Issues of Lexical Cognacy Encoding in IE-CoR

27th International Conference on Historical Linguistics

Tom Francis, Klaus Kim, Laura McLean, Muhammad Rehan
August 21, 2025

Introduction

- The recent study of Heggarty et al. (2023) is the most substantial investigation of Indo-European phylogeny to date.
- Their study used the newly-curated IE-CoR dataset that included 161 taxa (52 historical languages) and 170 meaning concepts (Heggarty, Anderson, and Bibiko 2023).
- The results engendered skepticism and disavowal from Indo-Europeanists (Kassian and Starostin 2025; Kroonen, Olander, Nørtoft, et al. 2023).
- Since data encoding can have severe repercussions on divergence-time estimates and the inference of tree topology (cf. Pronk 2011:144–5, Pereltsvaig and Lewis 2015:92–113; Rama et al. 2018, Heggarty 2021:376, Ringe 2022; Rönchen et al. 2024:7 with further references), we investigate the accuracy of lexical cognacy encoding for the ancient languages of IE-CoR.

Introduction

- Our investigation differs crucially from Kassian and Starostin (2025), who only verified the cognacy encoding of a select 100 items for Hittite and Gothic among the ancient Indo-European languages.
- We cast a broader net and systematically investigate the cognacy encoding in the Anatolian languages, Vedic Sanskrit, (Attic/New Testament) Greek, Italic languages, and Celtic languages (namely Old/Middle Welsh).
- Our investigation reveals systematic issues in character-state coding despite the development of robust criteria (Scarborough 2020) and the use of over 80 language experts for the assignment of character states.
- We highlight these issues by providing examples from ancient Indo-European languages that are clear from more rigorous philological attention.
- We argue that due to the issues of encoding lexical cognacy, lexical characters should not be solely relied upon for phylogenetic inference.

Roadmap

1. Background
2. Data & Discussion
3. Conclusions
4. Appendix

PIE Homeland and Divergence-time Estimation

- The two most common hypotheses for dating the divergence of Anatolian from the rest of the Indo-European tree are the Anatolian hypothesis (ca. 9,500-8,500 BP) and the Steppe hypothesis (ca. 6,500-5,500 BP; Atkinson & Gray 2006:91-92; Chang et al. 2015:195).
- Computational analyses based on prior lexical databases have produced conflicting results:
  - Atkinson and Gray (2006) used a modified version of the IELex data for their Bayesian analysis, which they argued supports the Anatolian hypothesis.
  - Chang et al. (2015) used a modified version of the IELex data and added ancestry constraints to their analysis, arguing that it supports the Steppe hypothesis.
- Heggarty et al.'s (2023a & b) model based on their IE-CoR dataset suggests a "hybrid model" in which Anatolian branches off ca. 8,120 BP, and there is another major division in the Indo-European tree roughly coinciding with the date assigned to the Steppe hypothesis.

Characters for Phylogenetic Inference

- Morphological characters have been thought to be most robust for phylogenetic inference by Indo-Europeanists (Ringe, Warnow, and Taylor 2002:65-70; Nakhleh, Ringe, and Warnow 2005:395–6; Hartmann 2023; Clackson 2007:6; Kim 2024:23[5])
    - Heggarty et al. (2023a:85-86) note that Ringe et al. supplemented his data with lexical characters but do not consider supplementing their own data with any other characters, likely because of the model complexity required for multi-character analysis.
- Phonological characters have also been used in various phylogenetic studies, mostly concerned with inferring rates of change rather than tree topology and divergence-time (Dockum 2018; Moran and Verkerk 2018; Moran, Grossman, and Verkerk)
- Syntactic and morphosyntactic characters (McMahon and McMahon 2005; Longobardi and Guardiano 2009, 2017; Ceolin et al. 2020; Creanga et al. 2024)
- Structural features (Nichols 1996, 2014; Dunn et al. 2005, 2008; Rexová, Bastin, and Frynta 2006; Wichmann and Saunders 2007; Greenhill et al. 2011, 2017)
- Phonotactics (Macklin-Cordes, Bowern, and Round 2021) and sound changes in cognate lists (character-based: Goldstein et al. 2025; distance-based: List et al. 2018)
- Paradigm complexity metrics (Herce and Bickel 2025)

