Phylogenetics and morphological reconstruction: A case study with clusivity and Pama-Nyungan (Australia)

Computational phylogenetics is the branch of diachronic linguistics that deals with trees reconstructed through evolutionary models and the inferences that can be made about language change using such trees (see e.g. Atkinson & Gray 2005; Bowern 2018). Over the last 20 years or so, trees created with phylogenetic methods have made it possible to study everything from lexical replacement to syntactic change and phonology (Greenhill 2023; Greenhill et al. 2017; Macklin-Cordes, Bowern & Round 2020; Jordan 2013). They have allowed us to make progress towards understanding population migrations, how languages split, and more generally how trees and networks are formed (Atkinson & Gray 2005; Atkinson et al. 2008; Bryant, Filimon & Gray 2005).

One line of work involves creating trees and networks (Bouckaert, Bowern & Atkinson 2018; Grollemund et al. 2015; Greenhill, Blust & Gray 2008; Gray, Drummond & Greenhill 2009; Gray & Jordan 2000; Kitchen et al. 2009; Chang et al. 2015), while another uses those trees to study the evolution of particular features and systems, both linguistic (Phillips & Bowern 2022; Skirgård 2024; Goldstein 2024) and cultural (Silva & Tehrani 2016; Jordan et al. 2009; Gavin et al. 2018). To date, much of that work involves lexical material (e.g. Gerardi et al. 2023), though there is also growing work on reconstruction in morphology and phonology (Macklin-Cordes, Bowern & Round 2020). In this talk, I think through the issues for reconstruction in morphology using phylogenetics, both for morphological theory and for investigating change, using clusivity marking (Bickel & Nichols 2005; Filimonova 2005) as an example.

Lexical data has characteristics that make it a good choice for phylogenetic trees. Rates of change in the lexicon vary greatly, meaning that it's possible to reconstruct branching through the family (conservative/slow changing material providing evidence for early branching, with more rapidly changing areas of the lexicon providing better information about more recent branching). Some aspects of the lexicon are highly borrowable, while others diffuse more rarely, providing ways to gauge the strength of tree structure (vs network-like evolution). Hartmann and Walkden (2024) show that syntactic data has, low strength of phylogenetic signal in comparison to phonology.

Where does this leave morphology? Morphology, of course, has a phonological, a syntactic, a semantic, and a lexical component. Morphemes occur in paradigms, allowing and requiring

us to study sets of items. This makes morphology a rich area for phylogenetics (as indeed, for tree-building in other ways, as has long been known); but it also brings up the weaknesses of phylogenetic methods for language. Because morphological characters are correlated (ie, occur in interlocking paradigms), they violate some of the assumptions needed for characters used in tree building. Because they can change semantics but not form (or vice versa), it can be difficult to diagnose changes in isolation. For example, the string sequence *inkama* is a licit verb in four Nyulnyulan languages (Bowern 2012); in each the morphemic structure *i-n-kama-ø* can be identified, but it doesn't mean the same thing in each language. Linguistic phylogenetic coding typically uses discrete data points, meaning that we cannot reconstruct values that are outside the attested data points.

In this talk, I explore some of these points with respect to clusivity marking in Pama-Nyungan, the language family that contains roughly three quarters of Australia's Indigenous languages (Bowern 2023). Clusivity as a category is, per Wichmann and Holman (2009) and Nichols (2003), stable; but Pama-Nyungan languages show varied marking of inclusive/exclusive distinctions, implying that clusivity has been subject to renewal in the family. While we can reconstruct clusivity to Proto-Pama-Nyungan based on both categorial reconstructions and the forms of dual and plural pronouns, the pronouns exhibit a range of changes (expansion of inclusive > exclusive and vice versa; changes in number reinforcing clusivity distinctions, and semantic change leading to suppletive first person pronoun marking).

References

Atkinson, Quentin D & R. D. Gray. 2005. Curious parallels and curious connections–phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54(4). 513–26.

Atkinson, Quentin D, Andrew Meade, C. Venditti, S. J. Greenhill & Mark Pagel. 2008. Languages evolve in punctuational bursts. *Science* 319(5863). 588–588.

Bickel, Balthasar & Johanna Nichols. 2005. Inclusive-exclusive as person vs number categories worldwide. In Elena Filimonova (ed.), *Clusivity: Typology and Case Studies of Inclusive-exclusive Distinction*, 49–72. Amsterdam: John Benjamins Publishing.

Bouckaert, Remco R., Claire Bowern & Quentin D. Atkinson. 2018. The origin and expansion of Pama–Nyungan languages across Australia. *Nature Ecology & Evolution*. Nature Publishing Group 2(4). 741–749. https://doi.org/10/cmg9.

