

Book Review

Frederik Hartmann. 2023. *Germanic phylogeny* (Oxford Studies in Diachronic and Historical Linguistics). Oxford: Oxford University Press, pp. 304. ISBN: 9780198872733.

Reviewed by **Chundra Cathcart**, Department of Comparative Language Science, Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland,
E-mail: chundra.cathcart@uzh.ch

<https://doi.org/10.1515/flin-2024-2048>

The Germanic subgroup of the Indo-European family is one of the most studied groups of languages, and among the best diachronically attested. At the same time, we have an incomplete understanding of several aspects of the Germanic languages' diversification and spread. *Germanic phylogeny* aims to fill various gaps in our understanding of the development of Germanic, bringing to bear quantitative and computational methods, both established as well as novel and ambitious, on outstanding questions regarding the higher-order subgrouping of the Germanic subgroup. In this book, quantitative modeling takes center stage, and I found *Germanic phylogeny* to be a highly thought provoking and invigorating contribution to the Bayesian modeling of linguistic diversification.

Bayesian statistics is a powerful and expressive probabilistic framework that allows practitioners to characterize the stochastic processes thought to generate the data we observe. Processes of this sort contain unobserved parameters which need to be inferred; the goal of Bayesian inference is to infer parameter values that are most compatible with the observed data (i.e., of highest posterior probability) in order to better understand the nature of the generative process. In some cases, the model defined by a generative process will have a tractable likelihood, in which case parameter values can be inferred using a well-understood method like Markov chain Monte Carlo (MCMC), an iterative process for estimating parameter values. If the model has a likelihood that is too costly to compute and evaluate over many iterations, then approximate Bayesian computation (ABC) can be used, which stores parameter values that generate simulated data that closely approximate properties (so-called summary statistics) of the observed data. The studies in the book make use of both types of model.

Chapter 1 is a short introduction to the current state of research on Germanic subgrouping and diversification with a brief mention of the opposition between the tree and wave model in historical linguistics. The extent to which the appearance of tree-like bifurcation between speech communities may obscure a number of processes has been discussed in reference to a number of subgroups in Indo-European and beyond (Garrett 2006; Ross 1988).

Chapter 2 gives a concise overview of the data used in the two studies presented in the book. The data employed build upon the catalog of features collected by Agee (2018), and comprise 479 innovations in eight pre-modern Germanic languages representing the oldest attested varieties of their respective lineage, two of which (Burgundian and Vandalic) are fragmentary and display a high degree of missing data. The binary characters or features in the dossier are based on innovative phonological, morphological, syntactic, and lexical changes, rather than synchronic patterns found in a language (e.g., word order, etc.). The individual features are given in a table in the Appendix, and are often telegraphically described, so one must consult Agee (2018) for a more in-depth description. These features (“change pattern[s] in the structure of a language”, p. 10) are apparently coded in a way such that innovations are treated as reversible, and denote correspondences with an earlier Proto-Germanic feature, rather than flagging the operation of changes during the history of a language (which may be obscured by later changes). As an example (p. 51), if a lineage undergoes the sound change trajectory $*o > *au > a$, the second change is seen as involving the “loss” of an innovation involving vowel breaking, and innovation loss does not necessarily correspond to change reversal per se (which should be impossible in the case of phonological mergers, etc.).

Chapter 3 introduces the reader to cladistics and phylogenetic methods, proceeding from simpler, more intuitive distance-based methods to Bayesian methods which assume that features change according to a substitution model. Missing is a discussion of the parsimony framework, which constructs phylogenetic topologies which minimize the number of parallel innovations occurring over lineages (see Felsenstein 2004: 97–121 for an assessment). This study (using data consisting of innovations and including only the oldest representatives of Germanic lineages) shares commonalities with other work using parsimony and related frameworks (Canby 2024; Nakhleh et al. 2005; Skelton 2014), so this omission is surprising.