Issues of Data: Sources

- We reviewed the general documentation for historical languages (Heggarty et al. 2023a:43-49) and the notes for each lexeme for our chosen historical languages.
- In our review, we found numerous philological and cognate errors.
- Many philological errors likely result from the exclusive reliance on decontextualized dictionaries for many historical languages (Hittite, Oscan, Umbrian, Latin, Greek: Ancient, Greek: New Testament, etc.) and a lack of any cited secondary sources and even primary texts in many cases (e.g., much of Old Welsh, Greek: Ancient, etc.).

Issues of Data: Greek and Latin Dictionaries

- For Greek: Ancient and Greek: New Testament, only the 1940 version of the LSJ dictionary from perseus.tufts.edu is cited on the IE-CoR website:
  - The 1996 supplement including extensive revisions is not included on this website.
  - The LSJ does not specialize in New Testament or early Judeo-Christian vocabulary (cf. Newman 2010, Lampe 1961, etc.) .
- For Latin, only the 1879 version of Lewis and Short's *Latin Dictionary* is included:
  - Although the *Thesaurus Linguae Latinae* (TLL) is still in progress, it includes many lexemes and provides a much more robust record of textual sources and meanings (suggesting changes for concepts like BLACK).

Issues of Defining Chronological and Geographic Language Categories: Greek

- Heggarty et al. (2023a:25-26) are explicit that their primary goal is to "estimate the time-depth of the first expansion and divergence of the Indo-European language family," focusing more on a deep phylogenetic signal rather than dialectal divergence.
    - Although relaxed in several noted instances, e.g., Norwegian (Bokmål and Nynorsk) and Nuristani languages, IE-CoR has a general rule requiring a greater than 4% difference in cognacy between dialects for both dialects to be included (2023a:25-26).
    - This necessarily entails a choice about which dialect is considered 'standard' or most important to the analysis (Milroy 2001). These sociolingulistic assumptions cannot be computationally tested since only one dialect is included.

The ancient Greek dialects in IE-CoR

- In the case of Greek, IE-CoR includes three historical Greek dialects: Greek: Mycenaean, Greek: Ancient, and Greek: New Testament.
- On IE-CoR's website, Greek: Ancient is defined as "Classical Attic," located in the "site of Classical Athens, the area where the Attic dialect of Ancient Greek originated."
    - Heggarty et al. argue that other ancient dialects of Greek do not differ enough lexically to meet IE-CoR's sampling threshold for inclusion and that "Ancient (Attic) Greek is at least very near to the direct ancestor to modern Greek," although it was "partially disrupted by the dialect-merging (*koiné)* phase during the Hellenistic period" (2023a:10; 43) .
- This explains their decision to include Attic Greek, but no explanation is provided for the inclusion of the narrow post-Hellenistic Greek: New Testament category ("based on a specific corpus, mostly the New Testament itself;" Heggarty et al. 2023a:44), or the exclusion of the vast corpus of Hellenistic Greek and Greek contemporaneous with New Testament Greek and ranging until the modern dialects.

Chronological and Geographic Language Categories: Greek

- Although Greek: New Testament is defined with some leeway in the supplemental materials (Heggarty et al. 2023a:44), their website defines the category as "strictly confined to the variety of Roman-period koiné that is attested in the New Testament corpus. Other varieties of Jewish koiné Greek (e.g. in the Septuagint translation of the Hebrew scriptures) are not included."
- Although the Septuagint is centuries older than the New Testament, the apparent omission of contemporary Jewish koiné authors causes some meanings that are otherwise unattested in the limited New Testament corpus to be omitted.
- For example, φθείρ 'louse' is attested in both Josephus (*Antiquitates Judaicae* II.300.2) and Philo of Alexandria (*De Providentia* fr. II.59.6 and *De Ebrietate* fr. V.12).
- The same φθείρ is listed for LOUSE for Greek: Ancient. Omitting 'louse' for Greek: New Testament removes an additional point of agreement between Greek: Ancient and Greek: New Testament, and it also removes a point of agreement between Greek: New Testament and the next attested stage of Greek in IE-CoR, the seven dialects of modern Greek that are included (cf. Greek: Modern Std ψείρα).

