supporting innovation is one that affects all the languages or dialects under investigation, but may affect others as well. The important thing is that all the languages or dialects in question innovated together. A conflicting innovation is one where at least one member of the set of languages or dialects participates in innovations with at least one language outside the set. However, the conflicting innovation cannot target all members of the set in question. These conflicting innovations support a different subgrouping and undermine the strength of the evidence for any given subgroup. Taking the set of dialects Bireeri and Safaree as an example (the Southern Con-

Table 22. Database of shared innovations among Boni dialects; shading indicates participation in an innovation

Innovation	Bireeri	Safaree	Jara	Kijee	Baddey	Kilii
#24	▨					▨
#25	▨					▨
#10	▨	▨	▨	▨		
#1	▨	▨	▨	▨	▨	
#2	▨	▨	▨	▨	▨	
#3	▨	▨	▨	▨	▨	
#4	▨	▨	▨	▨	▨	
#5	▨	▨	▨	▨	▨	
#6	▨	▨	▨	▨	▨	
#7	▨	▨	▨	▨	▨	
#8	▨	▨	▨	▨	▨	
#9	▨	▨	▨	▨	▨	
#19		▨	▨	▨	▨	
#20		▨	▨	▨	▨	
#21		▨	▨	▨	▨	
#22		▨	▨	▨	▨	
#23		▨	▨	▨	▨	
#11		▨	▨	▨		
#17	▨	▨		▨		
#18	▨	▨		▨		
#16	▨	▨				
#13	▨	▨				
#14	▨	▨				
#15	▨	▨				
#12	▨	▨				

vergence Area according to Heine), it is possible to observe that they share five exclusive innovations (#12–16). A further 12 innovations affect Bireeri and Safaree as well as other dialects, for a total of 17 supporting innovations (since exclusively shared innovations count as supporting innovations as well). Finally, Bireeri and Safaree have eight conflicting innovations (#11, 19–25) where one of the pair innovates with some other group of dialects instead. Thus, the values for the Southern Convergence Area are $\varepsilon = 5$, $p = 17$, $q = 8$.

Based on these values, the two final quantities which will serve in the creation of the glottometric diagram can be computed. The first is cohesiveness (κ), and it measures the tendency of a set of languages or dialects to innovate together rather than separately, on a scale from 0 to 1. It is equal to the number of supporting innovations over the total number of relevant innovations (that is, both supporting and conflicting ones; see Kalyan & François 2018):

$$\begin{aligned}\kappa &= (\text{supporting innovations}) / (\text{total relevant innovations})\\ &= p / (p+q)\\ &= 17 / (17+8)\\ &= 0.68\end{aligned}$$

Note that under the tree model, all subgroups must receive cohesiveness 1 because conflicting innovations are not recognized as valid. Cohesiveness is typically high in a tree-like subgroup (one which has differentiated through splitting and isolation) and low in a non-tree-like subgroup or dialect chain situation.

The second quantity to be calculated is subgroupiness (ς). This is simply equal to the number of exclusively shared innovations times the cohesiveness of the set (as defined in Kalyan & François 2018):

$$\begin{aligned}\varsigma &= (\text{cohesiveness}) \times (\text{exclusively shared innovations})\\ &= \kappa \times \varepsilon\\ &= 0.68 \times 5\\ &= 3.4\end{aligned}$$

Subgroupiness is a way of weighting the quantity of the evidence in favor of a subgroup by the quality of that evidence. Exclusively shared innovations are the only acceptable source of evidence in normal linguistic classification, so this method of analysis is orthodox in that regard. However, it adds the recognition that some language groups behave less cohesively than others, and penalizes them as a result. Note that subgroupiness, unlike cohesiveness, has no maximum value, and a larger database with more innovations will lead to larger subgroupiness scores.

When constructing a database of innovations, all types of shared innovations are acceptable. This includes sound changes (both regular and irregular), as well as innovations in morphology, syntax, semantics, and lexicon. Normally, these types of innovations are not taken to be equally indicative of subgrouping. Shared irregularities in morphological paradigms and uncommon sound changes are usually given great weight in language classifications, on the basis that they are less likely to spread through horizontal transmission or be replicated by chance. However, this often turns into a blunt tool for choosing which data are to be used to establish a "genuine subgroup," and which can be safely discarded. There is no need to discard common sound changes or shared loanwords in the construction of the database, because in the wave model, all innovations spread horizontally. The fact of a common change (like /s/ > /h/) spreading through contact is no less indicative of convergence and shared innovation than a rare sound change. Of course, instances of random parallel innovation, in which the same innovation occurs twice by chance, are still unacceptable as subgrouping evidence.

One solution, which would preserve the intuition that rare innovations are more solid indicators of language history than common changes or lexical loans, would be to weight innovations differently according to their type and rarity. However, it would be crucial to avoid the types of intuitive judgments about "common" and "uncommon" changes which we are taking such pains to avoid. In order to do this in an objective way, a large database of innovations from around the world would need to be compiled, but it would be extremely hard to ensure that this database is representative. In view of the difficulties of implementing such a weighting system correctly, it is preferable to simply stick to unweighted values. More information on the possibility of weighting, and reasons for not doing it, can be found in François (2014: 176–177).

3.2 Results

The full table of subgroupiness and cohesiveness values calculated for each group is given below. These values can be calculated from the database of innovations given in Table 23. The sets of dialects are listed in order of subgroupiness, and only groups which received a subgroupiness score greater than 0 are included.

In order to go from these raw values to a glottometric diagram, it is necessary to arrange the dialects in a convenient layout, then surround the various subgroups with lines representing bundles of isoglosses. The thickness of the encircling line is proportional to subgroupiness, while its darkness is proportional to cohesiveness. It can be very tricky to figure out the best configuration, since there is currently no automated way of generating these diagrams. The result is shown in the glottometric diagram given in Figure 3.

The dialects represented are Bireeri (BIR), Safaree (SAF), Kijee (KIJ), Jara (JAR), Baddey (BAD), and Kilii (KIL). On the top left is Oromo, represented here because it exerts an influence on Bireeri and Kilii especially. The blue patch extending from Oromo represents the influence which has extended to affect the peripheral Boni dialects Kilii and Bireeri especially strongly, and is an ad hoc addition to the diagram.

If the dialects are displayed on a map true to their geographic locations, the resulting visualization is called a glottometric map instead. The configuration of Boni dialects in a chain along the coast makes it relatively easy to transform the glottometric diagram into a glottometric map, shown in Figure 4.

Table 23. Cohesiveness and subgroupiness scores for all groups which have at least one exclusive innovation

	Exclusive Innovations	Supporting Innovations	Conflicting Innovations	Cohesiveness	Subgroupiness
Aweer	9	9	2	0.82	7.38
SCA	5	17	8	0.68	3.4
CCA	5	14	8	0.64	3.2
SCCA	2	12	8	0.6	1.2
CA	1	16	7	0.7	0.7
SA	1	10	7	0.59	0.59
Peripheral	2	2	17	0.11	0.22

Guide to names of dialect groups:

Aweer	Bireeri, Safaree, Jara, Kijee, Baddey
SCA	Bireeri, Safaree (Southern Convergence Area)
CCA	Safaree, Jara, Kijee, Baddey (Central Convergence Area)
SCCA	Bireeri, Safaree, Kijee (South Central Convergence Area)
CA	Safaree, Jara, Kijee (Central Aweer)
SA	Bireeri, Safaree, Jara, Kijee, Baddey (South Aweer)
Peripheral	Bireeri, Kilii

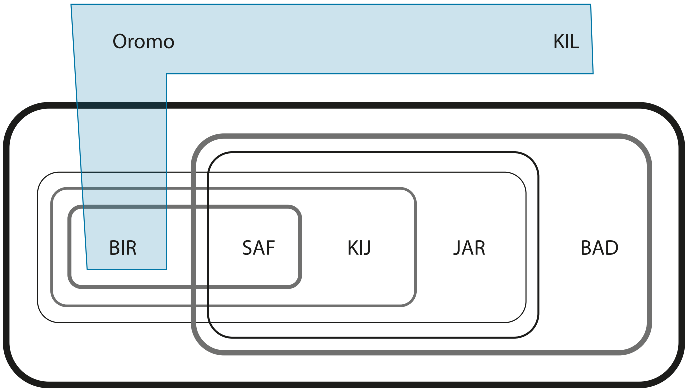


Figure 3. Glottometric diagram of the Boni dialects

3.3 Discussion

A few observations can be made about this glottometric diagram (and map). First of all, the thickest and darkest bundle of isoglosses separates Kilii from the remaining dialects. This is the only subgroup for which the available linguistic (and historical) evidence of a split is convincing, and this is reflected by the thickness and

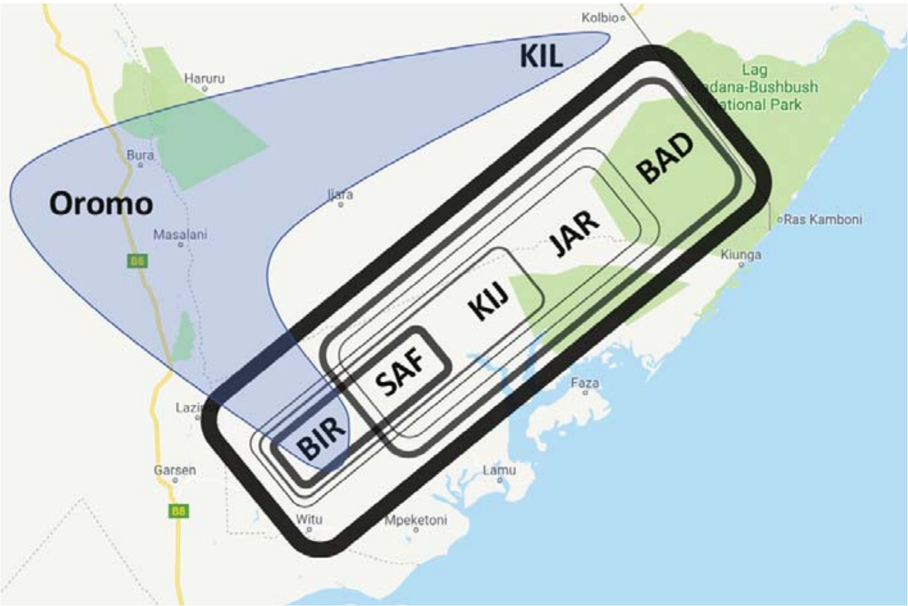


Figure 4. Glottometric map of the Boni dialects (map data copyright of Google 2018)

darkness of the line separating Kilii from the rest. Second, the lines which surround various subgroups of remaining dialects are rather light and thin. They all appear to be of a similar darkness because they all have cohesiveness scores in a narrow range between 0.59–0.7. This reflects the fact that the evidence for any one subgroup is quite slim, and any proposed subgroup can be undermined with other data. They all receive approximately equal levels of support from the evidence provided. Instead of making an arbitrary choice as to which set of evidence defines a "true subgroup" and which are "contact effects," this can be understood as the normal situation for a group of closely related dialects which remain in contact.

Much information that would be lost in a tree diagram is retained in this glottometric diagram. For instance, the close relationship of Bireeri and Safaree is obscured by the fact that Heine judges Bireeri to have branched off quite early. In the glottometric diagram, the dual nature of Safaree is clear. The outline of what Heine calls the Central Convergence Area (Safaree, Kijee, Jara, Baddey) is clearly visible in the diagram. Another beautiful pattern which can be observed is the influence spreading out from Bireeri. Heine himself mentions that there seems to be a set of innovations spreading from Bireeri which affect each successive neighbor less strongly. This pattern can clearly be seen in the glottometric diagram, where Bireeri has a series of lines surrounding it which grow thinner and lighter with distance. This illustrates very nicely the metaphor so often used by the wave model of linguistic change as ripples spreading in a pond. The subgroups proposed by Heine of South Aweer and Central Aweer are not surrounded by a prominent line, because the evidence supporting them is quite weak. Finally, Oromo was placed apart from the Boni dialects in a different shade, in order to represent the common innovations in Bireeri and Kilii which are due to both being in contact with Oromo (rather than directly in contact with each other). There is no convention for including this type of information in a glottometric diagram yet.

One potentially misleading aspect of the glottometric diagram is that Kilii does not seem to be joined to the other dialects by any bundle of isoglosses at all; this is because the object of this paper is not to establish the innovations which characterize all the Boni dialects, but rather those which characterize subgroups within Boni. With a bit of searching, it should be easy to find many innovations exclusively shared by Boni as a unit in contrast to the other Sam languages, Somali and Rendille, and hence to establish support for such a bundle of isoglosses.

4. Conclusion

In this paper, I first reviewed the evidence which Heine presents in support of his classification of Boni dialects. Two of the three proposed subgroups did not reach the level of support required to establish a secure genetic subgroup, so I put forth an alternative scenario to Heine's. I propose that Boni only underwent one split (between Kilii and the rest of the dialects) followed by in situ differentiation of the other dialects. The latter situation is an example of non-tree-like language differentiation, which has left a distinct pattern of overlapping innovation in the Boni dialects and is not suitably represented by a tree diagram. In order to represent it, I chose a glottometric diagram, which has certain advantages over the tree diagram.

First, building the glottometric diagram does not require a strict differentiation of vertical and horizontal transmission, which is always fraught with difficulty in a closely knit dialect group like Boni. Second, it retains much of the information which is thrown out by a tree diagram. It permits certain observations to be made which would be otherwise impossible with a tree diagram. Third, it provides a much more natural and realistic account of the current observed pattern of innovations. The tree model forces one to make unnecessarily strong predictions about splits and contact, when the evidence may not justify such inferences.

There are a few areas in which further progress can be made. To an audience familiar with tree diagrams, a new method of visualization can seem difficult to interpret. However, the problems with the tree diagram, especially in dialect chains and linkages, are too numerous to ignore, and there is a need for a new method of visualization. A practical matter which needs to be addressed is how to make the information about subgroupiness and cohesiveness as salient as possible to the reader; the thickness and darkness conventions used in this paper could continue to be improved. Additionally, it is still quite problematic to construct a glottometric diagram where too much information is missing from the database. It is hoped that future iterations of HG will address some of these remaining issues.

Abbreviations

BAD	Baddey
BIR	Bireeri
CA	Central Aweer
CCA	Central Convergence Area
HG	Historical Glottometry
JAR	Jara
KIJ	Kijee
KIL	Kilii
LDAK	Language and Dialect Atlas of Kenya
PB	Proto-Boni
SA	South Aweer
SAF	Safaree
SCA	Southern Convergence Area
SCCA	South-Central Convergence Area

References

Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.6

Bryant, David & Vincent Moulton. 2003. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution* 21:2.255–265. https://doi.org/10.1093/molbev/msh018

Bryant, David, Flavia Filimon & Russel D. Gray. 2005. Untangling our Past: Languages, Trees, Splits and Networks. *The Evolution of Cultural Diversity: Phylogenetic Approaches* ed. by Ruth Mace, Clare J. Holden & Stephen Shennan, 67–83. London: University College London Press.

François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. *The Routledge Handbook of Historical Linguistics* ed. by Claire Bowern & Bethwyn Evans, 161–189. New York: Routledge.

François, Alexandre. 2017. Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues. *Diffusion: implantation, affinités, convergence* ed. by Jean-Léo Léonard, 43–82. (= *Mémoires de la Société de Linguistique de Paris*, XXIV.) Louvain: Peeters.

Google (n.d.). [Map of Kenyan coast]. Retrieved on November 20th, 2018. Available at https://www.google.com/maps

Heine, Bernd. 1977. Bemerkungen zur Boni-Sprache (Kenia). *Afrika und Übersee LX*, ed. by Ernst Damman, Ludwig Gerhardt, Ernst Hammerschmidt, Emmi Kähler-Meyer, Hilke Meyer-Bahlburg & Jürgen Zwernemann 242–295.

Heine, Bernd. 1978. The Sam Languages: A History of Rendille, Boni and Somali. *Afroasiatic Linguistics*, 61–93.

Heine, Bernd & Wilhelm J. G. Möhlig. 1982. Boni Dialect. *Language and Dialect Atlas of Kenya: Vol. X* ed. by Bernd Heine and Wilhelm J. G. Möhlig. Berlin: Dietrich Reimer.

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. *Let's Talk About Trees: Tackling Problems in Representing Phylogenic Relationships Among Languages* ed. by Ritsuko Kikusawa & Lawrence Reid, 59–89. Osaka: National Museum of Ethnology.

Meinhof, Carl. 1912. *Die Sprachen der Hamiten*. Hamburg: Abhanglungen des Hamburgischen Kolonialinstituts.

Ross, Malcolm. 1988. *Proto-Oceanic and the Austronesian Languages of Western Melanesia*. Canberra: Pacific Linguistics.

Sasse, Hans-Jürgen. 1979. Entlehnung morphophonemischer Regeln im Boni. *Sprache und Geschichte in Afrika* 1.93–108.

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

Simons, Gary F. & Charles D. Fennig. 2017. *Ethnologue: Languages of the World, Twentieth Edition*. Dallas, Texas: SIL International. Available at http://www.ethnologue.com

Stiles, Daniel. 1988. Historical Interrelationships of the Boni with Pastoral Peoples of Somalia and Kenya. *Kenya Past and Present* 20.38–45.

Address for correspondence

Alexander Elias
Leiden University
Van Wijkplaats 4
2311 BX Leiden
The Netherlands
a.elias@hum.leidenuniv.nl

Subgrouping the Sogeram languages

A critical appraisal of Historical Glottometry

Don Daniels[1], Danielle Barth[1] and Wolfgang Barth[2]
[1] Australian National University |
[2] Centre of Excellence for the Dynamics of Language

Historical Glottometry is a method, recently proposed by Kalyan and François (François 2014; Kalyan & François 2018), for analyzing and representing the relationships among sister languages in a language family. We present a glottometric analysis of the Sogeram language family of Papua New Guinea and, in the process, provide an evaluation of the method. We focus on three topics that we regard as problematic: how to handle the higher incidence of cross-cutting isoglosses in the Sogeram data; how best to handle lexical innovations; and what to do when the data do not allow the analyst to be sure whether a given language underwent a given innovation or not. For each topic we compare different ways of coding and calculating the data and suggest the best way forward. We conclude by proposing changes to the way glottometric data are coded and calculated and the way glottometric results are visualized. We also discuss how to incorporate Historical Glottometry into an effective historical-linguistic research workflow.

Keywords: Historical Glottometry, subgrouping, language diversification, comparative reconstruction, Sogeram languages, Madang languages, Papuan languages

1. Introduction

Historical linguists have wrestled for a long time with the question of how best to represent the relationships among sister languages in a language family. The two principal proposals, the family tree model (Schleicher 1853) and the wave model (Schmidt 1872), have been known to linguists for a century and a half now. While most historical linguists consider them idealizations that do not capture all essential facts about language change, there is still no widespread agreement on the best way to conceive of and model language relationships within a family.

https://doi.org/10.1075/jhl.17011.dan
Journal of Historical Linguistics 9:1 (2019), pp. 92–127. issn 2210-2116 | e-issn 2210-2124

A promising recent proposal for modeling language relationships is a method called Historical Glottometry (Kalyan & François 2018, François 2014). In this article we employ Historical Glottometry to investigate the Sogeram languages of Papua New Guinea, and in the process provide a critical appraisal of the method.

The article proceeds as follows. In the rest of this section we review the relevant literature on language diversification and subgrouping (Section 1.1) and then introduce the Sogeram languages (Section 1.2). We then present our dataset of innovations and discuss the method of Historical Glottometry in Section 2. In Section 3 we present our conclusions, focusing in particular on how they differ from Kalyan's and François's and on how we think the method should be revised. We propose a different way of representing the data that more accurately shows patterns of innovation when isoglosses frequently cross-cut each other (Section 3.1); we argue that lexical innovations exert undue influence on François's dataset and propose a more appropriate way of coding lexical change (Section 3.2); and we discuss how to handle cases in which the analyst is uncertain whether a given language underwent a change or not (Section 3.3). We also argue that coding decisions need to be made much more transparent, and for that reason we have made our glottometric dataset available online with this publication. In the conclusion, we compare Historical Glottometry to other models of language diversification, illustrating how it can represent a variety of language diversification scenarios, including some that other models struggle with. We also propose some changes to the method (Section 4.1) and discuss how it can be integrated with other methods of historical-linguistic inquiry (Section 4.2).

1.1 Models of linguistic relatedness and subgrouping

Language diversification is a complicated process, and the difficulty of accurately modeling it has long been recognized. Just representing the spread of language-internal innovations is challenging enough, but there are also factors like contact-induced change and convergence to take into account. Two primary models have long been recognized: the family tree model, dating at least to the work of August Schleicher (1853); and the wave model, proposed some twenty years later by Johannes Schmidt (1872). The family tree model, being perhaps the more legible and easily understood of the two, has enjoyed significantly more popularity: papers in historical linguistics often give a family tree but only rarely provide an alternative representation of relatedness. A few examples include Loughnane & Fedden (2011), Willis (2011), Epps (2013), Fedden et al. (2013), Daniels (2014), Sarvasy (2014), DeLancey (2015), Ross (2015), and Foley (2017); note that the occasional exceptions to the trend, such as Heuvel & Fedden (2014), Korn (2016), and Campbell (2017), generally focus specifically on questions of subgrouping. Some of the authors on this list, such as Malcolm Ross, are noted critics of the family tree model, which only serves to underscore the representational power of family trees for succinctly expressing diachronic relationships.

But in spite of their popularity, family tree diagrams often present an inaccurate view of the historical developments within a language family – they are, essentially, "a vast oversimplification" (Matisoff 2002: 292). This is well known, and family trees are often accompanied by caveats about their limitations. Accordingly, the search for alternative models has continued. Perhaps the most well-known of these is Ross's (1988) "linkage" model, which aims to represent the often diffuse dialect networks from which modern Austronesian languages descended. This model has been expanded on by other authors, such as Pelkey (2015), who argued that the pattern of diversification evident in the Phowa languages of China, a subgroup of Tibeto-Burman, requires both the tree model and the wave model for a full account. He called this integrated model "hinge diversification" and considered it a sub-type of Ross's linkage.

Ross himself later expanded on the linkage model and tied its representations more explicitly to certain kinds of social events that can take place within a speech community (1997). The difference between what he terms "linkage breaking" (the gradual breakup of a dialect network) and "language fissure" (a sudden split of one speech community into two) is not fundamentally one of process, but one of degree. In linkage breaking, a community splits apart slowly and significant social ties are maintained between the various daughter communities. In language fissure, on the other hand, there is a sudden split in the community and the number of social links between the two networks "undergoes a sharp reduction in density" (Ross 1997: 218). Thus the difference is essentially one of speed: language fissure is a faster break-up of a community, linkage breaking a slower one.

An important consequence of this observation is that linkage breaking and language fissure are not discrete kinds of events, but rather points on a continuum of various language-diversification processes. This point has been made, among others, by Pawley (1999), who proposed a three-way typology between perfect subgroups, imperfect subgroups, and linkages. The first and last of these are, more or less, the outcomes of Ross's (1997) language fissure and linkage breaking, respectively; Pawley simply added a middle point between them and called it an "imperfect subgroup." He did not consider these three kinds of linguistic grouping to be discrete entities, though, but rather "points on what is, in reality, a continuum" (Pawley 1999: 129).

If the various ways that languages diversify can be conceived of as points along a continuum, then it would be ideal to be able to capture the tree-like-ness or linkage-like-ness of various subgroups in a scalar way. A recent proposal by Siva Kalyan and Alexandre François (Kalyan & François 2018; François 2014) offered just such a method, which the authors call Historical Glottometry. The idea is, in short, to create a dataset of all the innovations that have been reconstructed for a given language family and then to perform summarizing statistics on that dataset to produce what they call a "subgroupiness" score for each subgroup in the family. This score represents the tree-like-ness of that subgroup in a numerical way so that it can be compared with other subgroups. The subgroupiness values for a whole family can then be combined into what is essentially a wave diagram, but one in which the lines do not represent individual isoglosses but rather subgroupiness scores.

Kalyan & François (2018) have exemplified this method with the Torres-Banks languages, a group of Austronesian varieties spoken in northern Vanuatu with which François has extensive experience (e.g. François 2005, 2011, 2013, *inter alia*). In this article we present a glottometric examination of a different language family: the Sogeram languages of Papua New Guinea. In the course of this examination, we propose several suggestions for improving the method.

1.2 The Sogeram languages

The Sogeram languages are a group of ten languages spoken along the Ramu and Sogeram Rivers in central Madang Province, Papua New Guinea (see Figure 2). They belong to the Madang branch of the large Trans New Guinea family (Z'graggen 1971, 1975b; Pawley 2005, 2006; Ross 2005; Pawley & Hammarström 2017), and within Madang to the South Adelbert subgroup (Z'graggen 1980; Daniels 2010a, 2015). Four of them (Mand, Manat, Magɨ, and Kursav) are spoken in only one village each, and another two (Aisi and Sirva) are spoken in five or less; they are each spoken by no more than a few hundred people (Z'graggen 1975a). Larger languages, such as Apalɨ or Nend, often have significant dialect variation (Wade 1993; Daniels 2010a). This pattern of small communities exhibiting high linguistic diversity has presumably been characteristic of Sogeram communities since the breakup of the family and suggests that language contact and borrowing were common. This social situation seems broadly similar to that found in the Torres-Banks communities (François 2012), but presumably there have been substantial differences as well. Further research will hopefully be able to uncover what significance these differences have had for language diversification.

The current subgrouping of the Sogeram languages is based on the reconstruction in Daniels (2015); this subgrouping has remained unchanged in recent work on Sogeram (e.g. Daniels 2017a: 585), as shown in Figure 1. Daniels has also sometimes used a network diagram, shown in Figure 3, to represent the pattern of innovation diffusion observed in the family (e.g. Daniels 2017b: 90). The claims made by these diagrams will be revised in Section 4.2, as a result of the glottometric examination of the Sogeram data.