The remainder of the chapter presents a study carrying out phylogenetic inference with RevBayes (Höhna et al. 2016) using the previously described data. These data are a non-standard choice for the method, since Bayesian phylogenetic inference usually makes use of data from multiple chronological stages in order to estimate chronologies with precision. Furthermore, linguistic phylogenies are generally inferred on the basis of lexical root-meaning traits, which code whether a language expresses a given concept using a word from a particular cognate class, usually for several hundred concepts. For lexical data, the model of character evolution represents a stochastic process whereby a word in a meaning function is expected to be replaced over some interval of time (conceivably stemming from usage, etc.), and is famously reversible and homoplastic, involving parallelism (Chang et al. 2015), whereas the data analyzed here are arguably less prone to

homoplasy and reversibility. The best fitting of six models under comparison does not reveal strong support for most Germanic subgroups. The lack of resolution in the resulting phylogenies is not surprising – this is the case for Germanic even when more data are used, and this study uses a data set comprising only eight languages (phylogenetic studies of small families and subgroups have at least twice this amount; Goldstein 2024; Honkola et al. 2013; Kolipakam et al. 2018). Following these inconclusive results, the author directs us to alternative strategies for clarifying the dynamics of early Germanic diversification.

Chapter 4 is the longest in the book and introduces and motivates the book's major contribution, a Bayesian agent-based approach to modeling the development and spread of Germanic via interactions between idealized populations that undergo innovation and migration according to various stochastic processes. Phylogenetic models are phenomenological, characterizing the large-scale dynamics of change; deriving these dynamics from individual-level interactions is challenging (cf. Burridge and Vaux 2020). Agent-based models on the other hand explicitly model replicator dynamics at the individual level. In most cases, agent-based models test different change scenarios by varying a small number of parameters and observing how this impacts long-term outcomes of change (cf. Round et al. 2021). Here, a single model is constructed, and the results are used to address outstanding questions in the field.

In this setting, “agents” (representing small speech communities) occupy different “tiles” (representing geographic locations) on a simulation surface. At different timepoints (“ticks”) over the course of a model run, agents can spawn new agents (representing linguistic spread), actuate innovations as well as adopt and undo them depending on the distribution of linguistic features in their immediate neighborhood, migrate to adjacent tiles, and die out (removal of agents from tiles represents the encroachment of new speech communities upon tiles previously occupied by Germanic speech communities). At a high level, during a run, agents are initialized in the Germanic homeland (Figure 4.24, p. 109) and then migrate within the early Germanic territory, undergoing, spreading, and losing innovations. We can make inferences about an individual language's history by observing the behavior of agents in the geographical region where the language was spoken. Results of this procedure appear to show an early linguistic divergence of East Germanic followed by a long duration of linguistic similarity between North and West Germanic varieties, indicative of a dialect continuum.

Of course, a single simulation run will likely not produce a historical trajectory that results in patterns resembling the synchronic reality of a given language in the region where it was spoken. For this reason, the simulation is run a large number of times, and only runs where the simulated outcome comes close to approximating the linguistic data observed in different language locations are taken into account, a

technique adopted from ABC. Given the randomness involved, it is quite difficult to produce realistic results via simulations, so methods from machine learning are used in order to home in on regions of high-dimensional parameter space that produce more realistic results.

I consider myself fairly conversant in Bayesian modeling conventions, but nonetheless, midway through this chapter, things began to unravel for me, and likely will for even more diligent readers. The model presented is highly complex with many moving parts, and the author strives commendably to make its inner workings accessible to an audience with no prior exposure to such models. But despite this care, it is virtually impossible to understand all of the technical aspects of the model (such as to replicate it), and I found myself unable to fully critically evaluate the chapter's findings, e.g., to assess whether certain properties of model behavior were artifacts of design. As in most technical work, there are notational ambiguities. As an example, the equations on pp. 118–119 for the loss function used to characterize the goodness of fit of the simulated data to the observed data contain undefined variables. Certain unusual conventions, such as the use of a truncated Normal distribution to model probabilities rather than the more conventional Beta distribution, go unexplained. There is no discussion of whether the posterior distributions of parameters converge across runs (e.g., Gelman and Rubin 1992); certain posterior distributions are multimodal (e.g., Figures 4.45, 4.47), and it would be helpful to know how this multimodality impacts our interpretation of the results.

The author makes use of plate notation (Figures 4.32–4.34, 4.44) to graphically represent relationships between parameters and variables in the model. While this notation is aesthetically appealing, it can be unintuitive (Zinkov 2013), and should be accompanied by a prose description of the data generation process which defines variables and their distributions, and ideally contains a non-technical description of the phenomenon that each step of the process is intended to model (cf. Cathcart 2020: 60–61, 2022: 26). I was thrown off by some conventions; for instance, the diagram in Figure 4.32 contains rectangular “plates,” which usually represent repetitions or loops in the generative process, but closer inspection revealed that their purpose was simply to thematically partition the model into different modules. It was often not possible to determine the exact relationship between parameters in the model. As an example (pp. 123–124), it is clear that I_A (the innovation module) depends on the homoplasmy rate h and the occurrence time t_i , but the function through which these entities are related was not given. Two additions would have been incredibly helpful: (1) a table defining each parameter and variable mathematically and in plain language, and (2) pseudocode for a run of the simulation. As it stands, there is not enough information in this already quite detailed chapter in order to replicate its results. Furthermore, the book does not include a link to a repository with code, and at the time of writing, the author's GitHub page does not contain such a repository.