Hapaxes in the Historical Data: Oscan & Old Welsh

- One concern of collecting historical data that is not discussed in IE-CoR's supplementary materials is the prevalence of hapaxes (words attested only once) in the extant data. In some languages, hapaxes comprise a large portion of the lexical items in IE-CoR.
- For example, 13/34 meanings for Oscan have hapax lexemes and at least 14/69 meanings for Old Welsh have hapax lexemes.
- In both Oscan and Old Welsh, hapaxes present interpretative problems. Oscan hapaxes often do not meet the target parameters IE-CoR has provided for its meanings, and Old Welsh hapaxes are dependent on a mostly Latin text with Old Welsh marginalia and place names.

Hapaxes in the Historical Data: Oscan

- IE-CoR requires a language to have evidence for at least 35/170 meanings to be included in the dataset (2023a: 27).
- The Italic language Oscan, which has 35 words for 34 meanings (**iúkleí** and *zicolom* are listed as synonyms for DAY), is on the border of inclusion.
- In the case of Oscan hapax **pedú**, listed for the meaning FOOT, there is disagreement about its meaning, but none of its possible meanings align with IE-CoR's target parameters, including:
  - "The most generic noun for the foot as part of the human body"
  - "Avoid terms specific to the (size or length of a) 'foot' as a unit of measurement"

Interpretation of Oscan **pedú**

- Untermann (2000:523), the sole source cited for Oscan in IE-CoR, defines **pedú** as " 'Fuß' als längenmaß" ('Foot as unit of measurement').
- Franchi De Bellis (1988:116-118), citing a lacunose passage from Festus, interprets the word as equivalent to Latin *pali* 'stakes/fence' (cf. Untermann 2000:523).
- Recent translations from both McDonald (2022:127) and *Imagines Italicae* (2011:891 Abella 1; produced below) demonstrate the unlikelihood of **pedú** meaning 'foot' in the generic sense:
- "But between the Abellan and Nolan *slaags*, the surrounding road is all around (ablative **súllad**) of 10 feet (in width) [= **pedú X**]. At the mid-point of that road boundary markers stand."

Hapaxes in the Historical Data: Old Welsh

- The 14 hapaxes in the Old Welsh data are from the Llan Dav (*Liber Landavensis*); issues of this primarily Latin text ranging from dating to textual transmission to historical reliability have long been documented (cf. Davies 1973).
- Even Alexander Falileyev, the expert who compiled the Old Welsh lexical data for IE-CoR, excluded the Llan Dav data in his *Etymological Glossary of Old Welsh*, writing (2000: X): "The only text which was not used for the compilation of the present Glossary is the Book of Llan Dav, which still requires to be comprehensively discussed, and is a subject for research in its own right."
- Although Falileyev is clearly acquainted with these issues, IE-CoR makes no mention of the dubious nature of the source which leads to overconfidence in the Celtic phylogeny (cf. Heggarty et al. 2023a: 47, where only the date of Old Welsh is discussed).

Transcription and cognate-class assignment: Anatolian

- Because Kassian and Starostin (2025) discuss the Hittite entries, we focus here on Luwian.
- For the meaning concept EAR, "tummān(t)-"for Luwian and assigned to the cognate set of *$steh_3$-.
- In the notes, it is also mentioned that the Luwian word is cognate with Gk. στόμα 'mouth' (n.) and Av. *staman*- 'maw' (m.). There are a number of issues, however, with this entry.
    - Gk. στόμα is surely cognate with Luw. "tummān(t)-", but the evidence of Av. *staman*- 'maw' is more uncertain (cf. Melchert 2007/8:185; Tucker 2022:259 fn.7).
    - The word is only attested in a compound $^{(KÁ)}$*āš-tummant*- 'gate(way)', and -*tummant*- is better understood as 'split, aperture' (Melchert 2010: 56; 2024 s.v. $^{(KÁ)}$*āštummant*-).