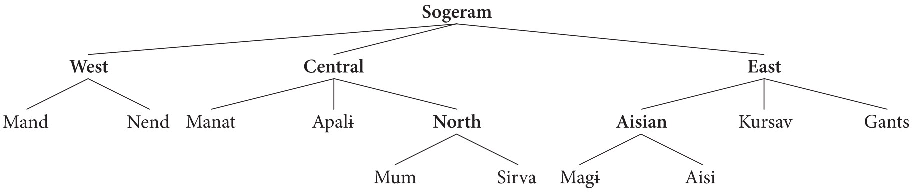


Figure 1. Sogeram subgrouping per Daniels (2017a)

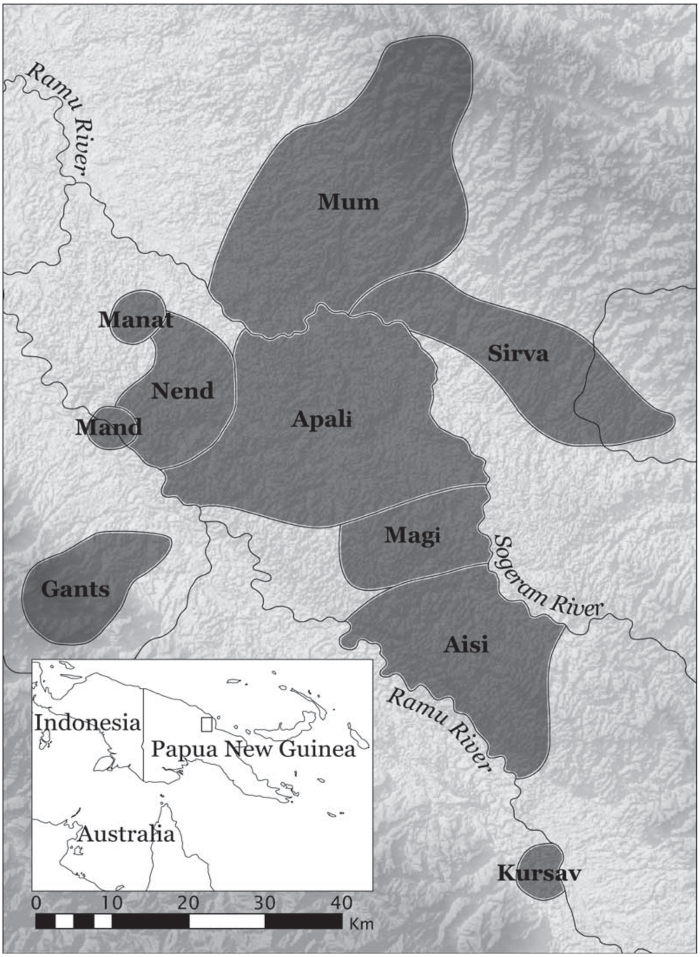


Figure 2. Map of the Sogeram languages

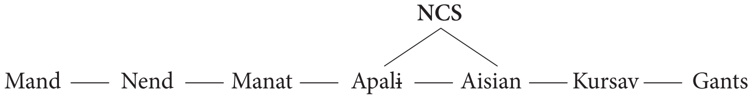


Figure 3. Sogeram contact chain per Daniels (2015)

2. The data and the method

In this section we discuss the data that we used in our examination of the Sogeram languages, as well as the method for calculating similarity between languages in Historical Glottometry.

2.1 The Sogeram innovations dataset

Daniels (2015) contains a reconstruction of Proto-Sogeram, with sections on phonology, morphology, syntax, and lexicon. We constructed a dataset based on this reconstruction, in which we entered every innovation that was shared by more than one Sogeram language. We counted a total of 196 such innovations; given that there are ten Sogeram languages, this means that the dataset consists of 1,960 individual data points. Each of these has one of three values: one (1) when the language reflects the innovation in question, zero (0) when it does not, and a blank cell (×) when it is not possible to be sure. Of the 1,960 data points, 578 (29.5%) are ones, 1,044 (53.3%) are zeros, and 338 (17.2%) are blanks.

A sample of the dataset is shown in Table 1. We considered ten different categories of change: regular sound change, irregular sound change, lexically specific sound change, lexically specific semantic change, lexically specific morphological change, creation of a new grammatical form, formal change to a bound morpheme, semantic change to a bound morpheme, loss of a bound morpheme, and syntactic change.

The category of regular sound change is fairly straightforward: any sound change that applied regularly throughout the whole lexicon in a language was considered regular. We are thus using "regular" to refer to sound changes that fulfill the Neogrammarian ideal, not to sound changes that are typical or expected. Initial plosive lenition, for example, applied to all word-initial plosives in Manat and Apalɨ. There are 22 regular sound changes in our dataset.

An irregular sound change is one that applied to some, but not all, lexical forms that it could have. For example, Mum and Sirva lost word-initial *i, but neither one did so from all words that had initial *i. It is naturally possible that any change that we classified as irregular was, in fact, regular and the conditioning environment simply has not yet been discovered, but there will be no way to address this issue until there has been considerably more research on the Sogeram

Table 1. Some Sogeram innovations

	Mnd	Nen	Mnt	Apa	Mum	Sir	Mag	Ais	Kur	Gaj
Initial plosive lenition	0	0	1	1	0	0	0	0	0	0
Loss of word-initial *i	0	0	0	0	1	1	0	0	0	0
Loss of *i in *iŋkwa 'give'	0	0	0	0	1	1	0	0	1	1
Creation of quotative marker	0	0	0	0	0	0	1	1	0	0
*kukra 'grow' > 'big'	1	1	0	0	0	×	0	0	0	0
Changed verb class for *imu 'cook'	1	1	0	×	×	×	×	×	0	×
New 1SG subject pronoun *nzɨ	0	1	1	×	0	0	0	0	0	0
Nasalization in *-ta 'SS.DELAY'	×	×	×	0	0	0	0	0	1	1
*-ɨt-ɨŋ '-IRR-1SG' > 'IMP'	0	0	1	1	1	0	0	0	0	0
Loss of *-u '2SG.IMP'	0	1	1	1	1	0	0	0	0	1
Postverbal negation construction	0	0	0	0	0	0	0	1	1	0
Loss of unaffixed demonstratives	1	1	1	0	0	0	0	0	0	0
NMLZ of *ura 'call out' > 'yelling'	0	0	0	0	0	1	×	1	0	0

language family and its phonological history. Our dataset contains 11 irregular sound changes.[1]

A lexically specific sound change is a change to the phonological properties of a single lexeme. This is obviously similar to – and might properly be conceived of as a subtype of – irregular sound change. For our purposes, we simply considered a sound change lexically specific if we observed it in only one lexeme and irregular if we observed it in more than one. Again, we face the same issue as with irregular sound change: it is possible that certain sound changes that only show up once in our dataset were actually fully regular, but that they affected rare phonological environments that only appeared in one form in our data. Similarly, we unfortunately see no way to address this possibility. Our data contain 38 lexically specific sound changes.

When the meaning of a particular lexeme changes in some languages, we classified that development as lexically specific semantic change. Thus *kukra 'grow' came to mean 'big' in Mand and Nend. Note that no reflex of this word has been

---

[1] There were also some changes that were regular in some languages and irregular in others. We counted these twice: a "regular sound change" counting only the languages in which the change was fully regular and an "irregular sound change" counting all languages that show the change, whether regularly or irregularly. Our data contain five such pairs.

found in Sirva, so we cannot know whether it underwent this semantic change or not. There are 13 instances of lexically specific semantic change in our data.

When we speak about lexical morphological change, we refer to a change in the morphosyntactic properties of an individual lexeme. So, for example, *imu ‘cook’ is reconstructed with a root-final *u, which affected the interaction of the root with certain verb suffixes. However, in Mand and Nend, this *u was replaced with an *a on analogy with *a-final verbs, which were the largest class of verb roots in Proto-Sogeram. Again, we lack a reflex of *imu in several languages, so we cannot know whether this word underwent this change in those languages or not. We record 21 instances of lexically specific morphological change, twelve of which take one of two forms: change to a verb’s verb class, as discussed for *imu ‘cook’; and change in the possessive prefix allomorph that an inalienably possessed noun takes.

We recorded 12 examples of grammaticalization – that is, the creation of a new grammatical form, defined for our purposes as bound affixes and pronouns. Thus the creation of a new 1SG pronoun *nzɨ in Nend and Manat is categorized as this sort of change. Note that Apalɨ also has an innovative 1SG pronoun *viaŋ* which is not cognate with *nzɨ. This means that we do not know whether Apalɨ participated with Nend and Manat in the creation of *nzɨ, since there is no way to tell whether *viaŋ* replaced *nzɨ or the original Proto-Sogeram 1SG pronoun *ya. Note that this category of change includes instances where we are sure of the etymology of the innovative form as well as instances, such as the creation of the Nend and the Manat pronoun, where we are not.

Formal change to a grammatical morpheme consisted of any irregular sound change that affected a bound morpheme. (For this category and the other two categories that involve grammatical morphemes, we considered pronouns as well as all bound morphemes “grammatical morphemes.”) For example, the alveolar stop in the ‘SAME-SUBJECT DELAYED’ suffix *-ta became a prenasalized *d in Kursav and Gants, when *d is not a reflex of *t in any other word or affix. We record 32 examples of this kind of change.

Semantic change to a grammatical morpheme involves some shift in the semantic or grammatical function of a grammatical morpheme but not in its phonological shape. This kind of change is fairly similar to grammaticalization, except that here both the start and end points are “grammatical morphemes” by our definition. For example, the two-suffix combination *-ɨt-ɨŋ ‘-IRR-1SG’ came to have imperative meaning instead of its original, more general irrealis meaning in Manat, Apalɨ, and Mum. There are 13 changes of this kind in our data.

We also considered the loss of a grammatical morpheme (again, defined for us as a bound form or a pronoun) as an innovation. We did not count the loss of a lexeme as an innovation because our lexical knowledge of the Sogeram languages is not complete enough to be able to distinguish between the loss of a lexeme and semantic shift in a lexeme. For example, the word *ina 'sun' is reconstructed on the basis of reflexes in Mand, Mum, and Sirva. This does not mean, however, that we can group the remaining Sogeram languages together on the basis of their all having lost this word. Many may still retain it, only with a meaning that does not occur in our data. However, with grammatical morphemes, which are generally much more high-frequency items, we considered the absence of a reconstructed form to constitute an innovation. Thus, for example, the 2SG.IMP suffix *-u was lost from Nend, Manat, Apalɨ, Mum, and Gants. We recorded 24 changes of this type.

The last major category we considered was syntactic change. This is change to any grammatical construction that we can reconstruct. For example, Daniels (2015: 325–330) reconstructs a verbal negation construction *[*ma* V], but Aisi and Kursav both have innovative constructions [V *ma*]. In addition to this kind of word order change, syntactic change can consist of a change in the distributional properties of a grammatical morpheme. For example, demonstratives can appear with or without suffixes in Proto-Sogeram (Daniels 2015: 277), but no longer can do so in Mand, Nend, or Manat. We counted seven cases of syntactic change.

Finally, we recorded three cases of lexicalization. These are cases where a combination of reconstructed morphemes – whether grammatical or lexical – comes to have a particular lexical meaning in the innovative languages. For example, Proto-Sogeram had a reduplicative derivational suffix that attached to verbs to form nouns. In Sirva and Aisi this suffix is no longer productive, but the nominalization of *ura 'call out', reconstructed as *ur~ura, is retained as a noun meaning 'yelling, shouting'. The Magɨ data are so meager that we gave Magɨ a blank for this change.

We thus arrive at the total of 196 separate changes, as summarized in Table 2. Because the presence or absence of a blank cell in a row has important implications for the computation of linguistic affiliation described below, we also note how many rows with and without blanks there are in each category.

2.2 Calculating relatedness: The method of Historical Glottometry

We now briefly describe how relatedness is calculated in Historical Glottometry, although we refer readers interested in a more detailed discussion to Kalyan and François's work (Kalyan & François 2018, François 2014). Historical Glottometry involves the calculation of three values, referred to as p, q, and ε. The value p represents the total number of supporting innovations; for any group of languages ABC, p is the total number of innovations that are present in all three languages, A, B, and C. They may, but do not have to, include other languages as well.

Table 2. The Sogeram innovations dataset

	With blanks	Without blanks	Totals
Regular parts of irregular sound change	0	5	5
Irregular sound change	0	11	11
Regular sound change	0	17	17
Lexically specific semantic change	12	1	13
Grammaticalization	4	8	12
Lexically specific sound change	35	3	38
Lexically specific morphological change	20	1	21
Formal change to a bound morpheme	20	12	32
Semantic change to a bound morpheme	5	8	13
Loss of a bound morpheme	1	23	24
Change in a syntactic construction	1	6	7
Lexicalization	3	0	3
Totals	**101**	**95**	**196**

The value q represents the number of contradictory innovations. So for a group ABC, an innovation counts towards q if it includes at least one of the languages A, B, or C; excludes at least one of the languages A, B, or C; and also includes a non-ABC language. In other words, any innovation that draws an isogloss that cuts across the ABC isogloss increases the q value for that group of languages.

The values of p and q are combined into another value, κ, referred to as the group's "cohesiveness," in the following way:

$$\kappa_{ABC} = \frac{p}{p+q}$$

In other words, the cohesiveness of ABC is the ratio of supporting innovations p to the total number of relevant innovations $p + q$ (that is, innovations that involve A, B or C in any way). In other words, when one of the ABC languages undergoes an innovation, the κ value indicates what proportion of the time the others undergo it as well.

The cohesiveness of the group is then multiplied by ε. This value is simply the number of *exclusively* shared innovations – innovations shared by A, B, and C and by no other language. This means that ε is always a subset of p. This final step is designed to weight groupings that share many innovations more heavily. Without this final step, the method would always make smaller groups seem stronger than larger groups simply due to the fact that all innovations to larger groups also include smaller groups within them. The result of $\kappa \cdot \varepsilon$ is referred to as a group's "subgroupiness" and is labelled ς. To summarize:

$$\varsigma_{ABC} = \varepsilon \cdot \frac{p}{p+q}$$

This value is taken to represent the genealogical strength of a given subgroup: the proportion of innovations shared by a group of languages, multiplied by the number of innovations shared exclusively by that group.

A very important issue in calculating these values is what to do with blank cells in our data. We followed the following principles in tallying the p, q, and ε values. An innovation did not count for the p value of a group ABC if that row contained a blank cell for any of A, B, or C; innovations only counted for p if they contained a 1 for all three languages. Innovations only contributed to the q value of a group ABC if they met all three of these criteria: (i) either A, B, or C has a 1; (ii) either A, B, or C has a 0; and (iii) at least one non-ABC language has a 1. Finally, the calculation of ε depends on the entire row lacking blanks, because ε is defined as the occurrence of the innovation in ABC *and* the non-occurrence of it outside of ABC. This final requirement has far-reaching implications, which we discuss below.

3. The glottometric view of the Sogeram languages

Because there have not yet been many examinations of language families using the methods of Historical Glottometry, it will be instructive to point out the ways in which the Sogeram data differ from the Torres-Banks data analyzed by Kalyan & François (François 2014; Kalyan & François 2018), especially to point out how the methodology of computing subgroupiness and visualizing glottometric results might need to be adjusted. In this discussion we focus on three primary differences between our Sogeram data and Kalyan & François's Torres-Banks data: (i) the apparently higher rates of cross-cutting in the Sogeram innovations; (ii) our comparative lack of emphasis on lexical innovations in constructing our dataset; and (iii) the larger number of blanks in the Sogeram data.

Before beginning this discussion, however, we present the results of our glottometric analysis in Table 4, which shows every subgroup for which one of the following is true: either ε is 2 or greater or ς is greater than or equal to 0.20 (these are essentially arbitrary cut-off points so that the number of displayed groupings is manageable). There are 18 groups that meet one or both of these criteria. The language abbreviations we use are given in Table 3.

Table 3. Language names and abbreviations

Language	ISO code	Our abbreviation
Mand	[ate]	Mnd
Nend	[anh]	Nen
Manat	[pmr]	Mnt
Apalɨ	[ena]	Apa
Mum	[kqa]	Mum
Sirva	[sbq]	Sir
Magɨ	N/A	Mag
Aisi	[mmq]	Ais
Kursav	[faj]	Kur
Gants	[gao]	Gaj

Table 4. Glottometric figures for the Sogeram languages

Group	ε	κ	ς
Mnd-Nen	9	0.76	6.85
Mag-Ais	8	0.83	6.60
Mum-Sir	6	0.58	3.51
Mnd-Nen-Mnt	6	0.28	1.66
Kur-Gaj	4	0.46	1.85
Ais-Kur	2	0.33	0.66
Mag-Ais-Kur	2	0.32	0.64
Sir-Gaj	2	0.25	0.50
Mnd-Nen-Apa	2	0.22	0.44
Apa-Mag-Ais	2	0.18	0.36
Mnd-Nen-Mnt-Apa	2	0.11	0.22
Nen-Mnt	1	0.32	0.32
Apa-Mum	1	0.30	0.30
Sir-Kur	1	0.28	0.28
Nen-Apa	1	0.28	0.28
Mnd-Mnt	1	0.28	0.28
Mnt-Apa	1	0.27	0.27
Sir-Mag-Ais	1	0.20	0.20

3.1 Cross-cutting isoglosses

These results can be represented on a glottometric diagram. François (2014), in his representation of the Torres-Banks languages, only shows those groupings which have a subgroupiness ς of one or greater. If we follow this principle, our results look as shown in Figure 4.

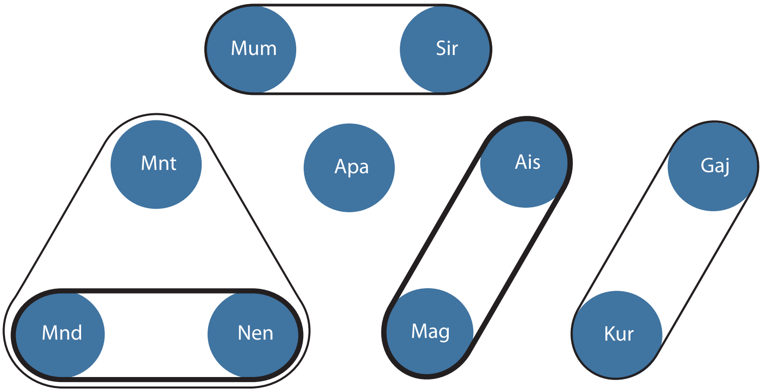


Figure 4. Glottometric diagram showing all isoglosses for which $\varsigma \geq 1$

Each circle represents a language. The position of the languages in the diagram is not significant, but corresponds roughly to geography. The thickness of a line around a group of languages corresponds to the subgroupiness value of that group.

Historical Glottometry was developed as a way of representing genealogical relationships between languages that allows innovations to cut across one another. Still, the frequency with which innovations do cut across one another can be expected to differ from family to family. Some families will exhibit a more tree-like structure (that is, they will have fewer cross-cutting isoglosses), while others will exhibit a less tree-like structure. The Sogeram family, at first glance, appears to have quite a tree-like structure – that is, there are no cross-cutting isoglosses in Figure 4, and sister languages do not appear to have influenced each other after diversification. But we believe this appearance is due, paradoxically, to the highly non-tree-like structure of the Sogeram family, and we propose that glottometric results should be represented differently from the way in which Kalyan & François (2018) have represented theirs.

The reason we believe the Sogeram family not to be tree-like has to do with Apalɨ. Figure 4 makes it seem as though Apalɨ, in spite of its very central geographic position in the Sogeram family, does not subgroup with any other Sogeram language. This would ordinarily be interpreted as meaning either that it diverged from the other Sogeram languages very early in the history of the family and is now a first-order subgroup, or that it is a very conservative language that simply has not innovated much at all.

But in fact Apalɨ is not particularly conservative; it undergoes 53 innovations in our dataset, which means that it has undergone innovations with other Sogeram languages at a fairly average rate. Figure 5 shows the number of innovations for each language that our dataset contains. The family average is 57.8 innovations, and the most conservative languages are Gants and Manat, with 42 and 45 innovations, respectively. The languages with above-average numbers of innovations in our dataset are those with close relatives: the Aisian languages Magɨ and Aisi and the West Sogeram languages Mand and Nend. This is to be expected, because languages with a close relative in the family underwent longer periods of common innovation with that relative before breaking up relatively recently. Languages without a close living relative, on the other hand, would be expected to reflect fewer innovations in common with other Sogeram languages, because they did not undergo as long a period of shared innovation before they split up from their closest relative in the family.

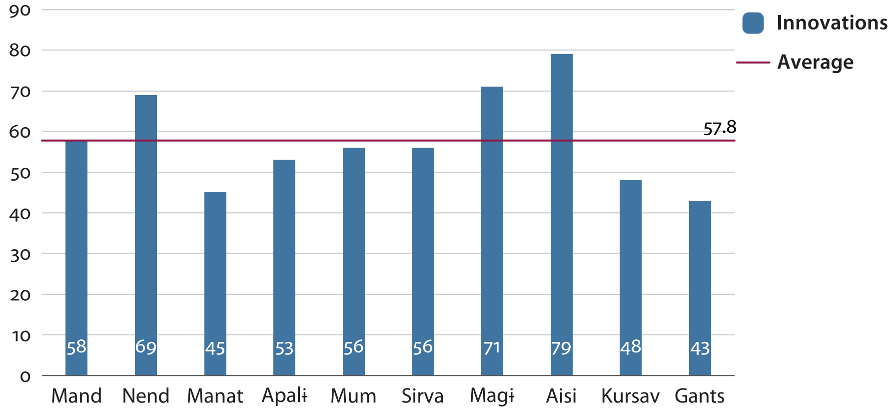


Figure 5. Number of innovations per language

We might expect the fact that Apalɨ commonly shares innovations with other languages to suggest that it ought to subgroup with something. However, an examination of the data reveals that Apalɨ is, in fact, grouping with too many other languages: it shares numerous innovations with Manat, Nend, Mum, Magɨ, and Aisi. This means that each of the groups that includes Apalɨ is contradicted by all of the other groups that include it. Due to how subgroupiness is calculated in Historical Glottometry, that means that none of these subgroups can attain a subgroupiness of 1.

If, instead, we define our cutoff in terms of exclusively shared innovations ε, then we begin to see the history of the Sogeram family more clearly. Figure 6 shows all the subgroups for which ε is 2 or greater, and here we see Apalɨ grouping with the languages to the west – Mand, Nend, and Manat – as well as to the east – Aisi and Magɨ. We also see many more cross-cutting isoglosses in the eastern part of the family: Kursav and Gants still group together, but Kursav also groups with both Aisian languages (Magɨ and Aisi). In Section 3.3 below, we discuss other possibilities for visualizing the data, suggesting that Apalɨ also groups with Mum and Sirva to the north.

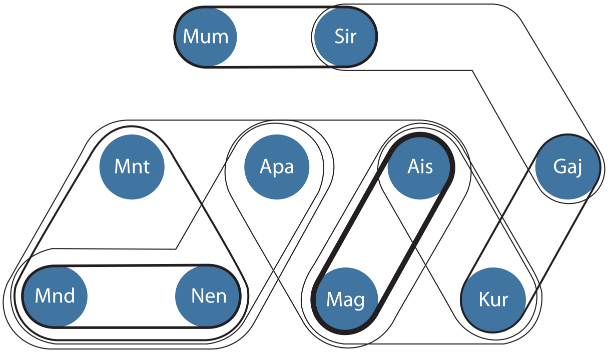


Figure 6. Glottometric diagram showing all isoglosses for which $\varepsilon \geq 2$

While Figure 6 is an improvement on Figure 4, it is still not perfect. Defining our cutoff in terms of ε runs a different risk: namely, the risk of over-representing groups of languages that simply happen to have shared several parallel innovations. This is what we believe has happened with the Sirva-Gants subgroup in this diagram. Sirva and Gants have a relatively low cohesiveness κ of 0.25, but that figure is bolstered by two innovations which they both underwent, both of which we actually consider to be best analyzed as parallel developments rather than shared ones (recall also from the map in Figure 2 that Sirva and Gants are spoken on opposite sides of the Sogeram language area). We were deliberately conservative in our coding, which meant that if two languages underwent similar developments, even if we were reasonably sure that they were parallel innovations rather than shared innovations, we coded them as shared innovations to avoid inserting our biases into the dataset. The two innovations from which the Sirva-Gants subgroup draws most of its strength are the loss of the participial suffix *-m and the reanalysis of the possessive pronouns as object pronouns.

The loss of participial *-m, in particular, is much more reasonably explained as parallel innovation than as a shared one, since the loss of inflectional categories is common cross-linguistically and Sirva and Gants are geographically very distant from one another. If we remove just this item from the dataset, the diagram changes to that shown in Figure 7. The strength of the Mum-Sirva and Kursav-Gants subgroups is strengthened slightly, because they are no longer contradicted by a Sirva-Gants isogloss, and the Sirva-Gants subgroup falls below the threshold we have set.

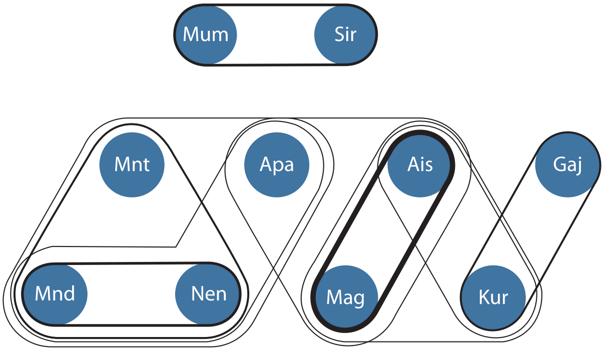


Figure 7. Glottometric diagram with one Sirva-Gants isogloss removed ($\varepsilon \geq 2$)

This point demonstrates that, although part of the value of Historical Glottometry comes from its ability to provide a quantitative perspective on language relationships, it is still necessary to engage qualitatively with the data and to be aware of which innovations are driving the patterns that the glottometric calculations present.

We turn now to a discussion of lexical replacement, which is perhaps the most significant way in which our dataset differs from François's (2014) dataset.

3.2 Lexical replacement

François's database of 474 Torres-Banks innovations (François 2014: 177) is summarized in Table 5, side by side with the comparable figures for our Sogeram dataset. François placed his changes into five broad categories, which are given in the left-hand column of Table 5. We place our changes into a larger set of more narrow categories, which are described in Section 2.1 above. For comparison with François's figures, we combined our categories into five groups to match the five categories he used. We did this in the following way: his "regular sound change" is our "regular sound change" plus "regular parts of irregular sound change." His "irregular sound change" is our "irregular sound change" plus "lexically specific sound change." His "morphological change" is our "lexically specific morphological change," "formal change to a bound morpheme," "semantic change to a bound morpheme," and "loss of a bound morpheme." His "syntactic change" is our "change in a syntactic construction" plus "grammaticalization." And his "lexical replacement" is our "lexically specific semantic change" plus "lexicalization." By combining our categories into groups to roughly match his, we can get a sense for how the dataset we constructed differs from his.

Table 5. A comparison of Torres-Banks and Sogeram innovations

	T-B	T-B%	Sogeram	Sogeram%
Regular sound change	21	4%	22	11%
Irregular sound change	116	24%	49	25%
Morphological change	91	19%	90	46%
Syntactic change	10	2%	19	10%
Lexical replacement	236	50%	16	8%
Totals	**474**	**99%**	**196**	**100%**

Two salient differences between our data and François's immediately present themselves: his data contain almost two and a half times as much irregular sound change as ours do, and they also contain over 14 times as many cases of lexical replacement as ours. In fact, lexical replacement forms approximately half of François's dataset, while it forms only eight percent of ours. We discuss these two differences and their implications below.

A third difference between our data and François's is the increased incidence of syntactic change in our data. We take this to be due to two things. First, Daniels's reconstruction focuses on syntactic change and questions of grammatical reconstruction (cf., e.g., Daniels 2014, 2015, 2017a), which has probably resulted in more cases of grammatical change being reconstructed and therefore more grammatical innovations being discovered. Second, François categorizes some instances of what we would call grammaticalization as morphological change instead (such as the creation of a dative preposition from a verb meaning 'follow'; François 2014: 178). These two factors increase the amount of syntactic change we report and decrease the amount François reports. However, the figures involved are rather small and we do not consider this a significant difference between the two datasets.