Code sharing (even of messy, unannotated code) is essential to making models like the one presented here transparent and reproducible (Barnes 2010), and is all the more important for highly elaborate models.

Chapter 5 reassesses views of the breakup of Germanic in light of the model's results. Chapter 6 sketches new directions for agent-based models of this sort, as well as additional applications. The author makes the important point that different historical linguistic scenarios call for different quantitative models. The prospect of linguists designing bespoke models to analyze the development of the linguistic groups in which they specialize is an exciting one indeed.

The points of criticism that I have made above should not be seen as detracting from my overall enthusiasm for what I found a highly engaging and inspiring book. I came away from reading *Germanic phylogeny* with renewed optimism regarding the issues that can be tackled by adopting the philosophy espoused in the book. Making methods of this sort accessible to a general audience is highly challenging, and the author's lucid exposition does this admirably. Quantitative and computational methods are increasingly important in diachronic linguistics, and it is clear that standard off-the-shelf tools are not appropriate for every phenomenon under study. This book provides a clear view of the benefits and challenges of designing novel approaches to modeling the highly multifaceted phenomenon that is linguistic evolution.

Acknowledgment: I am grateful to Muhammad Rehan and Steven Moran for helpful discussion. All errors and infelicities are my own responsibility.

References

- Agee, Joshua. 2018. *A glottometric subgrouping of the Germanic languages*. San Jose: San Jose State University MA thesis.
- Barnes, Nick. 2010. Publish your computer code: It is good enough. *Nature* 467. 753.
- Burridge, James & Bert Vaux. 2020. Brownian dynamics for the vowel sounds of human language. *Physical Review Research* 2. 013274.
- Canby, Marc. 2024. LinguiPhyR: A package for linguistic phylogenetic analysis in R. *Journal of Open Source Software* 9(101). 6201.
- Cathcart, Chundra. 2020. A probabilistic assessment of the Indo-Aryan Inner–Outer Hypothesis. *Journal of Historical Linguistics* 10. 42–86.
- Cathcart, Chundra. 2022. Dialectal layers in West Iranian: A hierarchical dirichlet process approach to linguistic relationships. *Transactions of the Philological Society* 120. 1–31.
- Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1). 194–244.
- Felsenstein, Joseph. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.

- Garrett, Andrew. 2006. Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 139–151. Cambridge: McDonald Institute Monographs.
- Gelman, Andrew & Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7. 457–511.
- Goldstein, David. 2024. Divergence-time estimation in Indo-European: The case of Latin. *Diachronica* 41. 1–45.
- Höhna, Sebastian, Michael J. Landis, Tracy A. Heath, Bastien Boussau, Nicolas Lartillot, Brian R. Moore, John P. Huelsenbeck & Fredrik Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65. 726–736.
- Honkola, Terhi, Outi Vesakoski, Kalle Korhonen, Jyri Lehtinen, Kaj Syrjänen & Niklas Wahlberg. 2013. Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* 26. 1244–1253.
- Kolipakam, Vishnupriya, Fiona M. Jordan, Michael Dunn, Simon J. Greenhill, Remco Bouckaert, Russell D. Gray & Annemarie Verkerk. 2018. A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science* 5. 171504.
- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81. 382–420.
- Ross, Malcolm D. 1988. *Proto Oceanic and the Austronesian languages of Western Melanesia*. Canberra: Pacific Linguistics.
- Round, Erich R., Sacha Beniamine & Louise Esher. 2021. The role of attraction-repulsion dynamics in simulating the emergence of inflectional class systems. <https://doi.org/10.48550/arXiv.2111.08465> (accessed 27 June 2022).
- Skelton, Christina. 2014. *A new computational approach to the Ancient Greek dialects: Phylogenetic systematics*. Los Angeles: University of California PhD thesis.
- Zinkov, Rob. 2013. Stop using plate notation. <https://www.zinkov.com/posts/2013-07-28-stop-using-plates> (accessed 10 April 2024).