Further problems with the encoding of CLuw. "*tummān*(*t*)-"

- A root *$steh_3$- cannot regularly lead to CLuw. "*tummān*(*t*)-" (cf. Vine 2019:229), which is, moreover, an incorrect reading for -*tu*[*m*)*man*(*t*)- (cf. Melchert 2007/8: 184[5]).
- Vine (2019:228) based on Wennerberg (1972) has convincingly argued that the Luwian word -*tummant*- and its cognates in other Indo-European languages reflect *(*s*)*temh*$_1$- (with s-mobile; cf. also Melchert 2007/8: 184-6), and thus the Luwian word should be assigned the cognate class *(*s*)*temh*$_1$- and not *$steh_3$-.
- Similarly, the entry for Hittite should also be modified to reflect the correct etymology for "*ištāman- / ištāmin-*" 'ear' (correctly *ištama*(*n*)-/*ištamin*- 'ear')

Target sense mismatches: Anatolian

- For Luwian, the "CHEST" word is encoded as CLuw. "tītan-", but the encoding of *tītan* as the word for "CHEST" violates the target sense outlined for this meaning concept:
    - "Provide the general term for this part of the body, that can be applied to both males and females, both children and adults. Certainly do not select a lexeme that refers only, or predominantly, to the female breasts. The English lexeme for this target sense is therefore *chest* (not *breast* — see also below), in French *poitrine* (not *sein*), and in Spanish *pecho* (not *seno*). Only if a language has no common term, select the word for *male* chest."
    - However, in its attestations, *tītan-* refers only to the female breast (cf. Melchert 2024, s.v. *tītan* and the translation 'breast, teat'; see the discussion in *EDIANA*, s.v; de Vaan 2019) and should not be selected by IE-CoR's own criteria.

Target sense mismatches: Anatolian

- In the IE-CoR target sense for the meaning concept "Sun", it is specifically mentioned, "Enter the literal term for the celestial body. Avoid loaded terms that inherently carry more specific senses of any sort, e.g. personifications of the sun, the sun seen in religious or mystical senses, the sun as a source of light, warmth and growth, etc". However, the word coded for sun is tiwat- < ti-wa-az>, which is also the word for 'sun god' in Luwian:
  - /*wa=an=tta Tarhunti* ***Tiwadi*** *Kubabaya=ha tanimi=ha=wa massani hanti sarra ariwi*/
  - "I (will) raise him up before the storm-god, the **sun-god**, and (the goddess) Kubaba, and every god" where Tiwadi is in conjunction with other gods Tarhunt and Kubaba.
  - Similarly,the word encoded for Hittite for the meaning concept 'Sun' is *ištanu*- in Hittite which, according to the references in IE-CoR itself, can refer to 'sun, sun-god(dess), solar deity; majesty' (https://iecor.clld.org/cognatesets/3334#8/40.010/34.620).

Revised etymologies and transcriptions: Anatolian

- "alaššama/i-" <a-la-aš-ša-me-in> is encoded as the word for the concept SEA and a reference is given to the discussion in Carruba ("For meaning, cf. Carruba in Kratylos 7.65").
- The meaning 'sea' is uncertain and not accepted by Luwian scholars nowadays.
- Thus the translation in *EDIANA* (s.v.) "wilderness (?)" and the extensive discussion about the semantics in Rieken and Yakubovich 2022.
- The phonetic value of some Luwian signs has been updated, but the database does not reflect our correct understanding of the sign values of Hieroglyphic Luwian signs: For example, in the lexeme for the meaning concept NAME, the transcription is "ataman-" with the signs <a-ta$_5$-ma-za>.
- However, the sign <ta$_5$> is read as phonetically [la/i] now (Rieken and Yakubovich 2010)
- Such errors (found in other languages as well) do not affect model outcome but undermine confidence in the validity of the data.