The most significant difference between our data and François's is the proportion of "lexical replacement" changes. François gives two examples of this kind of change: Proto-Oceanic *panako 'steal' was replaced with a new verb *^{m}balu in the Banks languages, and Proto-Oceanic *matiruR 'sleep' was replaced with *ŋoRo (etymologically 'snore') in five southern languages.

We consider this way of coding innovations problematic for a few reasons. First of all, the notion of "lexical replacement" implies a model of the lexicon in which a language will always have only one word for a given meaning, but this is obviously not the case. Languages tolerate quite a bit of synonymy.

More importantly, an examination of the notion of lexical replacement as François uses it reveals that it actually involves two separate changes – the innovation of a new word and the removal of an old one. Each of these is quite difficult to establish satisfactorily, especially given the often limited knowledge we have of the lexica of endangered minority languages. In order to argue, for example, that *panako 'steal' was replaced with *$^{\mathrm{m}}$balu in a subgroup ABC of a language family A–F, we have to be able to say two things. First, we must be able to assert that each of the languages A, B, and C lost the word *panako; in order to say this, we need to have extensive lexical knowledge of these languages, because it can easily be the case that these languages in fact do retain a reflex of *panako – it is simply one we have not heard yet. Second, we need to be able to say that none of the languages D, E, or F have a reflex of *$^{\mathrm{m}}$balu, which requires quite a similar level of familiarity with those languages. In other words, we need to have more or less complete lexical knowledge of every language in the family A–F. This level of knowledge does exist for some of the better-studied language families in the world, but we certainly do not believe it exists for the Sogeram languages, so we did not code innovations of this type.

One change that we did feel comfortable coding was semantic innovation, but even here there are issues. How much does the meaning of a lexical item have to shift for it to be considered an innovation? For example, Proto-Sogeram *mɨŋa 'get, hold, take' is retained in Mand as *ga-* 'grab'. It still occurs in many contexts that take a semantically neutral verb of holding, but its lexical meaning has become more restricted and it has been replaced as the most basic verb of holding or taking by *pi-* 'take'. This is a difficult case; luckily, we did not have to resolve it, because Mand is the only language to show this change. But the point is that semantic change is a matter of degree, and it is not clear how different a new meaning should be in order to count as an innovation. We were, consequently, quite conservative in our tally, counting only obvious cases of semantic change such as *kupra 'jump' > 'run' (in Aisi and Kursav) or 'good' > 'big' (in Manat and Mum).

Given this perspective, we can return to the two example innovations from François (2014) discussed above and see how we would have treated them differently. We would have conceptualized the replacement of Proto-Oceanic *matiruR 'sleep' with *ŋoRo (etymologically 'snore') as a semantic innovation that affected *ŋoRo 'snore' rather than a replacement that affected *matiruR 'sleep' (or that affected the semantic field 'sleep'). Additionally, we would not have included the innovation of a new word for 'steal' in our dataset at all, because we would not feel that our lexical knowledge of the Sogeram languages justified it.

We move now to a discussion of irregular sound change, which is another area in which our data differ significantly from François's. François refers to this category as "irregular (i.e. lexically-specific) sound change" (2014: 177), suggesting that he does not distinguish between a sound change that operates sporadically throughout the lexicon, affecting many lexemes but falling short of the Neogrammarian ideal, and a sound change that idiosyncratically affects one word. In fact, he treats all sound change that is not perfectly Neogrammarian as the latter type (Alexandre François, personal communication). This has the effect of magnifying the effect of irregular sound change relative to regular sound change, because each regular sound change is counted only once, while each irregular sound change is counted as many times as there are lexemes that instantiate it. This strikes us as an undesirable outcome. An irregular sound change that affects five lexemes in a set of languages is certainly evidence that those languages share a common heritage. But is that evidence five times stronger than one regular sound change shared by the same languages? And if further fieldwork reveals that the irregular sound change actually affected twenty lexemes, did the evidence just quadruple in strength again? In François's methodology, an irregular sound change that affected a grouping that wasn't otherwise well-supported by the data might appear with many lines of support in a glottometric dataset that counted every lexeme that reflected that change separately.

Rather, we believe irregular sound change should be weighted the same as regular sound change. Both are evidence that a particular way of speaking – a particular articulatory gesture or pattern of articulatory gestures – came to be preferred in a speech community. If evidence of that preference is retained in a set of modern languages, that suggests that, at some point in the past, they constituted a single community of practice. This is why we consider irregular sound change that is shared by a set of languages as evidence of common development for those languages *even if* the sound change is not found in the same lexemes. For example, Mum and Sirva both irregularly lost word-initial *i, but not always in the same words (Daniels 2015: 92). Sometimes both lost it (*iŋka 'see, perceive' > Mum, Sirva *ga-*), sometimes only Mum did (*isa 'bite' > Mum *sa-*, Sirva *isa-*), and sometimes only Sirva did (*intar- 'hear' > Mum *idar-*, Sirva *dari-*). Daniels interprets this distribution of the sound change as evidence that it took place when Mum and Sirva had already diverged into "a rather diffuse dialect network" (Daniels 2015: 90), in contrast to changes shared regularly by both languages, which probably took place when they were a more tightly integrated speech community. Whether that is the case or not, it seems clear to us that a shared irregular sound change is not stronger evidence of relatedness than a shared regular sound change, so we count them both equally.

We now turn to a discussion of the number of blank cells in our glottometric datasets.

3.3 Blank cells in glottometric datasets

Another way that our data differ significantly from François's (2014) is in the number of blank cells we entered. Our dataset has 1,960 data points in all, of which 338, or 17%, are blank. François's dataset, on the other hand, contains 290 blanks out of 8,058 data points – less than 4%. Part of this may be due to differing levels of familiarity with the language families we are dealing with: François began his work on the Torres-Banks languages in 1997, while Daniels did not begin his graduate work on the Sogeram languages until 2010.

Another factor, however, may be a difference in approach to coding glottometric datasets. We did not want our preconceptions to influence how we coded areas of uncertainty, so we were deliberately conservative in our coding. If there was any doubt regarding the participation of a language in an innovation, the cell was left blank. For example, Table 6 gives two examples of rows where we left cells blank. The first line is a case of vowel metathesis that affected the word for 'cross-cousin' in the two Aisian languages, Magɨ and Aisi. Kursav does not have a reflex of *-mintaŋ 'cross-cousin', so we left it blank. Because Kursav shares several innovations with the two Aisian languages, it is plausible to suppose that it underwent this metathesis with the Aisian languages and only later lost this word. Mand and Nend, the two westernmost languages, also lack a reflex of *-mintaŋ. But, in contrast to Kursav, they show virtually no evidence of sharing any innovations with the Aisian languages whatsoever. That means it would be fairly safe to assume that Mand and Nend did not undergo this metathesis with the Aisian languages. Nevertheless, we decided to leave the cells blank since we lack positive evidence to say that Mand and Nend did not undergo this change. A similar case is the nasalization of the same-subject delayed suffix *-ta, which became *-da in Kursav and Gants. The three languages at the other end of the Sogeram contact chain (Mand, Nend, and Manat) do not have a reflex of this morpheme. While it is almost certain that they did not undergo this change with Kursav and Gants, due to the fact that it is not completely certain, we leave these cells blank.

Table 6. Examples of uncertain changes

	Mnd	Nen	Mnt	Apa	Mum	Sir	Mag	Ais	Kur	Gaj
*-mintaŋ 'cousin' > *-mari	×	×	0	0	0	0	1	1	×	0
Nasalization in *-ta 'SS.DELAY'	×	×	×	0	0	0	0	0	1	1

What effect has our approach to coding had on our results? One effect is to lower the rate of exclusively shared innovations ε in our data. Because ε requires positive knowledge that every language in a subgroup underwent the innovation at hand *and* that every language outside the subgroup did not, any row which contains a blank cell cannot count towards any ε value. For example, the nasalization of *-ta was almost certainly shared exclusively by Kursav and Gants, but it does not count as an exclusively shared innovation in our glottometric calculations because we acknowledge the remote possibility that Mand, Nend, or Manat also underwent the change. If we entered zeros in these Mand, Nend and Manat cells, the Kursav-Gants subgroup would gain another exclusively shared innovation and would appear stronger. We acknowledge that this is a potentially undesirable result, since it deflates the number of exclusively shared innovations in our dataset. However, it is very difficult to draw a principled line between cases like the nasalization of *-ta, in which it is all but certain that the three western languages did not undergo the innovation, and cases like the metathesis in *-mintaŋ, where we simply cannot know whether Kursav underwent the change or not.

This conservatism in our coding means that over half the innovations that we coded contain at least one blank cell: of the 196 innovations we coded, 101 contain blank cells and only 95 do not.

It is also important to note that the uncertainty is not distributed evenly across the different kinds of change. Table 2 shows the categories of change in which these blank cells are contained. Sound changes and changes affecting bound morphemes are most likely to have every cell filled, while lexically specific changes are most likely to contain blank cells. The reasons for this are fairly obvious: if a sound change affected a particular phonological environment, it is quite likely that that phonological environment will be found in at least some lexemes in every daughter language; consequently, we will be able to say whether each daughter language underwent that change or not. Similarly, bound morphemes are relatively stable items diachronically, and so there are relatively high odds that a reflex of a bound morpheme will be kept in most daughter languages. Individual lexical items, however, are lost more easily, and when a language lacks a reflex of a particular word, we cannot usually tell whether that word underwent a particular semantic or phonological change before being lost.[2]

To see whether the higher rate of blank cells in our data had a potentially warping effect on our results, we ran the glottometric calculations on the subset of our data that does not contain any blanks. The results are given in Table 7 and Figure 8, and they comport relatively well with our overall results. Table 7 shows

---

[2] It is also, of course, possible that the language which has a blank cell has actually retained the lexeme at issue, but that it simply has not yet been encountered in fieldwork.

all groups for which ε is two or greater or ς is 0.30 or greater, while Figure 8 shows all isoglosses for which ε is two or greater.

Table 7. Glottometric figures for rows with no blanks

Group	ε	κ	ς
Mnd-Nen	9	0.66	5.93
Mag-Ais	8	0.75	6.00
Mum-Sir	6	0.57	3.43
Mnd-Nen-Mnt	6	0.30	1.80
Kur-Gaj	4	0.41	1.66
Ais-Kur	2	0.38	0.76
Mag-Ais-Kur	2	0.35	0.70
Sir-Gaj	2	0.26	0.52
Apa-Mag-Ais	2	0.19	0.38
Mnd-Nen-Apa	2	0.18	0.36
Mnd-Nen-Mnt-Apa	2	0.11	0.21
Sir-Kur	1	0.32	0.32
Nen-Mnt	1	0.31	0.31
Mnd-Mnt	1	0.31	0.31
Apa-Mum	1	0.30	0.30

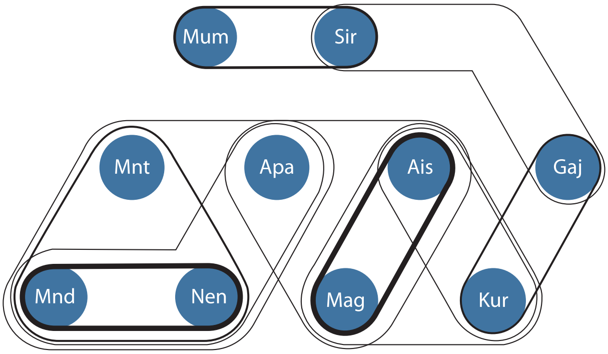


Figure 8. Glottometric diagram without rows containing blank cells (ε ≥ 2)

It is noteworthy that even though this diagram is based on less than half our data – the 95 rows that contain no blank cells – it matches the diagram in Figure 6 almost perfectly. Mand-Nend, Magɨ-Aisi, and Mum-Sirva are singled out as the strongest groups by far, with Mand-Nend-Manat and Kursav-Gants also coming out with fairly strong support. Weaker support is then shown for Aisi-Kursav, Magɨ-Aisi-Kursav, Sirva-Gants, Apalɨ-Magɨ-Aisi, Mand-Nend-Apalɨ, and a few other groupings. This suggests, hearteningly, that the groups that emerge from the rows that do not contain blank cells are not radically different from the groups that emerge from the dataset as a whole, and that our conservative way of coding the data has therefore not unnecessarily skewed our results.

Another coding possibility is to treat every blank cell as a zero – in other words, to assume that a language did not undergo any innovations unless there is positive evidence to the contrary. This compensates for the problem discussed in connection with Table 6 above, namely that we were quite conservative in our coding and that this lowered the ε values for several groups. However, this approach not only compensates for that problem, it overcompensates for it: now, innovations for which there is legitimate uncertainty are also treated as though they were definitively one way or the other in the computations. Nevertheless, this approach offers an instructive perspective on the data.

As mentioned, the approach raises the subgroupiness values even further, as the elimination of blank cells means that, now, every row counts towards the ε value of some subgroup. Running the calculations in this way gives the results shown in Table 8 and Figure 9.

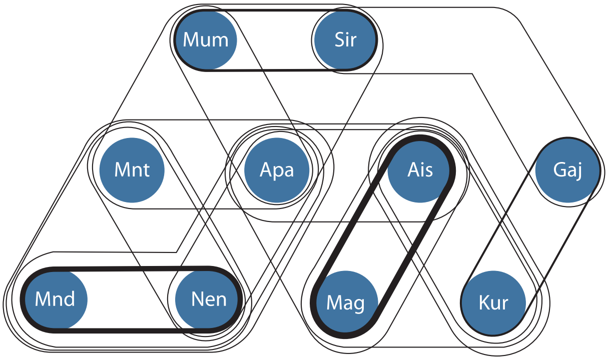


Figure 9. Glottometric diagram treating blank cells as zeros (ε ≥ 3)

The higher ε values in the table prompted us to use a higher cutoff in our diagram – we only drew isoglosses for which ε was 3 or greater – but the diagram is still busier than previous ones. It contains 16 isoglosses, compared with 11 in both Figure 6 and Figure 8. It also, interestingly, contains several isoglosses that here are quite strong yet which do not emerge strongly in other views of the data. The top five groupings remain unchanged, each with subgroupiness values over 1.5. Magi-Aisi-Kursav comes next, but then the seventh- and eighth-strongest groups are Apalɨ-Mum and Nend-Manat, neither of which is strongly supported when blanks are either left as blanks or are eliminated. Both of these groups are plausible – they each share four exclusive innovations, after all – but both are also often contra-

Table 8. Glottometric figures treating blanks as zeros

Group	ε	κ	ς
Mag-Ais	28	0.79	22.00
Mnd-Nen	25	0.67	16.78
Mum-Sir	12	0.51	6.16
Mnd-Nen-Mnt	7	0.25	1.75
Kur-Gaj	6	0.38	2.27
Mnd-Nen-Mnt-Apa	5	0.10	0.52
Mag-Ais-Kur	4	0.28	1.14
Apa-Mum	4	0.27	1.07
Nen-Mnt	4	0.27	1.07
Mnd-Nen-Apa	4	0.20	0.79
Ais-Kur	3	0.26	0.77
Nen-Apa	3	0.24	0.73
Sir-Gaj	3	0.21	0.62
Apa-Mag-Ais	3	0.16	0.47
Apa-Ais	3	0.15	0.44
Mnt-Apa-Mum-Sir	3	0.07	0.20
Mnt-Apa	2	0.24	0.48
Sir-Mag-Ais	2	0.19	0.39
Sir-Ais	2	0.17	0.35

dicted: Apalɨ-Mum is contradicted by the strong Mum-Sirva group and Manat-Nend by the even stronger Mand-Nend group. This frequent contradiction by very strong isoglosses probably explains their non-appearance in previous views of the data, and the fact that treating blank cells as zeros has raised the ε value of each of these groups probably explains their appearance now. Another group that appears in this diagram is Manat-Apalɨ-Mum-Sirva, which corresponds to Daniels's (2010a, 2015) Central Sogeram subgroup and which has not appeared on any diagram so far.

A final possibility is to fill in blank cells probabilistically, based on the distribution of ones and zeros in other parts of the dataset. When determining how likely a blank is to be a one or a zero, we can look at other rows in the dataset that have filled distributions in the same context. For these calculations we employed the following methodology. For every blank cell, we found all the 1s in its row (i.e., all the languages that shared that particular innovation). Then we found all other rows that contained 1s in the same cells (i.e., all the other innovations shared by those languages) and calculated the percentage of the time that the language in question had a 1 (i.e., the percentage of the time the language with the unknown innovation actually shared innovations with those languages). We repeated this process for every 0 in the original row (we calculated how often every language known to lack the innovation in question *did* innovate with the languages that had the innovation in question). This gave the probability of having the innovation for each language that actually lacked it. We found the highest probability out of the languages that lacked the innovation and used that as a baseline. Then we calculated the probability of the innovation being present in the blank cell on a scale from that calculated baseline to 100% (i.e., how much more likely it was that the language in question shared the innovation than any language with a 0 in the original row did).

We generated these probabilistic numbers for every blank in the dataset, then ran the glottometric calculations over many sample datasets; each time, we filled each blank cell with either a 1 or a 0, based on its probability value. The glottometric measures from all the samples could then be averaged.

This is a very promising possibility, but one that currently requires more development. A preliminary diagram based on such a calculation and showing all isoglosses for which $\varepsilon > 2.5$ is given in Figure 10.[3] The tabulated results are given in Table 9. The similarity to Figure 9 is striking, particularly in the appearance of the three groups discussed above: Nend-Manat, Apalɨ-Mum, and Manat-Apalɨ-Mum-Sirva. It is also noteworthy that the Mag-Ais-Kur-Gaj group, which is not strongly supported in other views of the data, is relatively well-supported here. This is Daniels's East Sogeram subgroup (Daniels 2010a, 2015, 2016).

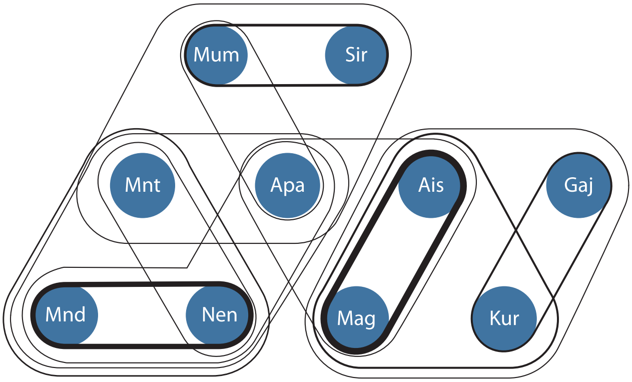


Figure 10. Glottometric diagram with probabilistically filled blank cells ($\varepsilon > 2.5$)

---

[3] Note that because values are calculated probabilistically, it is possible for ε values not to be integers, which is not normally the case. We are grateful to Siva Kalyan for providing the calculations that went into Figure 10 and Table 9.

Table 9. Glottometric figures with probabilistic blanks

Group	ε	κ	ς
Mag-Ais	24.90	0.82	20.34
Mnd-Nen	24.25	0.73	17.63
Mum-Sir	12.00	0.57	6.85
Mnd-Nen-Mnt	7.90	0.29	2.32
Mag-Ais-Kur	7.25	0.36	2.58
Kur-Gaj	6.00	0.43	2.59
Mnd-Nen-Apa	5.10	0.23	1.17
Mnd-Nen-Mnt-Apa	4.90	0.13	0.63
Mag-Ais-Kur-Gaj	4.45	0.17	0.74
Apa-Mag-Ais	3.60	0.17	0.61
Nen-Mnt	3.20	0.29	0.93
Apa-Mum	2.90	0.29	0.84
Mnt-Apa-Mum-Sir	2.50	0.08	0.19

4. Discussion and conclusions

Historical Glottometry is a powerful tool for understanding and visualizing the often very complicated relationships among sister languages. In combining a deep analytical understanding of the historical developments in a family with insightful descriptive statistics, it offers a potent combination of nuance and objectivity. For this, Kalyan & François's (2018) proposal deserves praise.

The diagram in Figure 10 accurately represents several kinds of language divergence. The North Central languages Mum and Sirva, for example, have split off quite cleanly from the other Sogeram languages: their subgroup is supported by a subgroupiness value of 6.85, while the only contradicting isogloss, Apalɨ-Mum, has a subgroupiness of only 0.85. This can be considered the outcome of "language fissure," as defined by Ross (1997).

The diagram also contains a prototypical example of a linkage in the East Sogeram languages, Magɨ, Aisi, Kursav, and Gants. The Magɨ-Aisi-Kursav group has a subgroupiness value of 2.58, while the Kursav-Gants group has a nearly identical value of 2.59. Thus, these languages seem to have descended from a dialect continuum in which innovations were as likely to be shared among Kursav and the Aisian languages as between Kursav and Gants. A linkage diagram à la Ross (1988) is truly an excellent way to represent these relationships, as shown in Figure 11. Note also that this linkage is not geographically expected: Kursav is not located in between Aisian and Gants, but rather southeast of both of them. Thus the comparative picture revealed by the data is not what might be expected based on simple geographic adjacency (for more discussion on this point in the Sogeram context, see Daniels 2017b).

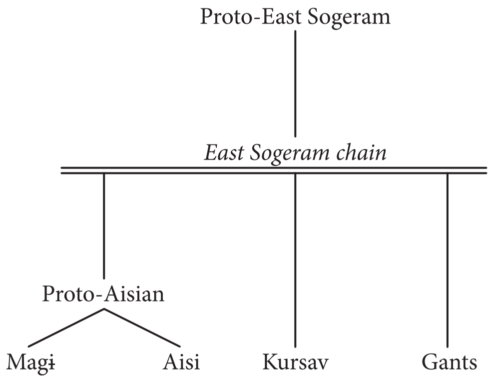


Figure 11. East Sogeram linkage diagram

Finally, glottometric diagrams can also capture linkage relationships that extend in more than two directions, which Ross's linkage diagrams cannot. The subgroups that Apalɨ participates in extend in three directions: north, west, and southeast.[4] The fact that glottometric diagrams can succinctly represent these kinds of complicated two-dimensional linkage patterns is one of the primary strengths of the model.

4.1 Improving the method

However, we do also have some suggestions for improving Historical Glottometry. First, we propose using exclusively shared innovations ε instead of subgroupiness

---

[4] While the question of *why* these languages diversified in the way they did is not the primary focus of our article, it is still worth briefly exploring possible motivations. Wade observed, speaking about the Apalɨ some 25 years ago, that "multilingualism is prestigious in the area" (Wade 1991: 14). It has been observed that as a language diversifies, the different dialects continue to be open to innovations from each other up to a certain point, but that eventually there is a drop-off in the amount of shared innovation they exhibit (Bowern 2013). It may be the case that high rates of multilingualism in the Sogeram area meant that this drop-off came later in their dispersal than it does in most cases of language diversification. If this was indeed the case, it would go some way in explaining the preponderance of cross-cutting isoglosses we see in the Apalɨ case, because speakers of pre-Apalɨ remained "open" to innovations from their neighbors for longer.

ς values for determining which isoglosses to draw in glottometric diagrams. Because frequent cross-cutting of isoglosses, as in the Apalɨ case, will depress cohesiveness κ values and thus also the subgroupiness values that are derived from them, representing only subgroups that have high subgroupiness scores will obscure cases where one language has innovated with several of its neighbors. Defining the cutoff in terms of exclusively shared innovations, however, gets around this by visualizing every subgroup for which there are at least two (or three, or however many the analyst deems appropriate) independent pieces of evidence for shared development.

Second, the question of how to handle blank cells needs work. For François (2014) it was not as crucial, as his data contain less than 4% blank cells. But our data contain 17% blanks – and at least one blank cell in over half the rows in our dataset – which significantly affects our ε values. Filling in blanks probabilistically seems to be the most promising way forward here, but there are several ways to go about that and we have only made a very small start.

Third, we believe there must be a much higher level of transparency about the coding decisions the analyst makes. The construction of a glottometric dataset is anything but a straightforward exercise in simple data entry. It is a deeply interpretive task that requires extensive qualitative analysis, because any glottometric dataset will inevitably be grounded, through and through, in a piece of comparative reconstruction. This fact needs to be recognized as central to the workings of Historical Glottometry: it is impossible to talk about innovations without talking about reconstructions.

This is why we focused so extensively in our discussion on the specific coding decisions that we made in constructing our glottometric dataset and also why we have made the dataset available online with this publication. This kind of commentary and transparency helps other researchers to evaluate our results and also helps them compare the Sogeram family with other families. It also allows debate about how best to represent the diachronic developments in a language family in a glottometric dataset. For example, we disagreed with François (2014) about how lexical innovations should be coded. Whether we are right or not in our criticism, the question has significant implications for our understanding of the Sogeram and Torres-Banks families, as we discussed in Section 3.2. This discussion was only made possible by the fact that we were able to compare our coding decisions to François's. Additional transparency about coding will enable more fruitful discussion along these lines.

These suggestions, we believe, will improve the method. But some weaknesses still remain, perhaps to be remediated in future work. First, Historical Glottometry cannot easily model language convergence (as discussed, for example, in Geraghty 1983). This phenomenon seems to have played a role in the development of the Sogeram family, particularly Apalɨ (Wade 1993), but we have not been able to address it here.[5] Second, and relatedly, the method has no way of incorporating the relative chronology of changes. In cases where it is possible to order innovations with respect to one another, this work must simply be done manually. Finally, the calculations make no distinction between highly unusual changes – which are generally seen as being more significant for subgrouping (Nichols 1996) – and typologically ordinary changes. It is not entirely clear how to avoid this, though. As Pelkey (2015: 402) correctly observes, "the introduction of weights in a quantitative analysis too easily becomes an outlet for comparativists to justify their own intuitions."

4.2 Historical Glottometry as part of the historical linguist's toolkit

We wish to wrap things up with a case study in how Historical Glottometry can complement other methods to provide a more complete historical-linguistic toolkit. The case study concerns a fairly narrow question – the issue of the genetic affiliation of Manat – but the lessons from it are more widely relevant.

One of the clearest results from our exploration of the Sogeram dataset is that, no matter how one runs the calculations, Manat groups more closely with Mand and Nend than it does with Apalɨ, Mum, and Sirva. In previous work, Daniels consistently put Manat in the Central Sogeram group with Apalɨ, Mum, and Sirva while pointing out that it shared some innovations with the West Sogeram languages Mand and Nend. The glottometric results indicate quite clearly that this was inaccurate: Manat belongs instead to a group with Mand and Nend and shares some innovations with the Central Sogeram langauges. It will be instructive to explore exactly how Daniels arrived at the conclusions he did so we can better understand the kinds of bias that can creep into the analytic process, and how a method like Historical Glottometry can correct that.

The first published subgrouping of the Sogeram languages was Daniels (2010a). At this point Gants was not generally considered a close relative of the Sogeram languages (but see Pawley 2006) and Magɨ had not yet been discovered by the linguistics community, so the paper dealt with the remaining eight Sogeram languages. Data on Mand (then known as Atemble) were extremely limited at that point but were nevertheless sufficient to identify Mand as a close relative of Nend's. Daniels proposed the subgrouping shown in Figure 12. He grouped Manat with Apalɨ on the strength of only one innovation, the lenition of word-initial plo-

---

[5] Rather than discussing possible convergence between the two dialects of Apalɨ that we felt Historical Glottometry could not easily handle, we have focused primarily on the Akɨ dialect.

sives, but considered that a “sufficiently unusual innovation to justify grouping them together” (2010a: 176).