Bowern, Claire. 2012. Nyikina paradigms and refunctionalization: A cautionary tale in morphological reconstruction. *Journal of Historical Linguistics* 2(1). 7–24. https://doi.org/10/ggdchr.

Bowern, Claire. 2018. Computational Phylogenetics. *Annual Review of Linguistics* 4(1). 281–296. https://doi.org/10/ggdchp.

Bowern, Claire (ed.). 2023. *The Oxford Guide to Australian Languages* (Oxford Guides to the World's Languages). Oxford, New York: Oxford University Press.

Bryant, David, F. Filimon & Russell Gray. 2005. Untangling our past: Languages, Trees, Splits and Networks. In R. Mace, C. Holden & S. Shennan (eds.), *The Evolution of Cultural Diversity: Phylogenetic Approaches*, 69–85. UCL Press.

Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1). 194–244.

Filimonova, Elena. 2005. *Clusivity: Typology and Case Studies of Inclusive-exclusive Distinction.* John Benjamins Publishing.

Gavin, Michael C., Patrick H. Kavanagh, Hannah J. Haynie, Claire Bowern, Carol R. Ember, Russell D. Gray, Fiona M. Jordan, Kathryn R. Kirby, Geoff Kushnick & Bobbi S. Low. 2018. The global geography of human subsistence. *Royal Society open science* 5(9). 171897. https://doi.org/10/gffh6v.

Gerardi, Fabrício Ferraz, Tiago Tresoldi, Carolina Coelho Aragon, Stanislav Reichert, Jonas Gregorio de Souza & Francisco Silva Noelli. 2023. Lexical phylogenetics of the Tupí-Guaraní family: Language, archaeology, and the problem of chronology. *PLOS ONE.* Public Library of Science 18(6). e0272226. https://doi.org/10.1371/journal.pone.0272226.

Goldstein, David. 2024. Divergence-time estimation in Indo-European: The case of Latin. *Diachronica* 41(1). 1–45. https://doi.org/10.1075/dia.22031.gol.

Gray, R. D., A. J. Drummond & S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913). 479–483.

Gray, R. D. & F. M Jordan. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405(6790). 1052–1055.

Greenhill, S. J., R. Blust & R. D. Gray. 2008. The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evolutionary bioinformatics online* 4. 271.

Greenhill, Simon J. 2023. Language Phylogenies: Modelling the Evolution of Language. In Jamshid J. Tehrani, Jeremy Kendal & Rachel Kendal (eds.), *The Oxford Handbook of Cultural Evolution.* Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198869252.013.61.

Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson & Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences* 114(42). E8822–E8829. https://doi.org/10/gchf46.

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.

Hartmann, Frederik & George Walkden. 2024. The strength of the phylogenetic signal in syntactic data. *Glossa: a journal of general linguistics*. Open Library of Humanities 9(1). https://doi.org/10.16995/glossa.10598.

Jordan, F. M. 2013. Comparative phylogenetic methods and the study of pattern and process in kinship. *Kinship systems: Change and reconstruction* 43–58.

Jordan, F. M, R. D. Gray, S. J Greenhill & R. Mace. 2009. Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B.*

Kitchen, A., C. Ehret, S. Assefa & C. J. Mulligan. 2009. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B: Biological Sciences* 276(1668). 2703–2710. https://doi.org/10.1098/rspb.2009.0408.

Macklin-Cordes, Jayden L., Claire Bowern & Erich R. Round. 2020. Phylogenetic signal in phonotactics. *arXiv:2002.00527 [cs, q-bio]*. http://arxiv.org/abs/2002.00527. (1 March, 2020).

Nichols, Johanna. 2003. Diversity and stability in languages. In Brian Joseph & Richard Janda (eds.), *The Handbook of Historical Linguistics*, 283–310. Oxford: Oxford University Press.

Phillips, Joshua & Claire Bowern. 2022. Bayesian methods for ancestral state reconstruction in morphosyntax: Exploring the history of argument marking strategies in a large language family. *Journal of Language Evolution* 7(1). 1–15. https://doi.org/10.1093/jole/lzac002.

Silva, Sara Graça da & Jamshid J. Tehrani. 2016. Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society Open Science* 3(1). 150645. https://doi.org/10.1098/rsos.150645.

Skirgård, Hedvig. 2024. Disentangling Ancestral State Reconstruction in historical linguistics: Comparing classic approaches and new methods using Oceanic grammar. *Diachronica* 41(1). 46–98. https://doi.org/10.1075/dia.22022.ski.

Wichmann, S. & E. W. Holman. 2009. Assessing temporal stability for linguistic typological features. *München: LINCOM Europa.* http://email.eva.mpg.de/~wichmann/WichmannHolmanIniSubmit.pdf.