Summary of issues

- Over reliance on lexical sources/dictionaries
- Dubious textual sources
- Omission of pertinent sources
- Hapaxes
- Compound words only
- Contradiction of both general rules as well as target parameters to include additional lexemes

Conclusions

- IE-CoR has made substantial and clear improvements over its lexical database predecessors.
- Despite these improvements, and despite enlisting over 80 scholars to encode the lexical data, issues still remain.
- A feature on the IE-CoR website that allows other experts to offer suggestions could be an effective way of cleaning up some problematic historical data.
- The presence of these issues in an undertaking the size of IE-CoR suggests that, even with greater philological attention, some shortcomings of lexical data cannot be overcome.
- We argue that, in order to offset these shortcomings, phylogenetic analyses must rely on more types of data beyond lexical data.

Thanks!

- Questions?
- Author correspondence:
  - Tom Francis (tdfrancis@g.ucla.edu)
  - Klaus Kim (klausklaus@g.ucla.edu)
  - Laura McLean (lauramclean@g.ucla.edu)
  - Muhammad Rehan (rehanmuh@g.ucla.edu)
- Thanks to UCLA Keck Humanistic Inquiry Research Award

Appendix: Issues of Incomplete Paradigms in the Historical Data: Oscan

- IE-CoR lists Oscan *valaemon* for the meaning GOOD.
- As noted by Untermann (2000:821-822), *valaemon* is a superlative form, but this is not listed in the notes sections of IE-CoR: "Latin script: valaemon (nom. or acc.sg.n, TB 10). Native script: **valaimas** (gen.sg.f., Cp 37)."
    - NB **valaimas** is an onomastic; *Imagines Italicae* (2011:1441) and Buck (1928:326) both gloss *valaemon* with Latin *optimum*.
    - Cf. Latin *valaemum* 'large pear"

Oscan *valaemon*

- For a typologically irregular adjective like ‘good’ (cf. Latin *bonus, melior/melius, optimus* and Umbrian *Cubrar* ‘good’ (?)), including a superlative form leads to false comparisons.
- This is not only pertinent to the phylogenetic outcome of Latino-Faliscan vs. Sabellic branches of Italic, but also to the phylogenetic outcome of the development of Romance languages from ancient Italic languages.
- Since Romance language adjectives with a suppletive comparative also use that same suppletive form for the superlative adjective (Van Peteghem 2021:20), including a (potentially) suppletive superlative form for the concept GOOD will also necessarily lead to false comparisons with Romance languages, which exclude suppletive superlative forms.

Issues in transcription and cognate-class assignment: Anatolian

- For the concept "WASH" in Luwian, *īlḫā(i)*- is encoded as one of two default words and traced back to the cognate set (īlḫā(i)- [Luwian]) but no mention is made of the fact that this is a derivative of PIE **lóh₃u̯-/lh₃u-́*:
    - If treated ultimately as a derivative of **lóh₃(u̯)-/lh₃(u)-* (the exact details of derivation vary, cf. Melchert 2011:127–8, Sasseville 2021:87), the cognate set assigned to the word would be the same as that assigned for Gk. λοέω 'to wash', Lat. *lauō* 'to wash.

Hitt. *āk-/akk-*

- For Hittite, the word encoded for the concept ‘to die, be killed’ is *āk-/akk-* and the word is assigned the cognate set “?*Heɬʔ-” with a comment made:
  - “etym. dubious, maybe related to Lat. NECARE 'to kill', Gr. NEKYS 'corpse' or Toch. AK(E) (Weeks 1985: 76). Acc to Puhvel, an IE root is probable and links the form to an obscure Venetic compound. Naturally this is not a particularly pleasing link.”
  - The etymological link has been provided by Melchert (2012), who takes *āki/akkanzi* to reflect a remodeling of a zero-grade **neḱ-*, which nicely connects the Hittite etymon with words like Lat. *necare*, *noceō* etc.