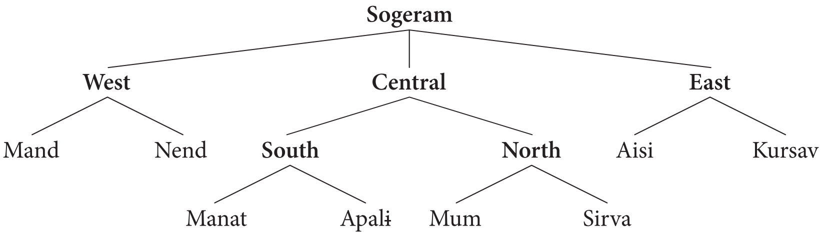


Figure 12. Sogeram subgrouping per Daniels (2010a)

Daniels placed Manat in the Central group on the basis of four innovations. The first was a reanalysis of the single labiovelar segment *k^w as two segments *kw. There was no direct evidence for this change, but it was posited to feed into a subsequent change that merged all tokens of *w with the following vowel into *u. The third change was a plosive voicing process that created prenasalized stops and voiced fricatives from voiceless stops; this change was shared sporadically with Nend. The final change was a sporadic vowel centering process that affected closed penultimate syllables.

Daniels also recognized two innovations that Manat shared with the West Sogeram languages: word-initial consonant loss, which was sporadic in Manat but regular in the West Sogeram languages; and an innovative set of pronouns, which was also found in Apalɨ.

All of this suggested to Daniels that Manat should tentatively be considered a Central Sogeram language, but he concluded that the question of its affiliation was still “not settled” (2010a: 185). The question continued to intrigue him, so after conducting more detailed morphosyntactic fieldwork on Manat he presented a paper at the Australian National University entitled “The Position of Manat Within the Sogeram Family” (Daniels 2010b) in which he reconstructed morphology and attempted to determine whether morphological innovations confirmed or contradicted the story he had found in the phonology. The morphological reconstruction was based on his Manat data plus published material on Nend (Harris 1990), Apalɨ (Wade 1997), and Mum (Wade 1993). It is now clear that the reconstruction was inaccurate in several respects, but the most serious error in this work was methodological. Because Daniels only had representatives of (what he considered) two primary subgroups of Sogeram, and because several of the reconstructions he proposed did not require any specific directionality in the innovations that they entailed, he was often unable to say which subgroup was innovative when they diverged in some way. So rather than counting innovations shared, he observed that Manat was separated from Nend by nine or ten isoglosses and from Apalɨ by only two or three. The implicit reasoning went as follows: (i) we already know that Manat either groups with the West Sogeram languages or the Central Sogeram languages; (ii) this morphological work shows Manat to be separated from Nend; (iii) therefore, since it does not group with Nend, it must group with the Central Sogeram languages Apalɨ and Mum.

The flaws in this reasoning are obvious. The assumption that Manat must group with one branch or the other, in particular, was unwarranted. And while Daniels (2010b) was aware of the problems and continued to consider Manat a less-than-perfect member of the Central Sogeram subgroup – for example, he disbanded the South Central node and placed Manat and Apalɨ as sisters under Central Sogeram – he also became less interested in pursuing the question further and in continually re-evaluating his conclusions about the genealogical affiliation of the language.

Subsequent years saw the additions of Gants (Daniels 2014) and then Magɨ (Daniels 2016) to the Sogeram group, but neither of these additions occasioned a reanalysis of the position of Manat. The opportunity for that, rather, came with the completion of Daniels's dissertation on the reconstruction of Proto-Sogeram (Daniels 2015). Here he revisited, with vastly improved data based on extensive fieldwork, the phonological and morphological reconstructions he had proposed in 2010. Although he did not notice it, the case for grouping Manat with the Central Sogeram languages gradually began to erode. The *k^w > *kw change and the *wV > *u change were combined in his analysis into a single change; the plosive voicing change was split into several separate changes, each of which affected a slightly different set of languages (and none of which perfectly singled out the putative Central Sogeram grouping of Manat, Apalɨ, Mum, and Sirva); and the support for the vowel centering change simply evaporated. This meant that the phonological evidence for the Central Sogeram group was now a single, rather unremarkable innovation: the labiovelar stop *k^w and the vowel that followed it merged and became a *ku sequence. The morphological case for grouping Manat with the Central Sogeram languages all but disappeared as a more complete reconstruction of Proto-Sogeram morphology emerged.

Meanwhile, small pieces of evidence that grouped Manat and the West Sogeram languages began to trickle in: the innovative set of pronouns identified in 2010 actually grew out of a few related, but distinct, innovations. A borrowing event triggered a very unusual sound change that only affected pronouns and verbs (Daniels & Brooks in prep.). A new word for 'who' appeared. The word for 'husband', *-mum, irregularly became *-mam. But none of these were significant enough, on their own, to trigger a re-evaluation of the subgrouping Daniels had already settled on. Besides, there were enough similar changes that grouped

Manat with its neighbors to the east: *ifu 'hit' changed verb class and became *ifa in Manat and Apalɨ. The 1sg irrealis suffixes *-ɨt-ɨŋ entered the imperative paradigm and became a 1sg imperative suffix in Manat, Apalɨ, and Mum. And enough of the old evidence for grouping Manat with Central Sogeram languages remained – especially the word-initial lenition change in Manat and Apalɨ that Daniels had considered unusual from the start.

Hindsight, as they say, is 20/20, and with its aid we can see what happened. An initial reconstruction with limited data weakly suggested the wrong subgrouping. A further reconstruction, again with limited data and also with poor methodology, weakly confirmed that subgrouping. Then, when better work with better data gradually revealed the right picture, confirmation bias prevented Daniels from seeing it, and he retained essentially the same subgrouping in his dissertation. He gave the family tree shown in Figure 1, reproduced here as Figure 13 (Daniels 2015: 51), and the contact chain diagram shown in Figure 3, reproduced as Figure 14 (Daniels 2015: 50), as the only visual representations of the relationships among the Sogeram languages.

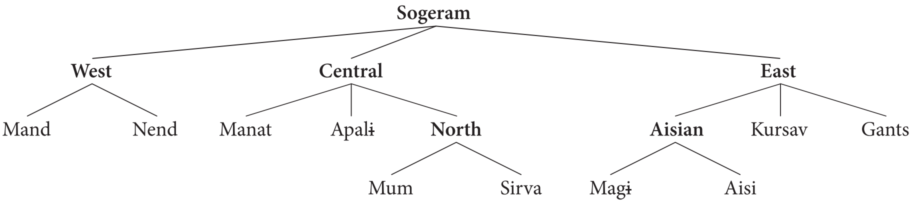


Figure 13. Sogeram subgrouping per Daniels (2015)

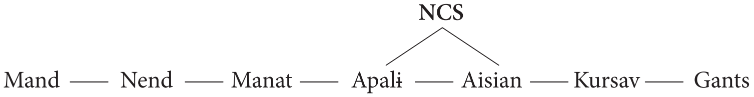


Figure 14. Sogeram contact chain per Daniels (2015)

Enter Historical Glottometry. While the dataset of innovations that we constructed was based entirely on the reconstruction in Daniels (2015), the objectivity of the calculations helped us to see what the reconstruction truly entailed: that Manat groups most closely with Nend and Mand. This, we believe, is where much of the value of the method lies.

Good historical reconstruction is an iterative process. Any reconstruction has implications for subgrouping and any subgrouping has implications for reconstruction. Since neither the reconstruction nor the subgrouping ever (in our experience) spring into being already fully formed, we must instead continually revisit and refine each. In refining one, we affect the other, so we must then refine that – and so the cycle continues. It is difficult, in this never-ending process of reevaluating what we know, to continue to treat working hypotheses as unproven, especially after they have served us faithfully for several years. This is why a method like Historical Glottometry makes such a valuable addition to the historical linguist's toolkit. It provides a helpful mathematical perspective on what the linguist already knows and, in so doing, helps break down preconceptions that have outlived their usefulness.

Acknowledgements

We are grateful to audiences at the Australian National University and the 93rd Linguistic Society of America meeting for helpful feedback on this research, and to Siva Kalyan, Alex François, and three anonymous reviewers for helpful feedback and discussion. All remaining errors and inaccuracies are our own. Fieldwork on the Sogeram languages was supported by U.S. Department of Education Javits Fellowship, the Living Tongues Institute for Endangered Languages, the UCSB Department of Linguistics, the UC Pacific Rim Research Program, NSF Grant BCS-1264157, and ELDP Grant IGS0221.

References

Bowern, Claire. 2013. Relatedness as a Factor in Language Contact. *Journal of Language Contact* 6:2.411–432. https://doi.org/10.1163/19552629-00602010

Campbell, Eric W. 2017. Otomanguean Historical Linguistics: Exploring the Subgroups. *Language and Linguistics Compass* 11:7.e12244. https://doi.org/10.1111/lnc3.12244

Daniels, Don. 2010a. A Preliminary Phonological History of the Sogeram Languages of Papua New Guinea. *Oceanic Linguistics* 49:1.163–193. https://doi.org/10.1353/ol.0.0068

Daniels, Don. 2010b. The Position of Manat Within the Sogeram Family. Talk presented at the Australian National University, Canberra, Australia, 2010.

Daniels, Don. 2014. Complex Coordination in Diachrony: Two Sogeram Case Studies. *Diachronica* 31:3.379–406. https://doi.org/10.1075/dia.31.3.03dan

Daniels, Don. 2015. A Reconstruction of Proto-Sogeram: Phonology, Lexicon, and Morphosyntax. University of California, Santa Barbara PhD dissertation.

Daniels, Don. 2016. Magɨ: An Undocumented Language of Papua New Guinea. *Oceanic Linguistics* 55:1.199–224. https://doi.org/10.1353/ol.2016.0004

Daniels, Don. 2017a. A Method for Mitigating the Problem of Borrowing in Syntactic Reconstruction. *Studies in Language* 41:3.577–614. https://doi.org/10.1075/sl.41.3.02dan

Daniels, Don. 2017b. Gants is a Sogeram Language. *Language and Linguistics in Melanesia* 35.82–93.

DeLancey, Scott. 2015. The Historical Dynamics of Morphological Complexity in Trans-Himalayan. *Linguistic Discovery* 13:2.60–79. https://doi.org/10.1349/PS1.1537-0852.A.463

Epps, Patience. 2013. Inheritance, Calquing, or Independent Innovation? Reconstructing Morphological Complexity in Amazonian Numerals. *Journal of Language Contact* 6:2.329–357. https://doi.org/10.1163/19552629-00602007

Fedden, Sebastian, Dunstan Brown, Greville Corbett, Gary Holton, Marian Klamer, Laura C. Robinson & Antoinette Schapper. 2013. Conditions on Pronominal Marking in the Alor-Pantar Languages. *Linguistics* 51:1.33–74. https://doi.org/10.1515/ling-2013-0002

Foley, William A. 2017. The Languages of the Sepik-Ramu Basin and Environs. In Bill Palmer, ed., 197–432.

François, Alexandre. 2005. Unraveling the History of the Vowels of Seventeen Northern Vanuatu Languages. *Oceanic Linguistics* 44:2.443–504. https://doi.org/10.1353/ol.2005.0034

François, Alexandre. 2011. Social Ecology and Language History in the Northern Vanuatu Linkage: A Tale of Divergence and Convergence. *Journal of Historical Linguistics* 1:2.175–246. https://doi.org/10.1075/jhl.1.2.03fra

François, Alexandre. 2012. The Dynamics of Linguistic Diversity: Egalitarian Multilingualism and Power Imbalance Among Northern Vanuatu Languages. *International Journal of the Sociology of Language* 2012:214.85–110. https://doi.org/10.1515/ijsl-2012-0022

François, Alexandre. 2013. Shadows of Bygone Lives: The Histories of Spiritual Words in Northern Vanuatu. *Lexical and Structural Etymology: Beyond Word Histories* ed. by Robert Mailhammer, 185–244. (= *Studies in Language Change*, 11.) Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614510581.185

François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. *The Routledge Handbook of Historical Linguistics* ed. by Claire Bowern & Bethwyn Evans, 161–189. New York: Routledge.

Geraghty, Paul. 1983. *The History of the Fijian Languages*. (= *Oceanic Linguistics*, Special Publication 19.) Honolulu: University of Hawai'i Press.

Harris, Kyle. 1990. Nend Grammar Essentials. *Two Grammatical Studies* ed. by John R. Roberts, 73–156. (= *Data Papers on Papua New Guinea Languages*, 37.) Ukarumpa: Summer Institute of Linguistics.

Heuvel, Wilco van den & Sebastian Fedden. 2014. Greater Awyu and Greater Ok: Inheritance or Contact? *Oceanic Linguistics* 53:1.1–36. https://doi.org/10.1353/ol.2014.0008

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the tree model: A framework for Historical Glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds), *Let's talk about trees: Genetic Relationships of Languages and Their Phylogenic Representation (Senri Ethnological Studies, 98)*. Ōsaka: National Museum of Ethnology. 59–89.

Korn, Agnes. 2016. A Partial Tree of Central Iranian: A New Look at Iranian Subphyla. *Indogermanische Forschungen* 121:1.401–434. https://doi.org/10.1515/if-2016-0021

Loughnane, Robyn & Sebastian Fedden. 2011. Is Oksapmin Ok? A Study of the Genetic Relationship Between Oksapmin and the Ok Languages. *Australian Journal of Linguistics* 31:1.1–42. https://doi.org/10.1080/07268602.2011.533635

Matisoff, James A. 2002. Genetic Versus Contact Relationship: Prosodic Diffusibility in South-East Asian Languages. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics* ed. by Alexandra Y. Aikhenvald & R. M. W. Dixon, 291–327. Oxford: Oxford University Press.

Nichols, Johanna. 1996. The Comparative Method as Heuristic. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change* ed. by Mark Durie & Malcolm Ross, 39–71. New York: Oxford University Press.

Palmer, Bill, ed. 2018. *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*. Berlin: De Gruyter Mouton.

Pawley, Andrew. 1999. Chasing Rainbows: Implications of the Rapid Dispersal of Austronesian Languages for Subgrouping and Reconstruction. *Selected Papers from the 8th International Conference in Austronesian Linguistics* ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 95–138. Taipei: Academica Sinica.

Pawley, Andrew. 2005. The Chequered Career of the Trans New Guinea Hypothesis: Recent Research and Its Implications. In Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide, eds., 67–107.

Pawley, Andrew. 2006. Madang Languages. *Encyclopedia of Language and Linguistics*, Vol. 7 ed. by Keith Brown, 429–432. Boston: Elsevier. https://doi.org/10.1016/B0-08-044854-2/04847-1

Pawley, Andrew, Robert Attenborough, Jack Golson & Robin Hide, eds. 2005. *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-speaking Peoples*. Canberra: Pacific Linguistics.

Pawley, Andrew & Harald Hammarström. 2017. The Trans New Guinea Family. In Bill Palmer, ed., 21–196.

Pelkey, Jamin. 2015. Reconstructing Phylogeny from Linkage Diffusion: Evidence for Cladistic Hinge Variation. *Diachronica* 32:3.397–433. https://doi.org/10.1075/dia.32.3.04pel

Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. (= *Pacific Linguistics C*, 98.) Canberra: Pacific Linguistics.

Ross, Malcolm. 1997. Social Networks and Kinds of Speech-Community Event. *Archaeology and Language I: Theoretical and Methodological Orientations* ed. by Roger Blench & Matthew Spriggs, 209–261. London: Routledge.

Ross, Malcolm. 2005. Pronouns as a Preliminary Diagnostic for Grouping Papuan Languages. In Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide, eds., 15–65.

Ross, Malcolm. 2015. The Argument Indexing of Early Austronesian Verbs: A Reconstructional Myth? *Historical Linguistics 2013: Selected Papers from the 21st International Conference on Historical Linguistics, Oslo, 5–9 August 2013* ed. by Dag T. T. Haug, 257–279. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.334.14ros

Sarvasy, Hannah. 2014. Four Finisterre-Huon Languages: An Introduction. *STUF – Language Typology and Universals* 67:3.275–295. https://doi.org/10.1515/stuf-2014-0017

Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur*, 786–787.

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

Wade, Martha. 1991. *An Overview of the Culture of the Apalɨ Speaking People*. Ms, Pioneer Bible Translators.

Wade, Martha. 1993. Language Convergence or Divergence: The Case of the Apalɨ (Emerum) Language. *Language and Linguistics in Melanesia* 24.73–93.

Wade, Martha. 1997. Switch Reference and Control in Apalɨ. *Language and Linguistics in Melanesia* 28.1–16.

Willis, David. 2011. Reconstructing Last Week's Weather: Syntactic Reconstruction and Brythonic Free Relatives. *Journal of Linguistics* 47:2.407–446. https://doi.org/10.1017/S0022226710000381

Z'graggen, John A. 1971. *Classificatory and Typological Studies in Languages of the Madang District*. (= *Pacific Linguistics C*, 19.) Canberra: Pacific Linguistics.

Z'graggen, John A. 1975a. *The Languages of the Madang District, Papua New Guinea*. (= *Pacific Linguistics B*, 41.) Canberra: Pacific Linguistics.

Z'graggen, John A. 1975b. The Madang-Adelbert Range Subphylum. *Papuan Languages and the New Guinea Linguistic Scene* ed. by Stephen A. Wurm, 569–612. (= *Pacific Linguistics C*, 38.) Canberra: Pacific Linguistics.

Z'graggen, John A. 1980. *A Comparative Word List of the Southern Adelbert Range Languages, Madang Province, Papua New Guinea*. (= *Pacific Linguistics D*, 33.) Canberra: Pacific Linguistics.

Address for correspondence

Don Daniels
Australian National University
9 Fellows Road
Acton, ACT 2601
Australia
don.daniels@anu.edu.au

Co-author information

Danielle Barth
Australian National University
College of Asia and the Pacific
Danielle.Barth@anu.edu.au

Wolfgang Barth
Australian National University
College of Asia and the Pacific
Wolfgang.Barth@anu.edu.au

Save the trees

Why we need tree models in linguistic reconstruction (and when we should apply them)

Guillaume Jacques and Johann-Mattis List
Centre des Recherches Linguistiques sur l'Asie Orientale | Max Planck Institute for the Science of Human History

Skepticism regarding the tree model has a long tradition in historical linguistics. Although scholars have emphasized that the tree model and its long-standing counterpart, the wave theory, are not necessarily incompatible, the opinion that family trees are unrealistic and should be completely abandoned in the field of historical linguistics has always enjoyed a certain popularity. This skepticism has further increased with the advent of recently proposed techniques for data visualization which seem to confirm that we can study language history without trees. In this article, we show that the concrete arguments that have been brought up in favor of achronistic wave models do not hold. By comparing the phenomenon of incomplete lineage sorting in biology with processes in linguistics, we show that data which do not seem as though they can be explained using trees can indeed be explained without turning to diffusion as an explanation. At the same time, methodological limits in historical reconstruction might easily lead to an overestimation of regularity, which may in turn appear as conflicting patterns when the researcher is trying to reconstruct a coherent phylogeny. We illustrate how, in several instances, trees can benefit language comparison, although we also discuss their shortcomings in modeling mixed languages. While acknowledging that not all aspects of language history are tree-like, and that integrated models which capture both vertical and lateral language relations may depict language history more realistically than trees do, we conclude that all models claiming that vertical language relations can be completely ignored are essentially wrong: either they still tacitly draw upon family trees or they only provide a static display of data and thus fail to model temporal aspects of language history.

Keywords: tree model, Historical Glottometry, phylogenetic networks, incomplete lineage sorting

https://doi.org/10.1075/jhl.17008.mat
Journal of Historical Linguistics 9:1 (2019), pp. 128–166. issn 2210-2116 | e-issn 2210-2124

1. Introduction

All languages develop by descent with modification (Darwin 1859): linguistic material is transferred from generation to generation of speakers, with slight modifications in pronunciation, denotation, and grammar potentially summing up to changes large enough that when two or more linguistic varieties have been separated in some way, be it by geographical or political separation of their speakers, they may become mutually incomprehensible. It is true that not all linguistic material is necessarily inherited from the parent generation. Linguistic material can easily be transferred across linguistic boundaries or diffuse across similar speech varieties. This, however, does not change the fact that the primary process of language transmission is through childhood acquisition of a first language (Ringe, Warnow & Taylor 2002: 61; Hale 2007: 27–48). The fact that languages that are mutually incomprehensible and markedly different can share a common genetic origin was one of the great insights of 19th-century linguistics, and even if lateral forces of diffusion are able to drastically change the shape of languages, this does not invalidate the crucial role that vertical transmission plays in language history. We follow Labov (2007: 347) in strictly distinguishing transmission of language via first language acquisition from diffusion via contact as two distinct processes.

In the following, we will aim to substantiate this viewpoint. We will start from a brief overview of the historical debate between proponents of tree models and proponents of wave models in the history of linguistics (Section 2), then introduce the core arguments of the new debate regarding trees and waves (Section 3). After this, we will defend tree thinking in historical linguistics by showing that even patterns which do not look tree-like at first glance can still be explained using a branching tree model (Section 4.1); we will also show that, conversely, some patterns that appear to demonstrate common inheritance may in fact go back to processes of language contact, which can be readily incorporated into a rooted network model in which a family tree model serves as a backbone representing inheritance with horizontal edges representing borrowing events (Section 4.2).

After presenting several examples that illustrate the benefits of trees in historical language comparison (Section 5) and also pointing to examples of their obvious shortcomings (Section 6), we conclude that both tree- and non-tree-like processes need to be taken into account when trying to draw a realistic scenario of language history. The logical and practical necessity of using both models for treelike and non-tree-like evolution shows that we cannot simply abandon the tree model in historical linguistics, but should rather work on integrating vertical transmission and horizontal diffusion in a common framework.

2. Dendrophobia and dendrophilia in linguistics

In order to get a clearer picture of the major arguments brought up to support or to dismiss the family tree model, it is useful to have a closer look at the origins of the tree model and the discussions that it sparked. In the following, we will give a brief overview on the development of tree thinking ("dendrophilia") and tree skepticism ("dendrophobia") in linguistics, from its beginnings up to the present day.

2.1 Tree thinking in Schleicher's work

Although he was not the first to draw language trees,[1] it was August Schleicher (1821–1866) who popularized tree thinking in linguistics. In two early papers from 1853 (Schleicher 1853a, 1853b) and numerous studies published thereafter (e.g. Schleicher 1861, 1863), Schleicher propagated the idea that the assumptions about language history could be best "illustrated by the image of a branching tree" (Schleicher 1853a: 787).[2] It should be noted that there was no notable influence from Darwin's writings in his work. It is more likely that Schleicher was influenced by stemmatics (for manuscript comparison, see Hoenigswald 1963: 8). Even today, historical linguistics has certain tendencies that resemble tendencies found in the field of stemmatics much more closely than they do in evolutionary biology. It seems that Schleicher's enthusiasm for the drawing of language trees had quite an impact on Ernst Haeckel (1834–1919) (Sutrop 2012), since – as Schleicher himself pointed out (Schleicher 1863: 14) – linguistic trees by then were explicit, drawing on and visualizing concrete data from real-world languages; this contrasts to Darwin's Origins, in which the data were abstract and illustrated hypothetical taxa (Darwin 1859).

Despite his seemingly radical idea to model language history as a process of diversification exclusively via branching and splitting, it is important to note that Schleicher was not a careless proponent of tree thinking. In his work we find many examples that show that he was aware of potential problems resulting from the tree model. In his open letter to Haeckel, Schleicher explicitly pointed to problems of language mixing, using Latin and its descendants as an example; he compared this mixing to plant hybrids in biology, identifying this hybridiza-

---

[1] The first trees and networks depicting language development date back to at least the 17th century; for details, see List et al. (2016), Morrison (2016), and Sutrop (2012).

[2] Our translation. Original text: "[Diese Annahmen, logisch folgend aus den Ergebnissen der bisherigen Forschung,] lassen sich am besten unter dem Bilde eines sich verästelnden Baumes anschaulich machen."

tion as a second factor leading to differentiation (Schleicher 1863: 18). In his earlier work, he explicitly mentioned language contact and borrowing of linguistic features as a process central in language history and development (Schleicher 1861: 6); he also emphasized the importance of distinguishing borrowed traits from inherited traits in language classification (Schleicher 1848: 30). Continuing the analogy between language development and species evolution, Schleicher also pointed to the difficulty in finding sharp borders between languages, dialects, and speech varieties ("Sprache, Dialekt, Mundarten und Untermundarten"); this fact finds a mirror in the difficulty in distinguishing between species and individuals (Schleicher 1863: 21). This last point in particular clearly demonstrates that Schleicher did not think that language splits were exclusively the product of abrupt separations of speakers and that he was aware of the idealizing aspect of the Stammbaum.

2.2 Tree skepticism in the work of Schmidt and Schuchardt

Schleicher's tree thinking, however, did not last long in the world of historical linguistics. By the beginning of the 1870s, Hugo Schuchardt (1842–1927) and Johannes Schmidt (1843–1901) published critical views, claiming that vertical descent was but one aspect of language evolution (Schmidt 1872; Schuchardt 1900). While Schmidt remained very vague in his criticism, Schuchardt was more concrete, pointing in particular to the problem of diffusion between very closely related languages: "We connect the branches and twigs of the family tree with countless horizontal lines and it ceases to be a tree" (Schuchardt 1900: 9).[3]

While Schuchardt's observations were based on his deep knowledge of the Romance languages, Schmidt drew his conclusions from a thorough investigation of shared cognate words in the major branches of Indo-European. In this investigation, he found patterns of words that were in a strong "patchy distribution" (see List et al. 2014) – that is, a distribution that showed many gaps across the languages under investigation, with only a few (if any) patterns that could be found across all languages. One seemingly surprising fact was, for example, that while Greek and Sanskrit shared about 39% of cognate vocabulary (according to Schmidt's count; see Geisler & List 2013) and Greek and Latin shared 53%, Latin and Sanskrit shared only 8%. Assuming that Greek and Latin had a common ancestor, Schmidt asserted, it was very difficult to explain the differences in the degree of vocabulary cognate between Greek and Sanskrit versus the vocabulary cognate between Latin and Sanskrit (Schmidt 1872: 24). Furthermore, this pattern

---

[3] Our translation. Original text: "Wir verbinden die Äste und Zweige des Stammbaums durch zahllose horizontale Linien, und er hört auf ein Stammbaum zu sein."

of patchy distributions seemed to be repeated in all branches of Indo-European that Schmidt investigated. Schmidt thus concluded:

> No matter how we look at it, as long as we stick to the assumption that today's languages originated from their common proto-language via multiple furcation, we will never be able to explain all facts in a scientifically adequate way.[4]
>
> (Schmidt 1872: 17)

Schmidt, however, did not stop with this conclusion but proposed another model of language divergence to take the place of the family tree model: "I want to replace [the 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"[5] (Schmidt 1872: 27). Since then, this new model, the so-called "wave theory" (*Wellentheorie* in German), has been energetically discussed in articles and books on historical linguistics, sometimes being promoted as the missing complement to Schleicher's *Stammbaumtheorie* (Campbell 1999: 187–200, Goebl 1983), sometimes being treated as its more realistic alternative (Gabelentz 1891: 194–195). Despite the apparent simplicity of the wave theory as reflected in its succinct presentation in handbooks of historical linguistics, the theory is the center of much confusion, both among linguists and among those without training in historical linguistics. This confusion is reflected not only in the discussions among dendrophilists and dendrophobists but also in the various attempts to visualize the waves. While Schmidt did not give a visualization in his 1872 book, he gave one three years later (Schmidt 1875: 199); this is shown in Figure 1 along with an English translation. It is difficult to interpret this figure, due not only to the scan quality (re-rendered from the original here) but also to its structure. It displays languages in a pie chart-like diagram in a quasi-geographic space. No information regarding the ancestral states of the relevant languages is given, and no temporal dynamics are shown. Being quasi-geographic, quasi-quantitative, and quasi-structured, the visualization is hard to understand, and the famous waves themselves are the last thing the figure brings to mind. Schmidt does not seem to ignore that evolution has a time dimension, but he does seem to deliberately neglect it when drawing his waves.

