

INTRODUCTION

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

Siva Kalyan¹, Alexandre François^{1,2} and
Harald Hammarström³

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).

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.

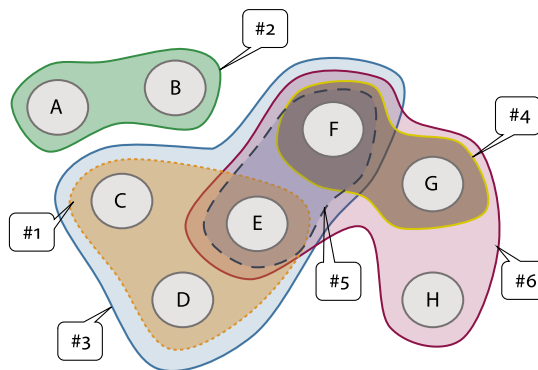


Figure 1. Intersecting isoglosses in a dialect continuum, or a “linkage”

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 (Bowerman 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](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 Phylogenetic 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. Schweschké & Sohn.
- Schleicher, August. 1873. *Die Darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Haeckel, o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena, Second Edition*. Weimar: Hermann Böhlau.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse 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