The State of the Character Debate: Lexical or Morphological, Phonological, etc. or All of the Above?

- The question of what types of characters (lexical, morphological, phonological, syntactic etc.) should be used in a dataset for phylogenetic analysis is a long standing question.
- In addressing this issue, Heggarty et al. (2023a: 85-86) note that Ringe's analysis required supplementing phonological and morphological data with lexical data; however, instead of taking this as impetus to use combined data, they interpret it as evidence that only lexical data is needed for such a dataset.
- The disadvantage of incorporating phonological and morphological data boils down to ease, cost, and model building that can take the varying rates at which linguistic characters change into account.

Bibliography

Atkinson, Quentin D., and Russel D. Gray. 2006. "How Old is the Indo-European Language Family? Illumination or More Moths to the Flame?" In Peter Forster and Colin A. Renfrew (eds.), *Phylogenetic Methods and the Prehistory of Languages,* 91–109. Cambridge: McDonald Institute for Archaeological Research.

Buck, C. D. 1928. *A Grammar of Oscan and Umbrian with a Collection of Inscriptions and a Glossary*. Boston: Ginn and Company.

Chang, Will, et al. 2015. "Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis." *Language* 91.1.194–244.

Ceolin, Andrea, Cristina Guardiano, Monica Alexandrina Irimia, and Giuseppe Longobardi. 2020. "Formal Syntax and Deep History." *Frontiers in Psychology* 11.488871. https://doi.org/10.3389/fpsyg.2020.488871.

Clackson, James. 2007. *Indo-European Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Crawford, Michael H., William M. Broadhead, James Clackson, Federico Santengelo, Sean Thompson, and Margaret Watmough. 2011. *Imagines Italicae: A Corpus of Italic Inscriptions.* 3 volumes. London: Institute of Classical Studies University London.

Creanga, Claudiu, Cristina Guardiano, Giuseppe Longobardi, Andrea Sgarro, and Gaia Sorge. 2024. "Indo-European and Its Closest Relatives: Are There Any?" Paper Presented at the *35th Annual UCLA Indo-European Conference*, Los Angeles, CA, October 25–26, 2024.

Davies, Wendy. 1973. "*Liber Landaensis*: Its Construction and Credibility." *The English Historical Review* 88.347: 335-351.

Bibliography

Dockum, Rikker. 2018. "Phylogeny in phonology: How Tai sound systems encode their past." In *Proceedings of the Annual Meeting on Phonology* (AMP). New York: Linguistic Society of America. https://doi.org/10.3765/amp.v5i0.4238.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink, and Angela Terrill. 2008. "Structural phylogeny in historical linguistics: methodological explorations applied in Island Melanesia." *Language* 84 (3).710–759. https://doi.org/10.1353/lan.0.0069.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, and Stephen C. Levinson. 2005. "Structural phylogenetics and the reconstruction of ancient language history." *Science* 309 (5743).2072–2075. https://doi.org/10.1126/science.1114615.

Falileyev, Alexander. 2011. *Etymological Glossary of Old Welsh*. Berlin: De Gruyter.

Franchi De Bellis, Annalisa. 1988. *Il Cippo Abellano*. Urbino: Università degli Studi di Urbino.

Goldstein, David M., Shawn McCreight, Éva Buchi, and John Huelsenbeck. 2024. "An Event-Based Model for Linguistic Phylogenetics." In Jonas Nölle, Limor Raviv, Kirstie Emma Graham, Stefan Hartmann, Yannick Jadoul, Mathilde Josserand, Theresa Matzinger, Katie Mudd, Michael Pleyer, Anita Slonimska, Sławomir Wacewicz, and Stuart Watson (eds.), *The Evolution of Language: Proceedings of the 15th International Conference (EvoLang XV)*, 220–22. Nijmegen: Max Planck Institute for Psycholinguistics.

Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray. 2010. "The shape and tempo of language evolution." *Proceedings of the Royal Society B: Biological Sciences* 277 (1693).2443–2450.