---

[4] Our translation. Original text: "Man mag sich also drehen und wenden wie man will, so lange man an der anschauung fest hält, dass die in historisches Zeit erscheinenden sprachen durch merfache gabelungen aus der Ursprache hervorgegangen seien, d.h. so lange man einen stammbaum der indogermanischen Sprachen annimmt, wird man nie dazu gelangen alle die hier in frage stehenden tatsachen wissenschaftlich zu erklären."

[5] Our translation, original text: 'Ich möchte an seine [*des Baumes*] stelle das bild der welle setzen, welche sich in concentrischen mit der entfernung vom mittelpunkte immer schwächer werdenden ringen ausbreitet.'

This confusion is also reflected in the scholarly literature. In the fifty years following Schmidt's publication, there was a wide range of various attempts to visualize the wave theory, ranging from Venn diagrams (Hirt 1905: 93) to early networks (Bonfante 1931: 174). The only publication known to us that retained Schmidt's pie-chart visualization was Meillet (1908: 134), in which the author applied it to Indo-European languages (see Geisler & List 2013 for details on early visualizations of the wave theory). After Schleicher's initial, rather pictorial, tree drawings, language trees began quickly to be schematized in historical linguistics. In contrast, the correct way to draw a wave remains disputed up to today. Some scholars have adopted the influential isogloss-map representation by Bloomfield (1973: 316) when they visualize the wave theory (Anttila 1972: 305, Burlak & Starostin 2005: 153–170, Holzer 1996: 13–48). Many scholars, however, still use alternative visualizations (Lehmann 1969[1962]: 124) or only mention the wave theory without further illustrations (Hock 1986). Visualization problems cannot be taken as primary arguments against a theory's validity. They may, however, reflect problems of internal coherence, and these problems of internal coherence are already reflected in the above-mentioned early interpretations of the Wellentheorie. It is therefore not surprising that Schmidt's wave theory provoked more negative than positive responses after its publication (Brugmann 1884; Hirt 1905).

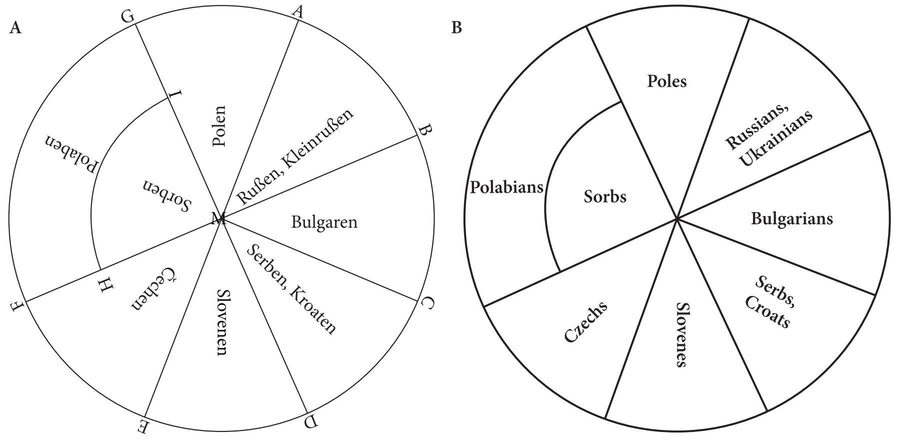


Figure 1. Schmidt's Wave Theory. A: Schmidt's visualization of the Wave Theory from 1875. B: English translation

2.3 Early arguments against the Stammbaumtheorie

Geisler & List (2013: 118–120) distinguish three different kinds of criticisms that have been raised against the family tree model (and in favor of the wave theory):

(i) practicability problems, (ii) plausibility problems, and (iii) adequacy problems. “Practicability problems” refers to the problems involved in using the tree model to analyze a given set of languages. Critics such as Schmidt (1872), mentioned above, point out in particular the issue of conflicting evidence as something that the tree model cannot accommodate. “Plausibility problems” refers to the realism of the family tree model, which is reflected in the obvious simplifications that the tree model necessitates. Critics addressing this point emphasize that languages do not necessarily split abruptly but slowly diverge, accompanied by complex waves of diffusion (Schuchardt 1900; Schmidt 1872). Questions of “adequacy” grow out of debates over the purpose of writing language history in historical linguistics. Critics complain that family trees reduce all the vital aspects of language history that are the substance of diversity within a language family to nothing more than the process of vertical descent. A similar argument has been brought up in biology, where the “tree of life” has been labeled the “tree of one percent,” due to the fact that only a minimal amount of the data seems to point to vertical descent (Dagan & Martin 2006).

Geisler & List (2013) emphasize that while all three types of criticism have been brought up against the family tree model, it is clear that their theoretical strength differs. Rejecting a model for reasons of practicability is straightforward, but this argument cannot be used to prove that a model is wrong or inadequate. An inability to find evidence for a tree in a given dataset is no proof that the family tree model is wrong, just like how an inability to distinguish borrowed from inherited traits (especially in deeper time depths) cannot be considered proof against the existence of tree-like divergence of languages. Geisler & List (2013) conclude that the stronger arguments against the family tree model are those that challenge its plausibility (particularly those that discuss the presumed split-process by which languages diverge) or its adequacy (particularly those that discuss its inability to provide a full picture of language history in all its complexity).

Putting adequacy to the side, the distinction between practicability and plausibility can be reframed as a distinction between the methodological and theoretical problems with the Stammbaumtheorie. Debates over practicability question the methodological possibility of inferring language trees from linguistic data (in essence, questioning the power of the methods available to us), while debates over plausibility question the adequacy of the model itself. While Schmidt’s arguments were largely methodological in nature, pointing to conflicts in the data (which are mostly based on a misunderstanding of the nature of scientific inference and phylogenetic reconstruction, as pointed out in detail in Geisler & List 2013), Schuchardt’s arguments are theoretical. He questions the process of divergence itself, claiming that languages do not split in an abrupt, binary fashion; rather, they slowly diverge, while at the same time exchanging material in a non-vertical manner. An even greater problem, though one not addressed in Schuchardt (1900), is the possibility of convergence. When convergence leads to hybridization, it clearly contradicts the core of the tree model (see Section 6), as rooted trees can be mathematically defined as directed, acyclic graphs in which all nodes have no more than one parent. Interestingly, Schleicher was well aware of these problematic theoretical aspects of the tree model. He explicitly pointed to the possibility of hybridization (Schleicher 1848), and he emphasized the often gradient transition underlying language divergence (Schleicher 1863: 21). On the other hand, he also deliberately ignored these aspects in the family tree model, giving a strict preference to divergence and vertical inheritance.[6]

Proponents of the wave theory, on the other hand, were much less clear about the different processes they sought to model. Do wave-like processes of language change reflect borrowing among closely related languages, or are they intended to reflect language change in general? While Schuchardt (1900) seems to distinguish the two, pointing to horizontal lines ("horizontale Linien") that make a network out of a tree, Schmidt (1872) is much less explicit, although he often invokes the idea of gradual transitions between language borders (Schmidt 1875: 200), thus emphasizing the gradualness of diversification rather than the interference of vertical and lateral processes in language change. Given the diversity of opinions and the lack of concreteness, it is difficult to determine a core theory to which scholars refer when mentioning the wave theory; while some see the wave theory as the horizontal counterpart of the family tree (Baxter 2006: 74), others see the wave theory as a theory explaining linguistic divergence (Campbell 1999: 188–191).

3. The new debate on trees and waves

Accompanying the "quantitative turn" in historical linguistics in the beginning of the 21st century (List 2014: 209–210), the debate over trees and waves has been recently revived. While most textbooks had previously treated the two models as working together to provide a complementary view of external language history[7]

---

[6] Yet he may have tried to visualize genetic closeness independently of elapsed time since separation, as can be seen from the tree in Schleicher (1861: 7), where he notes that the length of the lines indicated the divergence time, while the distance between the lines the degree of genetic closeness ("Die länge der linien deutet die zeitdauer an, die entfernung derselben von einander den verwantschaftsgrad"). This can be interpreted in such a way that Schleicher tried to include potential convergence after separation into his trees.

[7] External language history is here used in the sense of Gabelentz (1891: 179–290), who distinguishes it from internal language history, which points to different stages of one and the same language.

(Lehmann 1992; Anttila 1972) or as models of two different aspects of language change (Campbell 1999), more and more linguists now discuss the models as two opposing perspectives on language change (Heggarty, Maguire & McMahon 2010; François 2014). One reason for the revival of the discussion can certainly be found in the prevalence of trees in recent phylogenetic studies in historical linguistics (Gray & Atkinson 2003; Atkinson & Gray 2006; Ringe, Warnow & Taylor 2002; Pagel 2009). Both trees and waves had been playing a less prominent role in the field for some time prior to the "quantitative turn";[8] however, biological methods for phylogenetic reconstruction applied to large linguistic datasets can utilize a tree diagram to analyze and display large amounts of language data in a way that is much more transparent than the classical method of identifying shared innovations, leading to a recent increase in the use of tree diagrams.

Yet not long after the first biological software packages began being used for phylogenetic tree reconstruction in linguistics, new techniques for visualizing splits networks[9] provided by the SplitsTree software package (Huson 1998) offered scholars a fresh view on conflicts in their data. Often propagated as a reconciliation of tree and wave theory (Bryant, Filimon & Gray 2005; Ben Hamed & Wang 2006; McMahon & McMahon 2005) and easy to apply to linguistic distance data, splits networks have quickly become a very popular tool in historical linguistics (Gray, Bryant & Greenhill 2010; Heggarty, Maguire & McMahon 2010; Ben Hamed 2005; Bowern 2010).

3.1 Phylogenetic tree reconstruction after the quantitative turn

Classical phylogenetic tree reconstruction in historical linguistics is very similar to the process of cladistics in biology (Hennig 1950; see also Lass 1997: 105–171), insofar as it makes use of a small set of characters which are inherently weighted and represent unique innovations in order to uncover the phylogeny of a language family. The idea of unique innovations – changes that define a subgroup – is very old in linguistics and can be found in work as early as that of Karl Brugmann (1849–1919), although it was later scholars such as Isidore Dyen who popularized the principle in historical linguistics (see Chrétien 1963 and Dyen 1953). Brugmann himself justified the use of shared innovations in subgrouping as follows:

---

[8] Even Morris Swadesh was extremely wary of using his lexicostatistic method for producing family trees. Instead, he published a map on "interrelationships of American Indian languages" that came closer to the wave theory in its interpretation (Swadesh 1959: 23).

[9] Most of these techniques have been based on the NeighborNet algorithm (Bryant & Moulton 2004), but see Hurles (2003) for the earliest example of splits networks in linguistics known to us.

> The only thing that can shed light on the relation among the individual language branches […] are the specific correspondences between two or more of them, the innovations, by which each time certain language branches have advanced in comparison with other branches in their development.[10]
>
> (Brugmann 1967[1886]: 24)

The reason why linguists put such a great emphasis on shared innovations in subgrouping is obvious: while related languages can easily share features they have retained from their common ancestor, features which separate them from other languages in the same family and which may be interpreted as a new development can provide a strong argument for subgrouping. The problem, however, which is often downplayed in this context, is how to *identify* these exclusively shared innovations. If languages share common features (apomorphies in cladistic terminology), this does not necessarily mean that these features qualify as innovations (synapomorphies), since they could likewise have (i) been borrowed (see Section 4.2.1), (ii) been retained from the common ancestor of all languages (symplesiomorphies), (iii) emerged independently (homoplasies), or (iv) been erroneously annotated as shared features. Furthermore, differential loss or further development of features in subgroups may easily mask shared innovations, and consequently an innovation that was originally shared by a group of languages may give the impression of being patchily distributed. This is further complicated by the fact that variation of linguistic features occurs in all languages and might very well be traced back to the ancestral language. If this is the case and variation is later resolved randomly across the lineages, what looks like a shared innovation is in fact a shared retention or an independent development, a combination of (ii) and (iii), a possibility that will be further discussed in Section 4.1. None of these problems are new to historical linguistics: we can already find all of these points apart from the problem of variation in the proto-language in Brugmann (1884), who concludes that proposed innovations must be *frequent* enough to reduce the possibility of chance in order to be applicable to subgrouping (see also Dyen 1953).

It is difficult to give a concrete definition of frequency in historical linguistics, as scholars often intuitively weight characters, assigning more importance to certain kinds of evidence (e.g., form similarities in morphological paradigms; see Nichols 1996) than to other types (e.g., isolated lexical items or frequent sound change patterns which are likely to recur independently), and most debates

---

[10] Our translation. Original text: “Das einzige nun, was auf das Verhältnis der einzelnen Sprachzweige zu einander[, auf die Art des Hervorgangs der Einzelsprachen aus der idg. Ursprache] Licht werfen kann, sind die besonderen Übereinstimmungen zwischen je zwei oder mehreren von ihnen, die Neuerungen, durch die jedesmal gewisse Sprachzweige gegenüber den andern in der Entwicklung vorangeschritten erscheinen.”

regarding subgrouping center around the question of how different types of evidence should be weighted or how data should be interpreted. As an example, see the discussion in Sagart (2015), in which the author proposes that the innovations presented in Blust (1999) are better interpreted as retentions.

Phylogenetic approaches that had originally been developed for evolutionary biology offer a different approach to the problem; they use a larger pool of characters and explicit models of character evolution to automatically find the phylogenetic tree that best explains the data according to different criteria (likelihood, parsimony) while simultaneously determining which characters have been retained and which have been innovated (Greenhill & Gray 2012). Classical linguists often mistrust these methods, criticizing their "black box" character.[11] While the criticism is justified to some extent, it should be kept in mind that it is not the methods themselves which are non-transparent or inaccurate, but rather their application and the data they are applied to. Methods for phylogenetic reconstruction, be they based on parsimony, maximum likelihood, or Bayesian inference, are not black box methods per se. As the methods are based on an explicit modelling of evolutionary processes,[12] the tree or forest of trees they infer is based on detailed historical scenarios in which the history of each character in the data is calculated.

3.2 Linguistic data and data-display networks

As mentioned before, splits networks enjoy a considerable popularity in recent quantitative approaches in historical linguistics. Unfortunately, many scholars have failed to understand that splits networks are merely a tool for data display (Morrison 2010) and not a tool that directly produces a phylogenetic analysis. Splits networks are very useful for exploratory data analysis, notably:

> (1) the automatic extraction of previously unknown patterns with regard to groups of objects, without using known structures in the data; (2) the detection of anomalous objects in the dataset; and (3) providing a compact representation of the dataset, which can be easily visualized as a connected graph.
>
> (Morrison 2014: 2)

---

[11] This "character" is specifically reflected in the fact that multiple steps that lead to a certain conclusion (i.e. in phylogenetic reconstruction) are rarely shown to the users. Rather, users see only the aggregated results.

[12] This includes parsimony, since we are not talking about statistical modelling, but process modelling, which is usually a simple birth-death process in parsimony as well as in maximum likelihood and Bayesian inference.

However, they do not produce a hypothesis regarding the way in which languages or biological species have diverged or recombined, and they must be strictly distinguished from explicit evolutionary networks, which display "evolutionary relationships between ancestors and descendants" (Morrison 2011: 43).

The claim that splits networks are not equivalent to phylogenetic trees or phylogenetic networks often leads to confusion among scholars, as it does not seem to be very clear precisely what phylogenetic trees and networks are intended to represent. For us, the crucial difference between data-display approaches and true phylogenetic accounts is the lack of or presence of an explicit time dimension displaying events of divergence (or recombination)[13] of lineages. Whenever we are dealing with attested divergence, as, for example, in the case of mutually unintelligible languages which are obviously genetically related, we are dealing with at least one ancestral variety from which the attested varieties developed. How many further ancestral languages we assume at different stages of the development of the language family depends on the power of our methods, the time depths involved, and language-family-specific factors. However, if we ignore the ancestral varieties in our analysis completely, as is done with splits networks, we lose all temporal dynamics, and as a result end up with nothing more than a representation of the data rather than a concrete hypothesis on the development of the languages under investigation.

3.3 Shared innovations and Historical Glottometry

A very recent approach that has also attempted to dismiss the tree model is the theory of Historical Glottometry (François 2014; Kalyan & François 2018). Glottometry results from dissatisfaction with conflicting data in historical linguistics, dissatisfaction similar to that expressed in Schmidt (1872). Additionally, glottometry follows Ross (1988) in assuming that language divergence can proceed in terms of both concrete separation ("social split" according to François 2014) and dialect divergence. While the former process involves the complete separation of the speakers of a given language, mostly based on geographic dislocation of parts of a population, the latter involves the slow divergence of language varieties into dialect areas which may later result in a complete split and the loss of mutual intelligibility. Essentially, this argument resembles that of Schuchardt (1900), as it attacks the process of concrete language split as it is visually suggested by the tree model. Given the high diffusibility of linguistic features across mutually intelligible varieties, reconstructing a fully resolved tree showing language divergence

---

[13] When dealing with recombination of lineages (like under the assumption of language mixing; see Section 6), a tree model is not enough and a network has to be used.

in split processes may be difficult, if not impossible, in a scenario of language divergence. Ross (1988) uses the term “linkage” to refer to closely related language varieties that diffused rather than separated and uses specifically marked multifurcating nodes (polytomies) to highlight these varieties in his genetic subgrouping of Oceanic languages. Kalyan & François (2018) criticize this solution as unsatisfying, emphasizing that polytomies mask the fact that innovations can easily spread across dialect networks, thus creating intersecting, fuzzy subgroups. The solution proposed by Historical Glottometry is to use the classical comparative method to collect shared traits, intended to represent exclusively shared innovations, for the language family under investigation, then to display these traits as weighted isogloss maps in which weighting is represented by the thickness of a given isogloss. This is illustrated in Figure 2a, in which four hypothetical languages are given that are connected by three isoglosses, out of which two are in conflict with each other.

Three general problems with the method of Historical Glottometry need to be mentioned here. First, the resulting visualizations can by no means qualify as phylogenetic analyses, as they lack any time dimension. They are more similar to data display networks, and the fact that isoglosses are aggregated in numeric weights indicating isogloss strength makes them little more informative than splits networks produced with the NeighborNet algorithm. This does not mean that the measures proposed by glottometry do not have their specific value, but unlike the tree model, which displays a concrete evolutionary hypothesis, glottometric diagrams are mere tools for data visualization, as they do not allow ancestral languages to be included in the analysis.[14]

Second, the use of the term “innovation” in Historical Glottometry is logically problematic. According to the practice reported in François (2014), all instances in which a form in one language deviates in some respect from its reconstructed proto-form are interpreted as innovations. It seems to be further assumed that an innovation starts with its first introduction by a speaker and is diffused during the period of mutual intelligibility (François 2014: 178). Parallel innovations – i.e., innovations which look similar but happened independently of each other – are acknowledged as such (François 2016: 57), but when it comes to computing the diagrams, they are not distinguished from uniquely shared innovations; François

---

[14] Mathematically, the isogloss model proposed by glottometry corresponds to a hypergraph, in which edges can connect more than one vertex (Newman 2010: 122–123). Given that hypergraphs are equivalent to bipartite networks, it also seems that with the existing metrics applied in glottometry, not all mathematical possibilities are exhausted, and instead of weighting isoglosses using the cohesiveness value proposed in (François 2014), it might be interesting to look into different projections of bipartite into monopartite graphs (Newman 2010: 124–125).

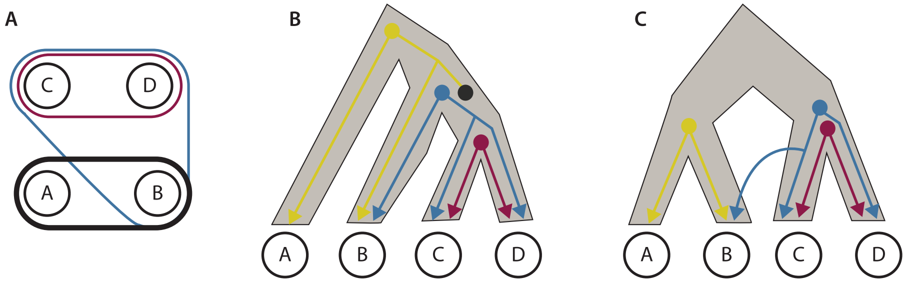


Figure 2. Historical glottometry, the family tree model, and an evolutionary network
Legend:

A: A glottometric diagram with weighted isoglosses drawn between languages sharing an innovation which are in apparent conflict with each other.
B: A demonstration of how the scenario in A can be explained with help of a family tree by assuming differential loss of the black isogloss.
C: An evolutionary network representing another possible explanation for the patterns; assumes that the blue innovation was borrowed from language C into language B.

(2014: 177) does not even attempt to distinguish between the two. Thus, innovations in glottometry represent two different processes: namely, (i) cases of unique deviation from linguistic traits inherited from the Ursprache (true shared innovations in the cladistic sense); and (ii) cases of parallel development.

Leaving aside the fact that using a proto-language to identify innovative traits silently acknowledges a tree-like divergence from the beginning, even if it turns out to be a star-phylogeny from which all descendants separated at once, this broad notion of shared innovations in the practice of glottometry bears practical and theoretical problems, especially given that the identified shared and parallel innovations in glottometry are used as an argument against tree-like patterns of separation of ancestral languages. Shared innovations in the cladistic sense are never in conflict with a tree, since they are *defined* as those elements which constitute the tree. They are rigorously distinguished from shared retentions, lateral transfer, and parallel developments (Fleischhauer 2009). Scholars often overlook this, since they interpret the term "shared innovation" as a *descriptive* term when the term in fact is meant to be *explanatory*. When labeling certain features as shared innovations, these scholars seem to provide a mere description of the data, while the term additionally denotes a judgment – an explanation for a certain phenomenon. The descriptive use of explanative terminology can be seen as a general problem in linguistic terminology, as reflected in terms like "pronominalization" (see Jacques 2016: 2), "polysemy" (see François 2008; List, Terhalle & Urban 2013), or "assimilation" (see List 2014: 32). In all these cases, the terms do not only describe a phenomenon but also explain it. While descriptively, "assimilation" could be seen as a process by which a sound becomes more similar to a neighboring sound, in most definitions scholars further add that this process is due to the *influence* of the neighboring sound (Campbell & Mixco 2007:16). The term is thus not only used to describe a phenomenon, but also to explain it. The same applies to the term "shared innovation." On the one hand, scholars use it to denote a set of similarities shared by a certain group of languages; on the other hand, they use it to denote synapomorphies, namely shared inherited similarities which define a subgroup in the cladistic tree.

The specific confusion involving the term "shared innovation," however, is not restricted to linguistics; it also occurs in biology (De Laet 2005). A cladistic analysis seeks to identify which out of a large pool of possible characters could be used to define a subgroup and thus potentially reflect true shared innovations. If a supposed set of innovations shows internal conflict with possible tree topologies, this means, from a cladistic perspective, that some of these innovations have been wrongly proposed. This is illustrated in Figure 2b, in which the data from Figure 2a are explained by differential loss of a shared character in one clade of a tree. Given that we can often hardly distinguish whether homologous characters in languages are due to independent change or inheritance, a fact which is explicitly admitted by François (2014), conflicts with possible tree topologies can by no means be taken as rigorous proof that a substantial amount of the data cannot be explained by a tree.[15] Interestingly, this was emphasized much earlier in the history of linguistics when Brugmann (1884) criticized the wave theory by Schmidt (1872), because Schmidt had similarly assumed that all exclusively shared traits could have originated only once, ignoring the possibility of erroneous judgments, parallel development, borrowing, shared retention, and chance.

Third, given that Kalyan & François (2018) admit that innovations develop *somewhere*, their approach is by no means less agnostic than the use of multifurcating tree topologies by Ross (1988), as we would assume that an innovation first occurs in a small community from which it spreads outwards. Theoretically, it may thus be possible to draw explicit pathways of diffusion which could be rendered as horizontal edges in an evolutionary network, as illustrated in Figure 2c. Since Historical Glottometry refuses to increase the level of explicitness in data-display, its analyses remain unsatisfactory, as historical linguistics should have more to offer than vague statements regarding shared traits between language varieties.

---

[15] A further problem, which is often ignored, is the possibility of erroneous annotations (see Wichmann 2017 for a more detailed account on false positives and false negatives in cladistic subgrouping).

4. Saving the trees from the critics

Given the logical necessity of allowing for divergence, a specific part of language history can be modeled with the help of a tree if specific processes like recombination (hybridization, creolization; see Section 6) can be excluded. That such a tree model does not necessarily represent all aspects of language history is obvious; even the strongest tree proponents would not deny this. Whether the amount of inheritance versus borrowing in language history is as low as it has been presumed for biology, a field in which tree critics have labeled the tree of life as the "tree of one percent" (Dagan & Martin 2006), is an interesting question worth being pursued further. However, given that we know that language varieties can diverge to such an extent that they lose mutual intelligibility, it is clear that a model for language history which handles divergence and splits of lineages is necessary. How these splits proceed in the end – whether they are best viewed as multifurcations after the split of a larger dialect continuum in several parts or as bifurcations – depends on our insights into the language family under investigation and into the processes of external language change in general.

When scholars point out that a given dataset lacks a tree-like signal, or that the tree-like signal for the subgrouping of a given language family is not strong, they often take this as direct evidence for large-scale language contact or linkage scenarios (Ross 1988). This, however, is by no means the only explanation for reticulations in datasets, and there are many other reasons why a given data selection may fail to reveal a tree (see the general overview in Morrison 2011: 44–66). The most obvious, and in cases of large datasets most frequent, reason is erroneous codings, which occur particularly in those cases where the data have not been thoroughly checked by experts in the field (Geisler & List 2010) or where automatic analyses have introduced a strong bias. Another obvious reason for reticulation in a dataset is the selection of the data. Commonalities in sound change patterns and grammatical features, for example, often do not represent true shared innovations, but rather independent development. Additionally, it is often very hard to distinguish between synapomorphy and homoplasy, especially for sound changes (Chacon & List 2015: 182–183); this is exacerbated by the fact that the majority of sound change patterns are extremely common, while rare sound changes are often very difficult to prove.

Apart from borrowing, dialect differentiation, data coding, and homoplasy, another often overlooked cause of conflicts in the data is the phenomenon of "incomplete lineage sorting" (Galtier & Daubin 2008). Incomplete lineage sorting is a well-known process in biology, in which polymorphisms (characters which are differently expressed in the same population, e.g., eye color) in the ancestral lineages are inherited by the descendant species when rapid divergence occurs

(Rogers & Gibbs 2014). Incomplete lineage sorting can explain, for instance, why 30% of the genes in a gorilla's genome have more similarity to the human or chimpanzee genome than the same genes in the human and chimpanzee genomes have to each other – surprising, given that human and chimpanzee are the closest relatives (Scally et al. 2012). In a recent study, List et al. (2016) proposed that incomplete lineage sorting may likewise occur in language history, given the multiple sources of polymorphisms in language change, ranging from near synonymy of lexical items via suppletive paradigms to word derivation.

Apart from these polymorphisms which may or may not be inherited across lineages before they are later randomly resolved, a further language-internal factor not mentioned by List et al. (2016) is that of sociolinguistic variation. This variation can occur in an entire population or even within a single speaker. The process of incomplete lineage sorting is further illustrated in Figure 3, where the two aspects – namely, sociolinguistic variation and language-internal variation – are contrasted. Note that in neither of these cases do we need to invoke strong language contact or situations of large-scale diffusion in dialect networks. Both patterns are perfectly compatible with a "social split" situation as invoked by François (2014), although they are based on fully resolved bifurcating trees. This shows that supposed reticulations or a lack of tree-like signal in the data do not necessarily prove the absence of tree-like patterns of divergence. They, rather, expose the weakness of our methods for finding the tree in the forest of individual histories of linguistic traits. In the following sections, we will illustrate this in more detail by showing how variation inherited from an ancestor language may be lost incompletely across lineages and by showing how the failure to identify true innovations may lead us astray when searching for convincing phylogenies.

4.1 Inherited variation and incomplete lineage sorting

Lexically-specific sound changes play an important role in Historical Glottometry, based on the assumption that they are "strongly indicative of genealogy, because they are unlikely to diffuse across separate languages" (François 2014: 178). Out of 474 shared traits which are classified as innovations in François (2014), 116 (24%) belong to this type. In view of the low diffusibility of such traits,[16] overlapping isoglosses constitute a major problem for the tree model from the point of view of supporters of glottometry. Regardless of whether lexically-specific sound changes have more difficulty crossing language boundaries than other types of

---

[16] This assertion remains to be demonstrated, but we accept it for the sake of argument.

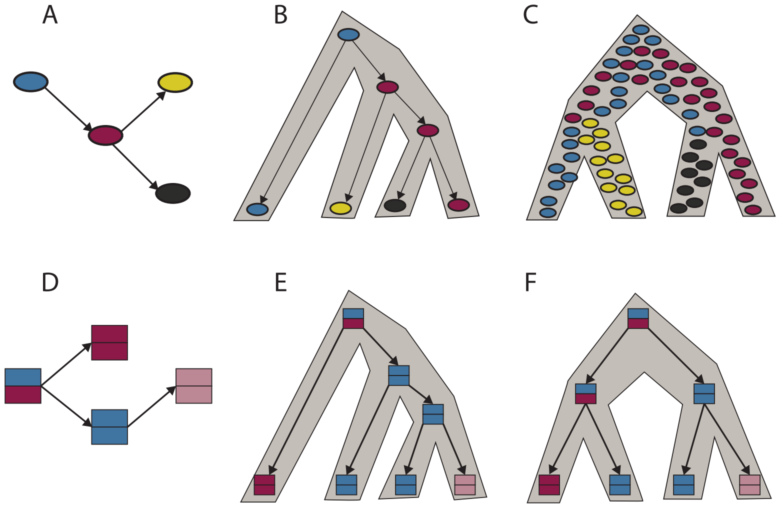


Figure 3. Incomplete lineage sorting due to sociolinguistic (A–C) and linguistic variation (D–F) and its impact on phylogenetic reconstruction and genetic subgrouping

Explanation:
A: Pattern of known directional evolution of a character (e.g., a sound change pattern)
B: A parsimonious tree resulting from the pattern in A
C: Alternative pattern assuming that the blue character already evolved in the ancestral language where it was used as a variant along with the original red character. Since the variation already occurred at the time of the ancestral language, it was inherited in the two descendant languages from which the character further developed. As a result, another tree topology can be reconstructed.
D: Example of a process of paradigm levelling
E: Parsimonious scenario of one tree topology, variant 1
F: Parsimonious scenario of one tree topology, variant 2

features, overlapping innovations can, as mentioned above, also be accounted for by assuming the existence of variation in the proto-language.[17]

Languages are never completely uniform, and fieldwork linguists working on unwritten languages commonly notice that even siblings can present significant differences in the pronunciation of certain words or even morphological paradigms (see, for instance, Genetti 2007: 29–30). While some innovations can spread quickly to the entire community (or at least to all members of a specific

---

[17] In this case, however, we can no longer speak of true innovations in the cladistic sense, given that, as mentioned above, the term "innovation" is explanatory and not descriptive and presupposes that a trait is uniquely shared by the subgroup that it defines.

age-group), in other cases it is possible for two competing forms (innovative vs. archaic) to remain used in the same speech community for a considerable period of time. This is observed in particular with sporadic changes, such as irregular metatheses, dissimilation, assimilation, or item-specific analogy.

When language differentiation occurs while forms are still competing, daughter languages can inherit the competing forms; subsequently, the innovative form may eventually prevail or disappear in a non-predictable way in each daughter language. If such a situation occurs, the distribution of the innovation will not directly match a particular node. This phenomenon is better illustrated by analogical levelling than by sporadic sound changes, as in the case of the former, the variation comes from well-understood morphological alternations that have been generalized in different ways in different language varieties, though the same account would be valid for the sporadic changes.

To illustrate how alternations and variation in the proto-language can blur the phylogeny, we take two examples from Germanic: the Proto-Germanic noun *knabō, *knappaz 'boy' and the dative second plural pronoun *izwiz or *iwiz.

4.1.1 *Alternations*

The reflexes of Proto-Germanic *knabō, knappaz 'boy',[18] an n-stem noun whose reflexes in the modern and ancient languages are particularly complex, can be found in Figure 4 (data from Kroonen 2011: 71, 128; Kroonen 2013: 294).

Using the attested ancient and modern forms with the known sound laws applied backwards, no fewer than four proto-forms have to be postulated: *knaban-, *knapan-, *knabban and *knappan-. Some languages have more than one reflex of this etymon (with diverging specialized meanings), and their distribution does not fit any accepted classification of the Germanic languages: for instance, while nearly all Germanicists agree on the existence of an Anglo-Frisian "Ingvaeonic" branch, we see that English sides with either German (in having a reflex of *knaban-) or with Dutch (the Old English reflex of *knapan-, lost in modern English) rather than with Frisian.

Unlike for most other language families, the detailed knowledge that has been accumulated regarding the history of Germanic languages allows us to go further than merely stating the presence of irregular correspondences: it is possible to account for them with a detailed model. It is now near-universally accepted that doublets such as these are due to the effect of Kluge's law (the change from *-Cn- to a geminate voiceless stop in pretonic position, *C being any pre-Germanic stop)

---

[18] The reflexes of this proto-form have developed distinct meanings in the attested languages, including 'squire', but this aspect is not considered here.

on the endings of n-stem nouns in Pre-Proto-Germanic (stage 0) (Kluge 1884; Kroonen 2011).

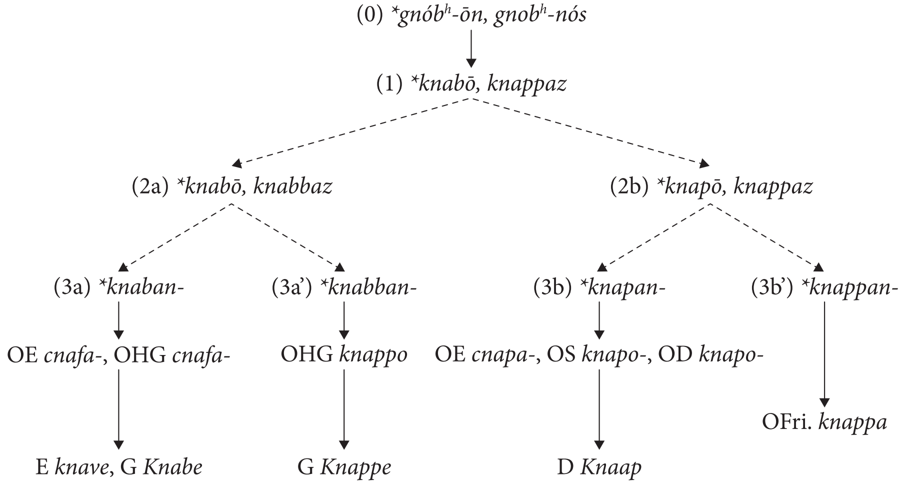


Figure 4. Several layers of variation: The etymon *knabō, knappaz 'boy' in Germanic

The paradigm of the noun 'boy' (as with all nouns of the same type) in Proto-Germanic (stage 1) had an alternation between *-b- and *-pp-. This complex alternation was variously levelled as *-b-/-bb- or *-p-/-pp- by stage 2; note that within a single language, not all items belonging to this declension class underwent levelling in the same way, and that some languages even have competing innovative (OE *cnapa* from *knapan-) and archaic (OE *cnafa* from *knaban-) forms for the same etymon (in this particular case, note that only the archaic form has been preserved with a different meaning in modern English *knave*). After simplification of the *-b-/-pp- alternation, all languages underwent a second wave of analogy, generalizing either the stem of the nominative (archaic *knaban- or innovative *knapan-) or that of the genitive (archaic *knappan- or innovative *knabban-), resulting in the four variants attested throughout the Germanic languages.

4.1.2 *Proto-variation*

Not all types of variations in the proto-language, however, can be straightforwardly accounted for by analogical levelings of paradigms, and in some cases alternative forms may have to be reconstructed back to the proto-language.

Germanic second person pronouns provide an example of this. The accusative and dative of the second person plural pronoun go back to two proto-forms: *izwiz (for Gothic *izwiz* and Old Norse *yðr* by dissimilation from *irwir; see Bugge 1855: 251) and *iwiz (Old English *ēow*, Old High German *iu*).

Some scholars argue that *iwiz is original and that *izwiz is an innovative form resulting from analogical leveling with the first plural accusative/dative *unsiz (through a stage *iwsiz followed by metathesis; see Kroonen 2013: 275). Since Gothic is incontrovertibly a member of the first branch of Germanic, this shared isogloss with Old Norse is clearly a problem if this form is a single-event innovation.

However, there is no clear consensus on the origin of these forms. Brugmann (1890: 804; see also Streitburg 1900: 265) argues instead that *izwiz and *iwiz are both ancient, as Proto-Indo-European had both *wes (Sanskrit second plural accusative-genitive-dative *vas*) and *swes (Welsh *chwi*). A particle *e (Greek *e-kêi* 'there', Sanskrit *a-sau* 'this') was added to both of these alternative forms, resulting in *ewes → *iwiz and *eswes → *izwiz respectively.[19]

Brugmann's idea implies that two proto-forms co-existed in Proto-Germanic for the accusative/dative of the second person plural. This is by no means a cross-linguistically uncommon state of affairs,[20] and this type of situation may account for irregularities in pronominal systems in other parts of the world (cf. François 2016).

4.1.3 *Concluding remarks*

We do not deny the potential value of item-specific changes of this type as evidence for studying phylogeny. However, it is obvious that isoglosses based on item-specific analogical levelling and sporadic sound change will overlap with each other, since competing forms can be maintained within the same language variety and only later be incompletely sorted across different lineages.

4.2 The problem of identifying lexical innovations

In order to identify inherited lexical innovations and distinguish them from recent borrowings, the method of Historical Glottometry uses a fairly uncontroversial criterion: etyma whose reflexes follow regular sound correspondences are considered to be inherited (François 2014: 176–178). Thus, whenever a common proto-form can be postulated for a particular set of words across several languages (which can thus be derived from this proto-form by the mechanical application of regular sound changes), it is considered in this model to be part of the inherited vocabulary and can be used, if applicable, as a common innovation.

---

[19] Proto-Indo-European *e shifts to Germanic *i in unaccented syllables.

[20] For instance, Japhug has several competing forms for the first and second person pronouns, as well as for the dative postposition, within a single variety and without counting dialectal variations (Jacques 2017: 624).

This approach, however, neglects an important factor: while regular sound correspondences are *necessary* for analyzing forms in related languages as cognates, i.e. originating from the same etymon in their common ancestor,[21] they are not *sufficient* due to the existence of undetectable borrowings and nativized loanwords.

4.2.1 *Undetectable borrowings*

Sound changes are not always informative enough to allow the researcher to discriminate between inherited word and borrowing. When a form contains phonemes that remained unchanged or nearly unchanged from the proto-language in all daughter languages (because no sound change, or only trivial changes, affected them), there is no way to know whether it was inherited from the proto-language or whether it was borrowed at a later stage.

This type of situation is by no means exceptional and can be found in various language families. We present here two examples of borrowings undetectable by phonology alone: ‘aluminum’ in Tibetan languages and ‘pig’ in some Algonquian languages. Amdo Tibetan *hajaŋ* ‘aluminum’ and Lhasa *hájã* ‘aluminum’ look like they regularly originate from a Common Tibetan form *ha.jaŋ.[22] This is, of course, impossible for obvious historical reasons, as aluminum came into use in Tibetan areas in the twentieth century at a time when Amdo Tibetan and Lhasa Tibetan were already mutually unintelligible. This word is generally explained (Gong Xun, personal correspondence) as an abbreviated form of *ha.tɕaŋ jaŋ.po* ‘very light’, but this etymology is not transparent to native speakers of either Amdo or Lhasa Tibetan. This word was coined only once[23] and was then borrowed into other Tibetan languages[24] and neighboring minority languages under Tibetan influence (as for instance Japhug *χajaŋ* ‘aluminum’). In this case, a phonetic borrowing from

---

[21] Note, however, that cognacy is a more complex concept than is usually believed (List 2016) and that even forms originating from exactly the same etymon in the proto-language may present irregular correspondences due to analogy.

[22] In Amdo Tibetan, Common Tibetan *h-*, *j-*, *-a* and *-aŋ* remain unchanged (Gong 2016). In Lhasa Tibetan, two sound changes relevant to this form occurred: a phonological high tone developed with the initial *h-*, and *-aŋ* became nasalized *ã*.

[23] We are not aware of detailed historical research on the history of this particular word, but in any case, it matters little for our demonstration whether it was first coined in Central Tibetan or in Amdo.

[24] In some Tibetan languages such as Cone, in which it is found as *hæ jãː* (Jacques 2014: 306), there is clear evidence that the word is borrowed from Amdo Tibetan and is not native (otherwise *yhæ jaː* would have been expected).

Amdo *hajaŋ* could only yield Lhasa *hájã*, since *h-* only occurs in high tone in Lhasa, and since final *-ŋ* has been transphonologized as vowel nasality.[25]

Several Algonquian languages share a word for 'pig' (Fox *koohkooša*, Miami *koohkooša* and Cree *kôhkôs*) that is ultimately of Dutch origin (Goddard 1974; Costa 2013). Hockett (1957: 266) pointed out that these forms must be considered to be loanwords "because of the clearly post-Columbian meaning; but if we did not have the extralinguistic information the agreement in shape (apart from M[enominee]) would lead us to reconstruct a [Proto-Central-Algonquian] prototype." The forms from these three languages could be regularly derived from Proto-Algonquian *koohkooša, a reconstruction identical to the attested Fox and Miami forms.[26]

Undetectable borrowings are also a pervasive phenomenon in Pama-Nyungan, where, with a few exceptions such as the Arandic and Paman groups, most languages present too few phonological innovations to allow easy discrimination of loanwords from cognates (Koch 2004: 46).

The same situation can be observed even if later sound changes apply to both borrowings and inherited words. Whenever borrowing takes place after the separation of two languages but before any diagnostic sound change has occurred in either the donor or the receiver language, or if the donor and the receiver languages have undergone identical sound changes up to the stage at which the borrowings occurred, phonology alone is not a sufficient criterion to distinguish between inherited words and loanwords.

A classic case is that of Persian borrowings in Armenian. As Hübschmann (1897: 16–17) put it, "in isolated cases, the Iranian and the genuine Armenian forms match each other phonetically, and the question whether borrowing [or common inheritance] has to be assumed must be decided from a non-linguistic point of view."[27] Table 1 presents a non-exhaustive list of such words, with the corresponding Proto-Iranian etyma.

The Armenian case shows that undetectable loans are not restricted to cases like those discussed above, in which a particular word contains only segments which have not been affected by sound changes at any point of the development

---

[25] Likewise, in the case of borrowing from Lhasa into Amdo, the rhyme *-aŋ* would be the only reasonable match for Lhasa *-ã*.

[26] However, it is true that, as shown in Taylor 1990, some Algonquian languages have forms that cannot regularly derive from a *koohkooša (for instance Ojibwe has *gookoozh* instead of expected *†gookoozh*), and that the ambiguity between cognate and loanword only exists with Fox, Miami and Cree. We wish to thank an anonymous reviewer for pointing out this fact.

[27] Our translation. Original text: "In einzelnen Fällen kann allerdings das persische und echt armenische Wort sich lautlich decken und die Frage, ob Entlehnung anzunehmen ist oder nicht, muss dann nach andern als sprachlichen Gesichtspunkten entschieden werden."

Table 1. Armenian words which cannot be conclusively demonstrated to be either borrowings from Iranian or inherited words on the basis of phonetic analysis

Armenian	Meaning	Proto-Iranian	Reference
naw	boat	*nāu-	Hübschmann (1897: 16–17, 201), Martirosyan (2010: 466, 715)
mēg	mist	*maiga-	Hübschmann (1897: 474), Martirosyan (2010: 466, 715)
mēz	urine	*maiza-	Hübschmann (1897: 474), Martirosyan (2010: 466, 715)
sar	head	*sarah-	Hübschmann (1897: 236, 489), Martirosyan (2010: 571)
ayrem	burn	*Haid-	Hübschmann (1897: 418), Martzloff (2016: 145)

from the proto-language into its daughter languages. Undetectable loans are also possible when a particular word is borrowed before any sound change which could affect its phonetic material occurred in either the giver or recipient language (or if both languages have identical sound changes for words of this particular shape), even if numerous sound changes occurred after borrowing took place. It is possible that post-borrowing sound changes even remove phonetic clues which could have allowed us to distinguish between loanwords and inherited words.

The situations illustrated above can be seen as clear evidence that undetectable borrowings can occur even when two language varieties are mutually unintelligible. Neglecting the distinction between inherited words and undetectable borrowings, as in the approach propagated by adherents of Historical Glottometry, amounts to losing crucial historical information; it does not seem justified to blame the family tree model for a shortcoming in our methods of linguistic reconstruction.

4.2.2 *Nativization of loanwords*

In the previous section, we discussed cases in which borrowing took place prior to diagnostic sound changes, thus making it impossible to effectively use sound changes to distinguish between loanwords and inherited words. There is, however, evidence that even when diagnostic sound changes exist, they may not always be an absolutely reliable criterion.

When a particular language contains a sizeable layer of borrowings from another language, bilingual speakers can develop an intuition of the phonological correspondences between the two languages and apply these correspondences to newly borrowed words, a phenomenon known as "loan nativization."

The best documented case of loan nativization occurs between Saami and Finnish (the following discussion is based on Aikio 2006). Finnish and Saami are only remotely related within the Finno-Ugric branch of Uralic, but Saami has borrowed a considerable quantity of vocabulary from Finnish, with some words being borrowed at a stage before most characteristic sound changes had taken place and others more recently. Table 2 presents examples of cognates between Finnish and Northern Saami that illustrate some recurrent vowel and consonant correspondences.

Table 2. Examples of sound correspondences in inherited words between Finnish and Northern Saami (data from Aikio 2006: 27)

Finnish	Northern Saami	Proto-Finno-Ugric	Meaning
käsi	*giehta*	*käti	hand
nimi	*namma*	*nimi	name
kala	*guolli*	*kala	fish
muna	*monni*	*muna	egg

The correspondence of final *-a* to *-i* and final *-i* to *-a* in disyllabic words found in the native vocabulary, as illustrated by the data in Table 2, is also observed in Saami words borrowed from Finnish. This includes recent borrowings, such as *mearka* from *merkki* 'sign, mark' and *báhppa* from *pappi* 'priest' (from Common Slavic *pǎpъ, itself of Greek origin), even though the sound change from Proto-Uralic to Saami leading to the correspondence *-a* : *-i* had already taken place at the time of contact. These correspondences are pervasive even in the most recent borrowings, to the extent that according to Aikio (2006: 36), "examples of phonetically unmarked substitutions of the type F[innish] *-i* > Saa[mi] *-i* and F[innish] *-a* > Saa[mi] *-a* are practically nonexistent, young borrowings included."

In cases such as *báhppa* 'priest', the vowel correspondence in the first syllable *á* : *a* betrays its origin as a loanword, as the expected correspondence for a native word would be *uo* : *a* as in the word 'fish' in Table 2 (Aikio 2006: 35 notes that this correspondence is never found in borrowed words).

However, there are cases in which recent loanwords from Finnish in Saami present correspondences indistinguishable from those of the inherited lexicon, as *barta* 'cabin' from Finnish *pirtti*, itself from dialectal Russian *pert'* 'a type of cabin'. These words show the same *CiCi* : *CaCa* vowel correspondence as the word 'name' in Table 2. Here, again, the foreign origin of this word is a clear indication that *barta* 'cabin' cannot have undergone the series of regular sound changes leading from Proto-Finno-Ugric *CiCi to Saami *CaCa*, and that instead the common vowel correspondence *CiCi* : *CaCa* was applied to Finnish *pirtti*.

Loan nativization can also occur between genetically unrelated languages. A clear example is provided by the case of Basque and Spanish (Trask 2000: 53–54, Aikio 2006: 21–23). A recurrent correspondence between Spanish and Basque is word-final *-ón* to *-oi*. Early Romance *-one (from Latin *-onem*) yields Spanish *-ón*. In Early Romance borrowings into Basque, however, this ending undergoes the regular loss of intervocalic *-n- (a Basque-internal sound change), and yields *-one → *-oe → *-oi*. An example of this correspondence is provided by Spanish *razón* and Basque *arrazoi* 'reason', both from Early Romance **ratsone* (from the Latin accusative form *ratiōnem*). This common correspondence has, however, been recently applied to recent borrowings from Spanish such as *kamioi* 'truck' and *abioi* 'plane' (from *camión* and *avión*). This adaptation has no phonetic motivation, since word-final *-on* is attested in Basque, and can only be accounted for as over-application of the *-oi : -ón* correspondence.

Nativization of loanwords is still a poorly investigated phenomenon and can only be detected in language groups whose historical phonology is already very well understood. While it has been documented for few languages as clearly as it has been for Saami and Basque, there is no reason to believe that this phenomenon is rare cross-linguistically. Its existence implies that sound laws cannot be used as an absolute criterion for distinguishing between inherited and borrowed common vocabulary (and thus between true shared innovations and post-innovation borrowings).

5. The benefit of trees in language comparison

In the previous section, we attempted to illustrate that not all patterns that look non-tree-like at first glance require a tree-free explanation, and that, at the same time, patterns that look like excellent examples of exclusively shared innovations may turn out to result from language contact. In addition, trees have several distinct advantages over more complex types of network representation, such as hybridization networks (Morrison 2011: 139), which makes the tree model preferable in the absence of evidence of its inapplicability (for more on this, see Section 6).

5.1 Parallel innovations

Trees can be used to detect cases of parallel innovations or features spread through contact. A typical example of such a situation is provided by Semitic. As shown in Table 3, Hebrew and Akkadian share no less than four common innovative sound changes in the evolution of their consonantal systems:

- *θ → *ʃ* (merging with *ʃ)
- *ð → *z* (merging with *z)
- *θ' → *s'* (merging with *s')
- *ɬ' → *s'* (merging with *s')

Table 3. Reflexes of Proto-Semitic coronals in a selected set of Semitic languages (Huehnergard 1997); innovative features shared by Akkadian and other languages are indicated in grey

Proto-Semitic	**Akkadian**	**Hebrew**	**Biblical Aramaic**	**Standard Arabic**
*t	*t*	*t*	*t*	*t*
*d	*d*	*d*	*d*	*d*
*θ	*ʃ*	*ʃ*	*t*	*θ*
*ð	*z*	*z*	*d*	*ð*
*s	*s*	*s*	*s*	*s*
*z	*z*	*z*	*z*	*z*
*ʃ	*ʃ*	*ʃ*	*ʃ*	*s*
*ɬ	*ʃ*	*ɬ*	*ɬ*	*ʃ*
*t’	*t’*	*t’*	*t’*	*tˤ*
*θ’	*s’*	*s’*	*t’*	*ðˤ*
*s’	*s’*	*s’*	*s’*	*sˤ*
*ɬ’	*s’*	*s’*	*ʕ*	*dˤ*

While phonology could seem, at first glance, to support grouping Akkadian and Hebrew together while excluding Aramaic and Arabic, the bulk of morphological and lexical innovations incontrovertibly support Akkadian being the first branch of the family and Aramaic and Hebrew being closer to each other than either of them is to Arabic (see, for instance, Hetzron 1976; Huehnergard 2006); this subgrouping is summarized in Figure 5. Bayesian phylogenetic analyses that have been proposed for Semitic confirm this insight (see, for instance, Nicholls & Ryder 2011). Here, the tree reconstructed from overwhelming, independently-collected evidence provides us with the near certainty that the innovative features shared by Hebrew and Akkadian are either parallel innovations or isoglosses transmitted through contact, and cannot be common innovations of these two languages.

5.2 Reconstruction of the Ursprache

Trees can be used to determine which features are reconstructible to the Ursprache and which are more likely to be later innovations. To illustrate this specific benefit of family trees, let us take the case of Semitic prepositions. Akkadian differs from the rest of the family in that its spatial prepositions are *in* and *ana*, while the other languages have forms going back to *l- and *b-. Geez (an Ethio-Semitic language, belonging to a sub-branch of West Semitic), however, has a cognate of Akkadian *in* : the preposition *ən*, which appears in some expressions (Huehnergard 2006: 16, Kogan 2015: 119). Additionally, Akkadian does have a frozen trace of the preposition *b- (Rubin 2005: 45–46). Since none of these four prepositions are the result of recent and obvious grammaticalization processes, there is no way without the tree model to decide which should be reconstructed to Proto-Semitic and which should not. Thanks to the Stammbaum in Figure 5, however, we know that since the prepositions *inV and *b- are attested (even as traces) in both Akkadian and West Semitic and are not recently grammaticalized, they can be safely reconstructed to Proto-Semitic.