Bibliography

Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael of the National Dunn, Stephen C. Levinson, and Russell D. Gray. 2017. "Evolutionary dynamics of language systems." *Proceedings Academy of Sciences* 114 (42).E8822–E8829.

Hartmann, Frederik. 2023. *Germanic Phylogeny*. Oxford: Oxford University Press.

Heggarty, Paul. 2021. "Cognacy databases and phylogenetic research on Indo-European." *Annual Review of LInguistics* 7.371–94.

Heggarty, Paul et al. 2023. "Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages." *Science* 381.eabg0818.

Herce, Borja, and Balthasar Bickel. 2025. "Paradigmatic complexity metrics as signals of phylogenetic relatedness: A proof of concept in Romance and Pamean diachrony." *Diachronica* 42 (1): 1–46. https://doi.org/10.1075/dia.23004.her.

Kassian, A. S., and G. Starostin. 2025. "Do 'Language Trees with Sampled Ancestors' Really Support a 'Hybrid Model' for the Origin of Indo-European? Thoughts on the Most Recent Attempt at Yet Another IE Phylogeny." *Humanities and Social Sciences Communications* 12.682.

Kim, Ronald I. 2024. "On the Phylogenetic Status of East Germanic." In Joseph F. Eska, Olav Hackstein, Ronald I. Kim, and Jean-François Mondon (eds.), *The Method Works: Studies on Language Change in Honor of Don Ringe*, 21–43. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-48959-4_2.

Kroonen, G., T. Olander, M. Nørtoft, et al. 2023. "Archaeolinguistic Anachronisms in Heggarty et al. 2023." *Science* (letter responding to Heggarty et al., "Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages," *Science* 381.eabg0818).

Bibliography

Lampe, G. W. H. 1961. *A Patristic Greek Lexicon.* Oxford: Clarendon.

List, Johann-Mattis, Mary Walworth, Simon J. Greenhill, Tiago Tresoldi, and Robert Forkel. 2018. "Sequence Comparison in Computational Historical Linguistics." *Journal of Language Evolution* 3 (2).130–44.

Longobardi, Giuseppe, and Cristina Guardiano. 2009. "Evidence for Syntax as a Signal of Historical Relatedness." *Lingua* 119 (11).1679–706. https://doi.org/10.1016/j.lingua.2008.09.012.

———. 2017. "Phylogenetic Reconstruction in Syntax: The Parametric Comparison Method." In Adam Ledgeway and Ian Roberts (eds.),*The Cambridge Handbook of Historical Syntax*, 241–72. Cambridge: Cambridge University Press

Macklin-Cordes, Jayden L., Claire Bowern, and Erich R. Round. 2021. "Phylogenetic Signal in Phonotactics." *Diachronica* 38 (2).210–58. https://doi.org/10.1075/dia.20004.mac.

McDonald, Katherine. 2022. *Italy Before Rome: A Sourcebook*. New York: Routledge.

Melchert, H. Craig. 2007/8. "Neuter Stems with Suffix *-(*e*)*n*- in Anatolian and Proto-Indo-European." *Die Sprache* 47 (2).182–91.

———. 2010. "The Word for 'mouth' in Hittite and Proto-Indo-European." *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 7.55–63.

———. 2011. "The PIE Verb for 'to pour' and Medial *h_3 in Anatolian." In Stephanie W. Jamison, H. Craig Melchert, and Brent Vine (eds.), *Proceedings of the 22nd Annual UCLA Indo-European Conference*, 127–32. Bremen: Hempen.

———. 2012. "Hittite *ḫi*-Verbs of the Type –āC_1i, -aC_1C_1anzi." *Indogermanische Forschungen* 117.173–85.

———. 2024. *A Dictionary of Cuneiform Luvian*. Ann Arbor: Beech Stave Press.

Bibliography

McMahon, April, and Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199279012.001.0001.

Milroy, James. 2001. "Language Ideologies and the Consequences of Standardization." *Journal of Sociolinguistics* 5:530–55.