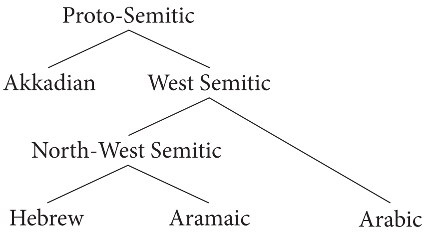


Figure 5. A simplified Stammbaum of Semitic languages

5.3 Directionality of change

As a byproduct of the reconstruction of particular features to the proto-language, trees can be used to determine the directionality of changes in ambiguous cases. While the directionality can sometimes be determined using the body of attested knowledge on sound changes (e.g. Kümmel 2007) or semantic changes (e.g., Urban 2011), there are still many isoglosses, particularly in inflectional morphology, whose interpretation as innovations or retentions is nearly impossible by direct comparison between languages.

As an example of the benefits of trees in determining the directionality of a semantic change, let us examine the root *ʔmr in Semitic (Kogan 2015: 233, 331, 544). This root is attested in various languages with a slightly different meaning; Table 4 provides its reflexes in several languages. The meaning of this root is highly divergent across these languages: it is a perception verb ('see', 'look at') in some languages and a verb of speech ('say', 'command') in others. It is not obvious at first glance which of the different meanings was the original one.

The family tree of Semitic, however, provides a scenario of how the meaning of this root evolved across the family. The use of this root as a perception verb is found in both Akkadian and Northwest Semitic (Ugaritic): perception is thus most likely to be the archaic meaning. Ugaritic, in which the root means both 'to look at' and 'say', represented an intermediate stage, where both meanings were still in competition (this may be a preservation of the Proto-West Semitic stage).

In Hebrew and Arabic, the use of this root as a perception verb has disappeared, and Arabic has further narrowed down its meaning to 'to command'.

Table 4. Reflexes of the root *ʔmr in several Semitic languages (Kogan 2015: 233, 331, 544)

Language	**Reflex**	**Meaning**
Akkadian	*amārum*	to see
Hebrew	*ʔāmar*	to say, to declare, to command
Ugaritic	*ʔmr*	to say; to look at
Arabic	*ʔamara*	to order

The pathway of semantic change (1) is a possible account of the evolution of the meaning of this root in Semitic, which is compatible with the tree in Figure 5.

(1) 'see, look at' → 'address' → 'say' → 'command'

In this particular case, the tree model does not only help us to solve an ambiguous question in Proto-Semitic reconstruction; it also provides evidence for a semantic change that might otherwise not have been clearly attested.

5.4 Common tendencies of language change

Many processes of linguistic change are overwhelmingly frequent and widespread. However, apart from highly controversial attempts to find a universal constant of lexical replacement rates (Swadesh 1955), most of the knowledge regarding change preferences in language history – be they family-specific, areal, or global – has never been explicitly modelled, since most scholars work from intuition about common tendencies. Language phylogenies and modern phylogenetic approaches, however, allow us to quantify the processes in various ways, and although most currently applied models lack linguistic realism, they offer a promising starting point for future efforts. In addition to intuitive accounts of frequency and cross-linguistic studies, such as the one regarding sound change by Kümmel (2007), there are other promising approaches: for example, phylogenetic approaches, in which the evolution of linguistic characters (phonetic, morphological, semantic) is modelled by inferring how the characters evolved along a given phylogeny, may yield interesting insights into common tendencies of language change. These approaches allow us to process larger amounts of data, but at the same time, they are not able to handle uncertainty in their inferences. Even less sophisticated approaches, such as weighted parsimony, can provide interesting insights into sound change patterns which frequently occur independently of each other along different branches of a tree (Chacon & List 2015). Static models of shared commonalities, like the isogloss-maps of glottometry, do not provide insights into the dynamics and tendencies of common processes of language change.

5.5 Language change and migration history

Trees can be used to make sense of population prehistory and can help to enhance the comparison of linguistic and archaeological evidence. Clues regarding the history and the spread of a language family can be obtained using the reconstructible vocabulary for particular nodes. For instance, the presence of a reconstructible etymon *kasp- ‘silver’ (Akkadian *kaspum*, Ugaritic *ksp,* and Hebrew *késeφ*, among other languages; Huehnergard 2012: 14–16) suggests that silver smelting could have been known to the speakers of Proto-Semitic, an idea supported by the evidence of cupellation in Syria as early as the 4th millennium BC (Pernicka, Rehren & Schmitt-Strecker 1998). Other metals, however, are only reconstructible to lower branches of the family; for instance ‘iron’ does not occur earlier than Proto-Canaanite (*barðill-, Hebrew *barzel*; cf. Kogan 2015: 287),[28] an observation compatible with the much later spread of iron technology (Yahalom-Mack & Eliyahu-Behar 2015).

Of course, as shown in Section 4.1, words that are compatible with the sound laws of inherited vocabulary may nevertheless be diffused by contact (especially a form like *kasp-, which remained unchanged in most of the ancient attested languages). As a result, “linguistic paleontology” should always be used with great caution. By using the tree model to advance our knowledge regarding directionality preferences in semantic shift and morphological change, as well as our knowledge regarding the strength of certain tendencies, we may be able to consolidate paleolinguistic evidence and finally put this highly controversial field on more solid ground.

6. The limits of the tree model

While the tree model has undeniable advantages and remains the most powerful model for understanding the vertical history of most languages, there undoubtedly remains a residue of cases in which this model is not applicable, even taking incomplete lineage sorting into account. These cases consist of instances in which one language results from the merger of two previously unintelligible languages (whether or not the two varieties are demonstrably related or not).

---

[28] Similar forms in other languages such as Akkadian *parzillum* ‘iron’ do not follow the regular correspondences and cannot be cognate.

The clearest and best documented example of this type is Michif, a contact language based on Canadian French and Plains Cree (Bakker 1997). Example (2), taken from Antonov (2015), illustrates the main features of this language (elements from French are in bold and those from Cree are underlined). Nearly all verbs and verbal morphology come from Plains Cree, except the verbs 'to be' and 'have'; these come from French, and retain the source language's complete irregular paradigms for these verbs, including French tense categories, as shown by (3). Most nouns and adjectives come from French. Some determiners are from French (the articles), but the demonstratives are from Cree; nouns can take the Cree obviative suffix *-(w)a,* and some nouns are compatible with possessive prefixes (like *o-* below).

(2) *o-**pâpa**-wa êtikwenn kî-wîkimê-yiw onhin **la***
3-father-OBV apparently PST-marry-3.OBV → 3.PROX this:AN:OBV DEF:FEM:SG
***fâm**-a*
woman-OBV
'Her father apparently married that woman…' (1: 8–9)

(3) ***stit=enn pchit orfelin***
BE:3SG:PST=INDEF:F:SG little orphan
'She was a little orphan' (1: 2)

The descent of a language like Michif, and potentially also the descent of less extreme contact languages, cannot be represented by the tree model, as the representation would require two roots (from languages belonging to unrelated families). A more complex type of network, a directed network with multiple roots, would be necessary to represent a language of this type. This might be fruitful, as the near-perfect division of the French and the Cree components of this language might allow for a meaningful representation of the nature of language mixture.

The applicability of the tree model on a global scale crucially depends on the rarity of languages like Michif. If, as the data available to us seem to show, this language is truly exceptional (because its genesis occurred in a very special setting that is unlikely to have existed at earlier stages of history), there are few obstacles against accepting the tree model to represent the vertical descent of languages.

7. Conclusion

In this paper, we have tried to save Schleicher's family tree model from being cut into pieces by critics speaking prematurely. We have shown that Schleicher himself was far more aware of the obvious insufficiencies of his tree model than is usually acknowledged in the literature, and that the wave theory by Schmidt, which is often praised as the alternative to the tree, never truly reached the level of sophistication necessary to depict the temporal dynamics of language history. After briefly introducing the new stage of the debate between proponents of trees and proponents of waves, we looked at Historical Glottometry, whose supporters stand as some of the strongest opponents of family trees in contemporary historical linguistics. We have shown, however, that Historical Glottometry suffers from the same problems as Schmidt's Wellentheorie, insofar as glottometry lacks temporal dynamics and is not capable of distinguishing true innovations from independently developed shared traits. We further substantiated this claim by illustrating how conflicts in linguistic data, which are taken as prima facie evidence against trees, can often be explained using a traditional family tree model, especially in cases where linguistic variation has been inherited from the ancestor language. On the other hand, we have shown how overlapping isoglosses, which are treated as evidence against tree-like evolution in language history, can likewise be explained by invoking classical processes of language contact. In order to further substantiate the claim that trees are worth being saved, we provided several examples of the usefulness of tree models in linguistic reconstruction, ranging from the detection of parallel innovations up to an enhancement of the methodology underlying linguistic paleography. We are aware that there are situations in language history, like language mixture, where trees cannot be used, but as long as these situations remain exceptional, we do not see any theoretical or practical justification for abandoning the family tree model as the standard to represent vertical aspects of language evolution.

Language history is incredibly complex and, even with more than 200 years of research into it, we have only seen the tip of the huge iceberg of possible processes in language evolution. No linguist would deny that not all aspects of language history are tree-like. Languages can split and branch when their speakers separate, but they do not necessarily do so; even after separation, languages may still easily exchange all kinds of linguistic material. We therefore agree with all tree skeptics that a language tree necessarily reduces linguistic reality, emphasizing only processes of vertical descent. On the other hand, however, we do not agree with the viewpoint that tree drawing per se is useless. Given our knowledge that we can, in theory, clearly distinguish processes of inheritance from processes of borrowing, we should make use of rooted phylogenies which distinguish vertical from lateral processes. While we explicitly acknowledge that integrated models which capture both vertical and lateral language relations may depict language history more realistically,[29] we do not accept the conclusion that vertical language change can be completely ignored. "Treeless" approaches, like Historical Glottom-

---

[29] These would be true evolutionary networks in the sense of Morrison (2011).

etry or splits networks, either silently still use family trees or only provide a static display of data and thus fail to model temporal aspects of language history.

Acknowledgements

This research was supported by the DFG research fellowship grant 261553824 (JML) and the ERC Starting Grant 715618 CALC (JML). We thank Dmitry Nikolaev, Gong Xun, John Cowan, Juho Pystynen, Julien Dufour, Martin Kümmel, Mathieu Ségui, Mikhail Zhivlov, Nathan Hill, Rémy Viredaz, Simon Greenhill, Thiago Chacon, Tiago Tresoldi, and Yoram Meroz for a vivid discussion after we circulated the first draft of the paper. We thank Russell D. Gray for providing additional pointers to missing bibliographic items and Malcom Ross for additional comments to our first draft. We also thank the anonymous reviewers for a challenging discussion.

Abbreviations

F	feminine
INDEF	indefinite
ILS	Incomplete Lineage Sorting
OBV	obviative
PST	past
SG	singular

References

Aikio, Ante. 2006. Etymological Nativization of Loanwords: A Case Study of Saami and Finnish. *Saami Linguistics* ed. by Ida Toivonen & Diane Nelson, 17–52. Amsterdam: Benjamins.

Antonov, Anton. 2015. The Obviative as a Differential Object Marker in Michif. Paper presented at 11th Biennial Conference of the Association for Linguistic Typology, Albuquerque, USA, August 1–3, 2015.

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? *Phylogenetic Methods and the Prehistory of Languages* ed. by Peter Forster & Colin Renfrew, 91–109. Cambridge/Oxford/Oakville: McDonald Institute for Archaeological Research.

Bakker, Peter. 1997. *A Language of Our Own: The Genesis of Michif, the Mixed Cree-French Language of the Canadian Metis*. Oxford: Oxford University Press.

Baxter, William H. 2006. Mandarin Dialect Phylogeny. *Cahiers de Linguistique – Asie Orientale* 35:1.71–114. https://doi.org/10.3406/clao.2006.1748

Ben Hamed, Mahé. 2005. Neighbour-Nets Portray the Chinese Dialect Continuum and the Linguistic Legacy of China's Demic History. *Proceedings of the Royal Society B* 272:1567.1015–1022. https://doi.org/10.1098/rspb.2004.3015

Ben Hamed, Mahé & Feng Wang. 2006. Stuck in the Forest: Trees, Networks, and Chinese Dialects. *Diachronica* 23:1.29–60. https://doi.org/10.1075/dia.23.1.04ham

Bloomfield, Leonard. 1973. *Language*. London: Allen & Unwin.

Blust, Robert. 1999. Subgrouping, Circularity, and Extinction: Some Issues in Austronesian Comparative Linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics* ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 31–94. Taipei: Academia Sinica.

Bonfante, Giuliano. 1931. I dialetti indoeuropei. *Annali del R. Istituto Orientale di Napoli* 4.69–185.

Bowern, Claire. 2010. Historical Linguistics in Australia: Trees, Networks and Their Implications. *Proceedings of the Royal Society B* 365:1559.3845–3854.

Brugmann, Karl. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen. *Internationale Zeischrift für allgemeine Sprachewissenschaft* 1.228–256.

Brugmann, Karl. 1890. *Grundriß der vergleichenden grammatik der indogermanischen sprachen*, Vol. 2. Strassburg: Karl J. Trübner.

Brugmann, Karl. 1967[1886]. *Einleitung und Lautlehre: vergleichende Laut-, Stammbildungs- und Flexionslehre der indogermanischen Sprachen*. Berlin, Leipzig: Walter de Gruyter.

Bryant, David, Flavia Filimon & Russell D. Gray. 2005. Untangling Our Past: Languages, Trees, Splits and Networks. *The Evolution of Cultural Diversity: A Phylogenetic Approach* ed. by Ruth Mace, Clare J. Holden & Stephen Shennan, 67–84. London: UCL Press.

Bryant, David & Vincent Moulton. 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution* 21:2.255–265. https://doi.org/10.1093/molbev/msh018

Bugge, Sophus. 1855. Die formen der geschlechtslosen persönlichen pronomina in den germanischen sprachen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 4:4.241–256.

Burlak, Svetlana Anatol'evna & Sergej Anatol'evič Starostin. 2005. *Sravnitel'no-istoričeskoe jazykoznanie*. Moscow: Akademia.

Campbell, Lyle. 1999. *Historical Linguistics: An Introduction (Second Edition)*. Edinburgh: Edinburgh University Press.

Campbell, Lyle & Mauricio Mixco. 2007. *A Glossary of Historical Linguistics*. Edinburgh: Edinburgh University Press.

Chacon, Thiago Costa & Johann-Mattis List. 2015. Improved Computational Models of Sound Change Shed Light on the History of the Tukanoan Languages. *Journal of Language Relationship* 13:3.177–204.

Chrétien, C. Douglas. 1963. Shared Innovations and Subgrouping. *International Journal of American Linguistics* 29:1.66–68. https://doi.org/10.1086/464714

Costa, David J. 2013. Borrowing in Southern Great Lakes Algonquian and the History of Potawatomi. *Anthropological Linguistics* 55:3.195–233. https://doi.org/10.1353/anl.2013.0013

Dagan, Tal & William Martin. 2006. The Tree of One Percent. *Genome Biology* 7:10.118. https://doi.org/10.1186/gb-2006-7-10-118

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

De Laet, Jan E. 2005. Parsimony and the Problem of Inapplicables in Sequence Data. *Parsimony, Phylogeny, and Genomics* ed. by Victor A. Albert, 81–116. Oxford: Oxford University Press.

Dyen, Isidore. 1953. Review of the book *Malgache et maanjan: Une comparaison linguistique* by Otto Chr. Dahl. *Language* 29:4.577–590. https://doi.org/10.2307/409983

Fleischhauer, Jens. 2009. A Phylogenetic Interpretation of the Comparative Method. *Journal of Language Relationship* 2.115–138.

François, Alexandre. 2008. Semantic Maps and the Typology of Colexification: Intertwining Polysemous Networks Across Languages. *From Polysemy to Semantic Change* ed. by Martine Vanhove, 163–215. (= *Studies in Language Companion Series*, 106.) Amsterdam: Benjamins. https://doi.org/10.1075/slcs.106.09fra

François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. *The Routledge Handbook of Historical Linguistics* ed. by Claire Bowern & Bethwyn Evans, 161–189. London: Routledge.

François, Alexandre. 2016. The Historical Morphology of Personal Pronouns in Northern Vanuatu. *Faits de Langues* 47.25–60.

Gabelentz, Hans Georg C. 1891. *Die Sprachwissenschaft*. Leipzig: T. O. Weigel.

Galtier, Nicolas & Vincent Daubin. 2008. Dealing With Incongruence in Phylogenomic Analyses. *Proceedings of the Royal Society B* 363:1512.4023–4029. https://doi.org/10.1098/rstb.2008.0144

Geisler, Hans & Johann-Mattis List. 2010. Beautiful Trees on Unstable Ground: Notes on the Data Problem in Lexicostatistics. *Die Ausbreitung des Indogermanischen. Thesen aus Sprachwissenschaft, Archäologie und Genetik* ed. by Heinrich Hettrich. Wiesbaden: Reichert.

Geisler, Hans & Johann-Mattis List. 2013. Do Languages Grow on Trees? The Tree Metaphor in the History of Linguistics. *Classification and Evolution in Biology, Linguistics and the History of Science: Concepts – Methods – Visualization* ed. by Heiner Fangerau, Hans Geisler, Thorsten Halling & William Martin, 111–124. Stuttgart: Franz Steiner Verlag.

Genetti, Carol. 2007. *A Grammar of Dolakha Newar*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110198812

Goddard, Ives. 1974. Dutch Loanwords in Delaware. *A Delaware Indian Symposium* ed. by Herbert C. Kraft, 153–160. Harrisburg: Pennsylvania Historical and Museum Commission.

Goebl, Hans. 1983. Stammbaum und Welle. *Zeitschrift für Sprachwissenschaft* 2:1.3–44.

Gong, Xun. 2016. A Phonological History of Amdo Tibetan Rhymes. *Bulletin of the School of Oriental and African Studies* 79:2.347–374. https://doi.org/10.1017/S0041977X16000070

Gray, Russell D., David Bryant & Simon J. Greenhill. 2010. On The Shape and Fabric of Human History. *Philosophical Transactions of the Royal Society B* 365:1559.3923–3933. https://doi.org/10.1098/rstb.2010.0162

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree Divergence Times Support the Anatolian Theory of Indo-European Origin. *Nature* 426:6965.435–439. https://doi.org/10.1038/nature02029

Greenhill, Simon J. & Russell D. Gray. 2012. Basic Vocabulary and Bayesian Phylolinguistics. *Diachronica* 29:4.523–537. https://doi.org/10.1075/dia.29.4.05gre

Hale, Mark. 2007. *Historical Linguistics: Theory and Method*. Malden/Oxford/Victoria: Blackwell.

Heggarty, Paul, Warren Maguire & April McMahon. 2010. Splits or Waves? Trees or Webs? How Divergence Measures and Network Analysis Can Unravel Language Histories. *Philosophical Transactions of the Royal Society B* 365:1559.3829–3843. https://doi.org/10.1098/rstb.2010.0099

Hennig, Willi. 1950. *Grundzüge einer Theorie der phylogenetischen Systematik*. Berlin: Deutscher Zentralverlag.

Hetzron, Robert. 1976. Two Principles of Genetic Reconstruction. *Lingua* 38:2.89–108.

Hirt, Herman. 1905. *Die Indogermanen*, Vol. 1. Strassburg: Trübner.

Hock, Hans Henrich. 1986. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.

Hockett, Charles F. 1957. Central Algonquian Vocabulary: Stems in k-. *International Journal of American Linguistics* 23:4.247–268. https://doi.org/10.1086/464417

Hoenigswald, Henry M. 1963. On the History of the Comparative Method. *Anthropological Linguistics* 5:1.1–11.

Holzer, Georg. 1996. *Das Erschließen unbelegter Sprachen*. Frankfurt am Main: Lang.

Huehnergard, John. 1997. *A Grammar of Akkadian*. Atlanta: Scholars Press.

Huehnergard, John. 2006. Proto-Semitic and Proto-Akkadian. *The Akkadian Language in its Semitic Context: Studies in the Akkadian of the Third and Second Millennium BC* ed. by G. Deutscher & N. J. C. Kouwenberg, 1–18. Leiden: Nederlands Instituut voor het Nabije Oosten.

Huehnergard, John. 2012. *An Introduction to Ugaritic*. Peabody: Hendrickson Publishers.

Hurles, Matthew E., Elizabeth Matisoo-Smith, Russell D. Gray & David Penny. 2003. Untangling Oceanic Settlement: The Edge of the Knowable. *Trends in Ecology and Evolution* 10:10.531–540. https://doi.org/10.1016/S0169-5347(03)00245-3

Huson, Daniel H. 1998. SplitsTree: Analyzing and Visualizing Evolutionary Data. *Bioinformatics* 14:1.68–73. https://doi.org/10.1093/bioinformatics/14.1.68

Hübschmann, Heinrich. 1897. *Armenische Grammatik, 1 Theil: Armenische Etymologie* Leipzig: Breitkopf und Härtel.

Jacques, Guillaume. 2014. Cone. *Phonological Profiles of Little-studied Tibetic Varieties* ed. by Jackson T.-S. Sun, 265–371. Taipei: Academia Sinica.

Jacques, Guillaume. 2016. Tangut, Gyalrongic, Kiranti and the Nature of Person Indexation in Sino-Tibetan/Trans-Himalayan. *Linguistics Vanguard* 2:1.1–13. https://doi.org/10.1515/lingvan-2015-0033

Jacques, Guillaume. 2017. Japhug. *The Sino-Tibetan Languages (Second Edition)* ed. by Graham Thurgood & Randy LaPolla, 614–634. London: Routledge.

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the tree model: A framework for Historical Glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds), *Let's talk about trees: Genetic Relationships of Languages and Their Phylogenic Representation (Senri Ethnological Studies, 98)*. Ōsaka: National Museum of Ethnology. 59–89.

Kluge, F. 1884. Die germanische Consonantendehnung. *Beitrage der deutschen Sprache und Literatur* 9.149–186.

Koch, Harold. 2004. A Methodological History of Australian Linguistic Classification. *Australian Languages: Classification and the Comparative Method* ed. by Claire Bowern & Harold Koch, 17–60. Amsterdam: Benjamins. https://doi.org/10.1075/cilt.249.06koc

Kogan, Leonid. 2015. *Genealogical Classification of Semitic: The Lexical Isoglosses*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781614515494

Kroonen, Guus. 2011. *The Proto-Germanic n-stems: A Study in Diachronic Morphophonology*. Amsterdam: Rodopi.

Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden: Brill.

Kümmel, Martin J. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert Verlag.

Labov, William. 2007. Transmission and Diffusion. *Language* 83:2.344–387.

Lass, Roger. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620928

Lehmann, Winfred Philipp. 1969[1962]. *Einführung in die historische Linguistik*. Heidelberg: Carl Winter.

Lehmann, Winfred Philipp. 1992. *Historical Linguistics (Third Edition)*. London: Routledge.

List, Johann-Mattis. 2014. *Sequence Comparison in Historical Linguistics*. Düsseldorf: Düsseldorf University Press.

List, Johann-Mattis. 2016. Beyond Cognacy: Historical Relations Between Words and Their Implication for Phylogenetic Reconstruction. *Journal of Language Evolution* 1:2.119–136. https://doi.org/10.1093/jole/lzw006

List, Johann-Mattis, Shijulal Nelson-Sathi, Hans Geisler & William Martin. 2014. Networks Of Lexical Borrowing and Lateral Gene Transfer in Language and Genome Evolution. *BioEssays* 36:2.141–150. https://doi.org/10.1002/bies.201300096

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* 11:39.1–17.

List, Johann-Mattis, Anselm Terhalle & Matthias Urban. 2013. Using Network Approaches to Enhance the Analysis of Cross-Linguistic Polysemies. *Proceedings of the 10th International Conference on Computational Semantics – Short Papers*, 347–353. Stroudsburg: Association for Computational Linguistics.

Martirosyan, Hrach K. 2010. *Etymological Dictionary of the Armenian Inherited Lexicon*. (= *Leiden Indo-European Etymological Dictionary Series*, 8.) Leiden: Brill.

Martzloff, Vincent. 2016. Arménien geri 'captif' : étymologie et phraséologie à la lumière d'un parallèle en albanien du Caucase. *Wékwos* 2.109–179.

McMahon, April & Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.

Meillet, Antoine. 1908. *Les dialectes indo-européens*. Paris: Librairie Ancienne Honoré Champion.

Morrison, David A. 2010. Review of the book *Sequence Alignment: Methods, Models, Concepts, and Strategies*, ed. by Michael S. Rosenberg. *Systematic Biology* 359:3.363–365.

Morrison, David A. 2011. *An Introduction to Phylogenetic Networks*. Uppsala: RJR Productions.

Morrison, David A. 2014. Phylogenetic Networks: A New Form of Multivariate Data Summary for Data Mining and Exploratory Data Analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 4:4.296–312. https://doi.org/10.1002/widm.1130

Morrison, David A. 2016. Genealogies: Pedigrees and Phylogenies Are Reticulating Networks Not Just Divergent Trees. *Evolutionary Biology* 43:4.456–473. https://doi.org/10.1007/s11692-016-9376-5

Newman, M. E. J. 2010. *Networks: An Introduction*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Nicholls, Geoff K. & Robin J. Ryder. 2011. Phylogenetic Models for Semitic Vocabulary. *Proceedings of the 26th International Workshop on Statistical Modelling* ed. by David Conesa, Anabel Forte, Antoni López-Quílez & Facundo Muñoz, 431–436.

Nichols, Johanna. 1996. The Comparative Method as Heuristic. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change* ed. by Mark Durie & Malcolm Ross, 39–71. New York: Oxford University Press.

Pagel, Mark. 2009. Human Language as a Culturally Transmitted Replicator. *Nature Reviews Genetics* 10.405–415. https://doi.org/10.1038/nrg2560

Pernicka, Ernst, Thilo Rehren & Sigrid Schmitt-Strecker. 1998. Late Uruk Silver Production by Cupellation at Habuba Kabira, Syria. *Metallurgica Antiqua: In Honour of Hans-Gert Bachmann and Robert Maddin* ed. by Thilo Rehren, Andreas Hauptmann & James D. Muhly 123–134. Bochum: Deutsches Bergbau Museum.

Ringe, Donald, Tandy Warnow & Ann Taylor. 2002. Indo-European and Computational Cladistics. *Transactions of the Philological Society* 100:1.59–129. https://doi.org/10.1111/1467-968X.00091

Rogers, Jeffrey & Richard A. Gibbs. 2014. Comparative Primate Genomics: Emerging Patterns of Genome Content and Dynamics. *Nature Reviews Genetics* 15:5.347–359. https://doi.org/10.1038/nrg3707

Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. (= *Pacific Linguistics C*, 98.) Canberra: Pacific Linguistics.

Rubin, Aaron D. 2005. *Studies in Semitic Grammaticalization*. Winona Lake: Eisenbrauns.

Sagart, Laurent. 2015. "East Formosan" and the PAN Palatals. Talk presented at the 13th International Conference on Austronesian Linguistics, Taipei, Taiwan, July 18–22, 2015. Available at https://www.academia.edu/22871609/

Scally, Aylwyn, Julien Y. Dutheil, LaDeana W. Hillier, Gregory E. Jordan, Ian Goodhead, Javier Herrero, Asger Hobolth, Tuuli Lappalainen, Thomas Mailund, Tomas Marques-Bonet, Shane McCarthy, Stephen H. Montgomery, Petra C. Schwalie, Y. Amy Tang, Michelle C. Ward, Yali Xue, Bryndis Yngvadottir, Can Alkan, Lars N. Andersen, Qasim Ayub, Edward V. Ball, Kathryn Beal, Brenda J. Bradley, Yuan Chen, Chris M. Clee, Stephen Fitzgerald, Tina A. Graves, Yong Gu, Paul Heath, Andreas Heger, Emre Karakoc, Anja Kolb-Kokocinski, Gavin K. Laird, Gerton Lunter, Stephen Meader, Matthew Mort, James C. Mullikin, Kasper Munch, Timothy D. O'Connor, Andrew D. Phillips, Javier Prado-Martinez, Anthony S. Rogers, Saba Sajjadian, Dominic Schmidt, Katy Shaw, Jared T. Simpson, Peter D. Stenson, Daniel J. Turner, Linda Vigilant, Albert J. Vilella, Weldon Whitener, Baoli Zhu, David N. Cooper, Pieter de Jong, Emmanouil T. Dermitzakis, Evan E. Eichler, Paul Flicek, Nick Goldman, Nicholas I. Mundy, Zemin Ning, Duncan T. Odom, Chris P. Ponting, Michael A. Quail, Oliver A. Ryder, Stephen M. Searle, Wesley C. Warren, Richard K. Wilson, Mikkel H. Schierup, Jane Rogers, Chris Tyler-Smith & Richard Durbin. 2012. Insights into Hominid Evolution from the Gorilla Genome Sequence. *Nature* 483:7388.169–175. https://doi.org/10.1038/nature10842

Schleicher, August. 1848. *Zur vergleichenden Sprachengeschichte*. Bonn: König.

Schleicher, August. 1853a. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3.786–787.

Schleicher, August. 1853b. O jazyku litevském, zvlástě ohledem na slovanský června. *Časopis Českého Museum* 27.320–334.

Schleicher, August. 1861. *Kurzer Abriss einer Lautlehre der indogermanischen Ursprache*. Weimar: Böhlau.

Schleicher, August. 1863. *Die Darwinsche Theorie und die Sprachwissenschaft*. Weimar: Hermann Böhlau.

Schmidt, Johannes. 1872. *Die Verwantschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.

Schmidt, Johannes. 1875. *Zur Geschichte des indogermanischen Vocalismus. Zweite Abteilung*. Weimar: Hermann Böhlau.

Schuchardt, Hugo. 1900. *Über die Klassifikation der romanischen Mundarten: Probe-Vorlesung, gehalten zu Leipzig am 30 April 1870*. Graz.

Streitberg, Wilhelm. 1900. *Urgermanische grammatik*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Sutrop, Urmas. 2012. Estonian Traces in the Tree of Life Concept and in the Language Family Tree Theory. *Journal of Estonian and Finno-Ugric Linguistics* 3.297–326.

Swadesh, Morris. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21:2.121–137. https://doi.org/10.1086/464321

Swadesh, Morris. 1959. The Mesh Principle in Comparative Linguistics. *Anthropological Linguistics* 1:2.7–14.

Taylor, Allan R. 1990. A European Loanword of Early Date in Eastern North America. *Anthropological Linguistics* 32:3/4.187–210.

Trask, Larry. 2000. Some Issues in Relative Chronology. *Time Depth in Historical Linguistics, Vol. 1* ed. by Colin Renfrew, April McMahon & Larry Trask, 45–58. Cambridge: The McDonald Institute for Archaeological Research.

Urban, Matthias. 2011. Asymmetries in Overt Marking and Directionality in Semantic Change. *Journal of Historical Linguistics* 1:1.3–47. https://doi.org/10.1075/jhl.1.1.02urb

Wichmann, Søren. 2017. Genealogical Classification in Historical Linguistics. *Oxford Research Encyclopedia of Linguistics*, 1–19. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.78

Yahalom-Mack, Naama & Adi Eliyahu-Behar. 2015. The Transition from Bronze to Iron in Canaan: Chronology, Technology, and Context. *Radiocarbon* 57:2.285–305. https://doi.org/10.2458/azu_rc.57.18563

Address for correspondence

Johann-Mattis List
Department of Linguistic and Cultural Evolution
Max Planck Institute for the Science of Human History
Kahlaische Str. 10
07745 Jena
mattis.list@shh.mpg.de

Co-author information

Guillaume Jacques
INALCO-CRLAO
Centre des Recherches Linguistiques sur
l'Asie Orientale

When the waves meet the trees

A response to Jacques and List

Siva Kalyan[1] and Alexandre François[1, 2]
[1] Australian National University |
[2] LaTTiCe (CNRS; ENS-PSL; Paris 3-USPC)

1. Introduction

We thank Jacques and List (henceforth, J&L) for their paper in defense of the family tree model in historical linguistics. They present powerful arguments in favor of the view that the tree model should be taken as the default model of language genealogy – except in cases of hybridization, which are rare.[1]

We find J&L's use of the notion of "incomplete lineage sorting" (Section 4.1) to be illuminating, and have learned much from their discussion of "undetectable borrowings" and loanword nativization (Section 4.2). However, we believe they have misunderstood the aims of Historical Glottometry (François 2014, 2017; Kalyan & François 2018), the model of language diversification that it assumes, and our reasons for making certain methodological choices when applying it. By clarifying these points, we hope to show that our approach to language diversification – and that of other researchers who subscribe to a wave-based approach to language genealogy – is in fact largely compatible with that of J&L.

2. Two ways of understanding the family-tree model

We believe that there is considerable difference between J&L's understanding of the family tree model and the way in which it is traditionally understood and applied in historical linguistics. Historical Glottometry (and other instantiations of the wave model) arose as a critique of historical linguists', not phylogeneticists', use of the tree model; thus, in order to understand the aims of Historical Glottometry and evaluate how well it fulfills these aims, it is necessary to clarify how historical linguists traditionally use and understand the tree model.

---

[1] See, however, Nichols (1992) and Donohue (2013: 222–224) on the prevalence of areal diffusion of structural features.

https://doi.org/10.1075/jhl.18019.kal
Journal of Historical Linguistics 9:1 (2019), pp. 167–176. issn 2210-2116 | e-issn 2210-2124

In the traditional understanding (see François 2014: 163–167 for more detailed discussion), a population speaking an internally-homogeneous language separates into two or more sub-populations. Each sub-population undergoes linguistic changes, then splits into sub-populations of its own, and so on recursively through time. Crucially, as long as a language community remains in existence – i.e., until it splits into two separate sub-populations – the traditional tree model assumes that every (viable) language change must always have spread to the *entire* population; in biological terms, every linguistic innovation is assumed to have "gone to fixation" before any further event of lineage splitting.[2] As a consequence, every innovation is assumed to be passed on to all of the descendants of the language in which it occurs – and *only* to those descendants. This leads to the useful result known as "Leskien's principle" (after Leskien 1876: vii, though the principle was first clearly stated by Brugmann 1884: 231): namely, that a subgroup of a language family – a set of languages that descend from a single intermediate proto-language – can only be identified by verifying whether those languages are defined by a set of *exclusively-shared innovations*. This principle of subgrouping by exclusively-shared innovations is an essential part of the Comparative Method underlying the practice of historical linguistics, and is also reflected in the assumptions of Bayesian gain-loss phylogenetic methods (see Greenhill & Gray 2012: 525–526).

However, it is often the case (both for languages and for species: see Baum & Smith 2013: 146–151) 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. Moreover, since different segments of the original population may have undergone different partially-diffused innovations, different innovations may be passed on to different subsets of descendants. This is the fundamental observation of the wave model: that innovations within a language (i.e. among closely related dialects) typically show overlapping patterns (Bloomfield 1933: 317; de Saussure 1995 [1916]: 273–278).

There are two ways to apply the insights of the wave model to the subgrouping of a language family:

- The first is to define a "subgroup" as "all of the languages descended from a single intermediate proto-language" and accept that some subgroups may not be definable in terms of exclusively-shared innovations, but only in terms of a chain or network of overlapping innovations.
- The second is to adhere to the definition of "subgroup" used in the traditional application of the tree model – namely, as a set of languages defined by

---

[2] As noted by Baum & Smith (2013: 79), this is often a reasonable assumption in biology, as the rate of fixation of novel traits is extremely rapid compared to the rate of lineage splitting.

> exclusively-shared innovations – and accept that subgroups (thus defined) may intersect rather than being strictly nested. This is the approach that we take in Kalyan & François (2018), where we explicitly reconcile the wave model with the Comparative Method.

We tend to prefer the latter, more easily operationalized definition of "subgroup," while J&L adhere to the former one (as does, e.g., Ross 1997 in his discussion of "innovation-defined" vs. "innovation-linked" subgroups). We see value in both approaches.

In both the traditional understanding of the tree model and the more nuanced understanding outlined above, it is taken for granted that languages diversify as a result of successive population splits. The main difference, then, is whether each population split necessarily leads to the emergence of a homogeneous daughter language (as assumed in the traditional family-tree model: see Ross 1997: 212) or whether the daughter language may exhibit dialectal variation from the very beginning due to arising from a segment of a dialect network. Thus, we do not contest J&L's statement that our approach "silently acknowledges … tree-like divergence … even if it turns out to be a star-phylogeny"; in fact, we wholeheartedly acknowledge it.

It is in fact possible to restate our views in terms of the framework proposed by J&L. If we allow for the partial diffusion of an innovation through a proto-language, then we are effectively allowing for variation in the proto-language that is differentially resolved in each descendant. This is nothing more than *incomplete lineage sorting*, but of a kind that J&L do not discuss: namely, one in which the variation in the proto-language is conditioned areally rather than morphologically or socially. This term introduced by J&L provides historical linguists with a useful, biologically-inspired way of talking about the phenomena highlighted by the wave model.

3. Methodology and goals of Historical Glottometry

In the remainder of this response, we address J&L's specific remarks on Historical Glottometry (in their Section 3.3), and attempt to clarify our position where we feel it has been misunderstood.

3.1 Definition and identification of "shared innovations"

J&L criticize our use of the term "shared innovation" for two reasons. Firstly, they object to our using the term agnostically for both "true" shared innovations as well as innovations that might turn out to be cases of parallel development, and suggest that our dataset should consist of only those innovations that are securely known to belong to the first category. Secondly, they claim that if we were to restrict our dataset in this way, we would not find any overlapping innovations: that shared innovations in a tree can never overlap, because each must occur in a single node.

We believe that it is necessary for us to be agnostic at the data-assembly stage about whether the innovations we identify are shared or parallel; the very purpose of the glottometric method is to *infer* which sets of innovations are likely to be "true" shared innovations (by virtue of following a consistent pattern) and which are likely to be parallel innovations (by virtue of following a pattern that is infrequently attested).[3] In other words, the assumption we make is that cases of parallel innovation will "come out in the wash" – and indeed, in our work, we have found that innovations that follow a geographically haphazard pattern (and are thus likely to be cases of parallel innovation) are invariably associated with subgroups that have low rates of cohesiveness and subgroupiness.

We agree that each (genealogically-relevant) innovation must occur at a single node in the tree – i.e., in a single ancestral speech community; however, this does not mean that innovations within this speech community cannot partially overlap. In fact, if we allow for the possibility of an innovation partially diffusing through a proto-language (as discussed in the previous section), then this is what we would often expect.

3.2 Reading diachrony in glottometric diagrams

J&L criticize glottometric diagrams on the grounds that they are "pure data display" and thus carry no diachronic information (other than the trivial information that all the lects displayed are descended from the same proto-language). While it is true that glottometric diagrams do not directly show a temporal dimension, we believe that, like trees, they encapsulate hypotheses regarding the relative chronology of lineage-splitting events.

This becomes apparent once we realize that a glottometric diagram not only summarizes the innovations that have taken place in a language family, but also represents the mutual intelligibility relations among the dialects of the proto-language: the more isoglosses there are that connect a group of dialects, and the more "subgroupy" these isoglosses are, the more mutually intelligible those dialects were.

Let us define a "language" as a set of dialects connected by links of mutual intelligibility and disconnected from other dialects. In a dialect continuum, what are initially mutually intelligible lects diversify progressively from each other by

---

[3] We admit that this point would be clearer if we referred to the innovations in our data as "*potential* shared innovations" whose status as "shared" or "parallel" would need to be determined by the degree to which they are supported by other potential shared innovations. We thank J&L for highlighting this potential source of confusion.

undergoing local innovations. For a while, the innovations taking place inside the continuum increase the difference across dialects yet do not jeopardize their overall mutual intelligibility. In spite of its emerging internal fragmentation, the language as a whole may remain alive for a period and still undergo its own innovations on a larger scale. In a glottometric diagram, this is shown by the coexistence of isoglosses of smaller scope (change affecting local dialects) with those of global scope (change affecting the language as a whole).

As changes accumulate over time, the links of mutual intelligibility among the dialects become weaker and ultimately disappear (with the weakest links disappearing first). In visual terms, this is equivalent to successively removing the weakest isoglosses from the glottometric diagram. As the weakest isoglosses are successively removed, the glottometric diagram "breaks apart" into disconnected sets of isoglosses; this is equivalent to the proto-language breaking up into mutually unintelligible daughter languages. This approach constitutes one possible way to formalize the notion of *linkage breaking* (Ross 1997: 222–228) – we can observe how a former dialect network progressively breaks into smaller dialect networks in a recursive manner until we reach the languages that are currently observable. This process is partly reminiscent of the diachrony that can be read in a tree diagram – as subgroups split successively into smaller subgroups – except that the initial stages of the evolution involve a dialect continuum showing intersecting isoglosses, a situation which is itself incompatible with a tree as traditionally conceived.

In more formal terms, we propose the following definitions:

i. A **glottometric diagram** is a *weighted hypergraph* (see J&L's footnote 14) whose nodes are dialects, whose edges are isoglosses (i.e., sets of dialects defined by one or more exclusively-shared innovations), and whose edge weights are the subgroupiness values of these isoglosses.
ii. A **language** is a *connected component* of such a weighted hypergraph – in other words, a set of dialects that are chained together by isoglosses and disconnected from any other dialects in the diagram.
iii. The **chronology** of a language family is found by successively removing the weakest edges from the hypergraph and at each stage noting how the dialects are partitioned into connected components (i.e., into languages).

We can illustrate these ideas using a glottometric diagram of the languages of North Vanuatu (Figure 1). The glottometric map in Figure 2 (from François 2017: 72) plots the same results onto a geographical map.[4]

---

[4] For details of the dataset and methodology used to produce these diagrams, see François (2014), Kalyan & François (2018). From left to right in Figure 1 (Northwest to Southeast in

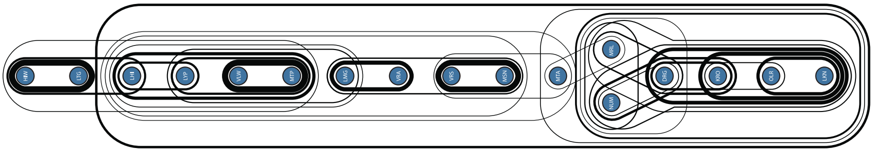


Figure 1. Glottometric diagram of the Oceanic languages of North Vanuatu

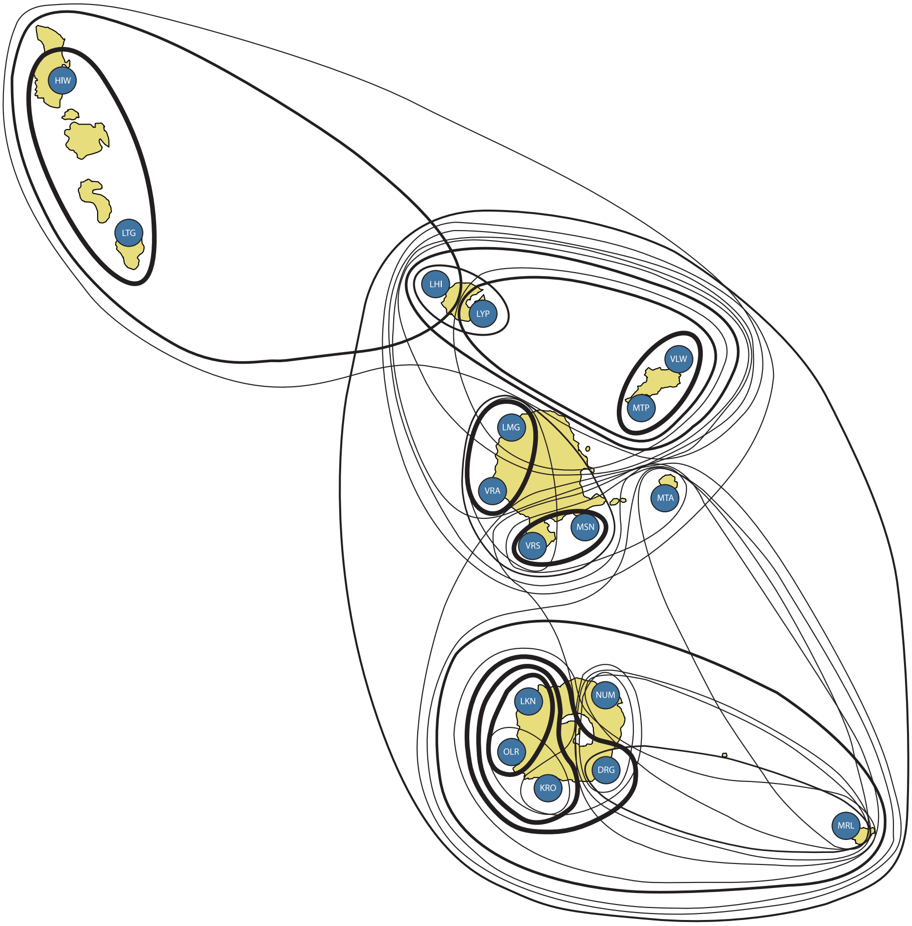


Figure 2. Glottometric map of the Oceanic languages of North Vanuatu

---

Figure 2), the language names read as follows: HIW Hiw, LTG Lo Toga, LHI Lehali, LYP Löyöp, VLW Volow, MTP Mwotlap, LMG Lemerig, VRA Vera'a, VRS Vurës, MSN Mwesen, MTA Mota, NUM Nume, DRG Dorig, KRO Koro, OLR Olrat, LKN Lakon, MRL Mwerlap.

As can easily be seen, the isoglosses covering these 17 languages form a single connected set, reflecting the fact that the dialects that they descend from were all initially mutually intelligible. Due to the strength of the isogloss connecting the three westernmost dialects ({HIW-LTG-LHI}), all of the dialects remain connected even if the 21 weakest isoglosses are removed from the diagram. But as soon as we remove the 22nd-weakest isogloss (namely, {HIW-LTG-LHI}), the glottometric diagram breaks apart, as shown in Figure 3.

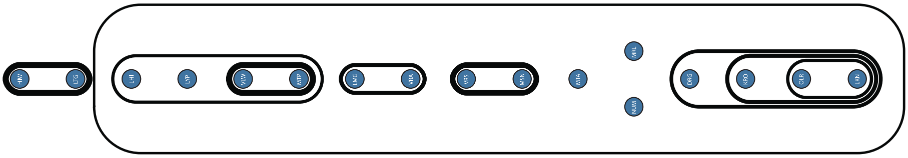


Figure 3. Glottometric diagram of the Oceanic languages of North Vanuatu, with the 22 weakest isoglosses removed

Here, the two westernmost dialects (the Torres group, {HIW-LTG}) have split off from the remainder (the Banks group). We interpret this as the first break in mutual intelligibility that occurred among the languages of North Vanuatu, resulting in a division of the original proto-language into two mutually-unintelligible intermediate proto-languages. Continuing in this manner, we drop the next weakest isogloss (i.e., the next-weakest link of mutual intelligibility), namely the one encompassing the Banks group. Figure 4 shows the situation that must have prevailed when that subgroup broke up, i.e., when the various dialects within it lost mutual intelligibility.

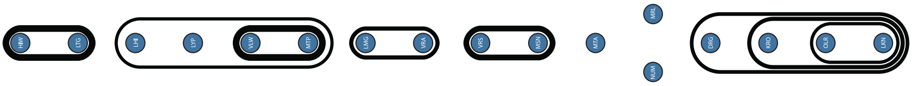


Figure 4. Glottometric diagram of the Oceanic languages of North Vanuatu, with the 23 weakest isoglosses removed

The result was then a set of eight languages, i.e. eight sets of dialects ({HIW-LTG}, {LHI-LYP-VLW-MTP}, {LMG-VRA}, {VRS-MSN}, {DRG-KRO-OLR-LKN}, {MTA}, {NUM}, and {MRL}), which, for some time, continued to evolve each as a single language community – as indicated by the amount of shared innovations characterizing each one. The process of *linkage breaking* illustrated here continued for several generations, eventually leading up to the different languages we know today.

In sum, Historical Glottometry does make it possible to infer a chronology of lineage-splitting events;[5] this becomes apparent once we realize that a glottometric diagram not only provides a synoptic overview of the innovations that occurred across the history of a language family, but also provides a map of the mutual-intelligibility relations among the dialects of the proto-language.

4. Conclusion

We hope to have shown that Historical Glottometry does not challenge the family tree model once *incomplete lineage sorting* has been taken into account. Our approach is actually meant as a critique of the tree model as traditionally understood, where an innovation must necessarily affect the *whole* of the population in which it occurs. The crucial observation of the wave model – that innovations frequently overlap – pertains to the developments that occur in a proto-language before it splits up (i.e., when it is still a network of mutually intelligible dialects). Thus, this observation is naturally captured within the framework of "incomplete lineage sorting" proposed by J&L; the only difference in our argument is that we allow for "variation" in the proto-language that is not only morphological or sociolinguistic, but also areal (dialectal). Finally, glottometric diagrams, far from being "pure data display," do in fact encode information about the order in which lineage splits most likely occurred.

This is not to say that we see no room for further improvement and refinement in Historical Glottometry. In particular, unlike standard approaches in computational phylogenetics, we do not currently have a generative model that could be used for estimating dates of innovations and population splits. One approach would be to apply existing models of incomplete lineage sorting (e.g. Pagel & Meade 2004; Wen, Yu & Nakhleh 2016) to linguistic data.[6] Another approach, which we are currently pursuing, is to directly develop a generative model of the spread of linguistic innovations in a network, adapting Madigan & York's (1995) work on Bayesian graphical models of discrete data. We believe that many promising avenues of research open up once we look beyond the restrictive assumptions of the family-tree model as traditionally understood – regardless of whether we draw our inspiration from the wave model of Schmidt (1872) and Schuchardt (1900) or from the insights of phylogenetic systematics (Maddison 1997; Galtier & Daubin 2008).

---

[5] This type of inference should be used with caution, however: the inferred chronology is dependent on the ranking of isoglosses by subgroupiness, which itself may be sensitive to small changes in the data (particularly for low subgroupiness values).

[6] See Verkerk (this issue) for an initial attempt in this direction.

Acknowledgements

This work contributes to the research program ”Investissements d’Avenir,” overseen by the French National Research Agency (ANR-10-LABX-0083): LabEx *Empirical Foundations of Linguistics*, Strand 3 – “Typology and dynamics of linguistic systems.”

References

Baum, David A. & Stacey D. Smith. 2013. *Tree Thinking: An Introduction to Phylogenetic Biology*. New York: Macmillan.

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

Brugmann, Karl. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen. *Internationale Zeitschrift für allgemeine Sprachwissenschaft* 1, 226–256.

Donohue, Mark. 2013. Who Inherits What, When? Toward a Theory of Contact, Substrates, and Superimposition Zones. *Language Typology and Historical Contingency: In Honor of Johanna Nichols* ed. by Balthasar Bickel, Lenore Grenoble, David A. Peterson & Alan Timberlake, 219–240. (= *Typological Studies in Language*, 104.) Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.104.09don

François, Alexandre. 2011. Social Ecology and Language History in the Northern Vanuatu Linkage: A Tale of Divergence and Convergence. *Journal of Historical Linguistics* 1:2.175–246. https://doi.org/10.1075/jhl.1.2.03fra

François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. *The Routledge Handbook of Historical Linguistics* ed. by Claire Bowern & Bethwyn Evans, 161–189. London: Routledge.

François, Alexandre. 2017. Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues. *Diffusion: implantation, affinités, convergence* ed. by Jean-Léo Léonard, 43–82. (= *Mémoires de la Société de Linguistique de Paris*, XXIV.) Louvain: Peeters.

Galtier, Nicolas & Vincent Daubin. 2008. Dealing with Incongruence in Phylogenomic Analyses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363:1512.4023–4029. https://doi.org/10.1098/rstb.2008.0144

Greenhill, Simon J. & Russell D. Gray. 2012. Basic Vocabulary and Bayesian Phylolinguistics: Issues of Understanding and Representation. *Diachronica* 29:4.523–537. https://doi.org/10.1075/dia.29.4.05gre

Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. *Let’s Talk About Trees: Genetic Relationships of Languages and Their Phylogenic Representation* ed. by Ritsuko Kikusawa & Lawrence A. Reid, 59–90. (= *Senri Ethnological Studies*, 98). Ōsaka: National Museum of Ethnology.

Leskien, August. 1876. *Die Declination im Slavisch-Litauischen und Germanischen*. Leipzig: S. Hirzel.

Maddison, Wayne P. 1997. Gene Trees in Species Trees. *Systematic Biology* 45:523–536. https://doi.org/10.1093/sysbio/46.3.523

Madigan, David & Jeremy York. 1995. Bayesian Graphical Models for Discrete Data. *International Statistical Review* 63:2.215–232.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226580593.001.0001

Pagel, Mark & Andrew Meade. 2004. A Phylogenetic Mixture Model for Detecting Pattern-Heterogeneity in Gene Sequence or Character-State Data. *Systematic Biology* 53:571–581. https://doi.org/10.1080/10635150490468675

Ross, Malcolm D. 1997. Social Networks and Kinds of Speech-Community Event. *Archaeology and Language 1: Theoretical and Methodological Orientations* ed. by Roger Blench & Matthew Spriggs, 209–261. London: Routledge.

de Saussure, Ferdinand. 1995 [1916]. *Cours de linguistique générale*. Paris: Éditions Payot & Rivages.

Schmidt, Johannes. 1872. *Die Verwantschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.

Schuchardt, Hugo. 1900. *Über die Klassifikation der romanischen Mundarten. Probe-Vorlesung, gehalten zu Leipzig am 30. April 1870*. Graz.

Wen, Dingqiao, Yun Yu & Luay Nakhleh. 2016. Bayesian Inference of Reticulate Phylogenies Under the Multispecies Network Coalescent. *PLOS Genetics* 12:5.e1006006. https://doi.org/10.1371/journal.pgen.1006006

Address for correspondence

Siva Kalyan
Department of Linguistics
School of Culture, History and Language
College of Asia and the Pacific
Australian National University
9 Fellows Road
Acton, ACT 2601
Australia
siva.kalyan@anu.edu.au

Co-author information

Alexandre François
CNRS-ENS-LaTTiCe
Australian National University
alexandre.francois@ens.fr

Journal of Historical Linguistics

VOLUME 9 NUMBER 1 2019

ISSN 2210-2116 / E-ISSN 2210-2124

John Benjamins Publishing Company