Moran, Steven, Eitan Grossman, and Annemarie Verkerk. 2020. "Investigating Diachronic Trends in Phonological Inventories using BDPROTO." *Language Resources and Evaluation* 55:79–103. https://doi.org/10.1007/s10579-019-09483-3.

Moran, Steven, and Annemarie Verkerk. 2018. "Differential Rates of Change in Consonant and Vowel systems." In Chris Cuskley, Marina Flaherty, Holly Little, Luke McCrohon, Annalena Ravignani, and Tim Verhoef (eds.), *The Evolution of Language* (*EVOLANG XII*). NCU Press. https://doi.org/10.12775/3991-1.077.

Nakhleh, Luay, Don Ringe, and Tandy Warnow. 2005. "Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages." *Language* 81 (2).382–420.

Nakhleh, L., T. Warnow, D. Ringe, and S. N. Evans. 2005. "A Comparison of Phylogenetic Reconstruction Methods on an IE Dataset." *Transactions of the Philological Society* 3 (2).171–192.

Newman, Barclay M. 2010. *Greek-English Dictionary of the New Testament*. Stuttgart: Deutsche Bibelgesellschaft.

Nichols, Johanna. 1996. "The Comparative Method as Heuristic." In Mark Durie and Malcom Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 39–71. Oxford: Oxford University Press

———. 2014. "Derivational Paradigms in Diachrony and Comparison." In Martine Robbeets and Walter Bisang (eds.), *Paradigm Change in the Transeurasian Languages and Beyond*, 61–88. Amsterdam: John Benjamins.

Pereltsvaig, Asya and Martin W. Lewis. 2015. *The Indo-European Controversy*. Cambridge: Cambridge University Press.

van Petegham, Marleen. 2021. "Comparatives and Superlatives in the Romance Langauges." *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press, 1-34.

Bibliography

Rama, Taraka, Johann-Mattis List, Johannes Wahle, and Gerhard Jäger. 2018. "Are Automatic Methods for Cognate Detection Good Enough for Phylogenetic Reconstruction in Historical Linguistics?" In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 393–400. New Orleans, LA: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-2063.

Rexová, Kateřina, Yvonne Bastin, and Daniel Frynta. 2006. "Cladistic Analysis of Bantu Languages: A New Tree Based on Combined Lexical and Grammatical data." *Naturwissenschaften* 93 (4).189–94. https://doi.org/10.1007/s00114-006-0088-z.

Rieken, Elisabeth, and Ilya Yakubovich. 2010. "The New Values of Luwian Signs L 319 and L 172." In Itamar Singer (ed.), *ipamati kistamati pari tumatimis: Luwian and Hittite Studies Presented to J. David Hawkins on the Occasion of his 70th Birthday*, 199–219. Tel Aviv: Emery and Claire Yass Publications in Archaeology.

Rieken, Elisabeth and Ilya Yakubovich. 2022. Zu den Reflexen der Wurzel **al-* in den anatolischen Sprachen. In Hannes Fellner, and Melanie Malzahn (eds.), *Zurück zur Wurzel. Struktur, Funktion und Semantik der Wurzel im Indogermanischen*, 267–80. Bremen: Hempen.

Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. "Indo-European and Computational Cladistics." *Transactions of the Philological Society* 100 (1).59–129.

Rönchen, Philipp, Tilo Wiklund, and Harald Hammarström. 2024. Likelihood Calculation in a Multistate Model of Vocabulary Evolution for Linguistic Dating. *Language Dynamics and Change* 14.1–41.

Bibliography

Sasseville, David. 2020. *Anatolian Verbal Stem Formation: Luwian, Lycian and Lydian*. Leiden: Brill.

Scarborough, Matthew. 2020. "Cognacy and Computational Cladistics: Issues in Determining Lexical Cognacy for Indo-European Cladistic Research." In Matilde Serangeli, and Thomas Olander (eds.), *Dispersals and Diversification: Linguistics and Archaeological Perspectives on the Early Stages of Indo-European*, 179-208. Leiden: Brill.

Tucker, Elizabeth. "Lexical Variation in Young Avestan: The Problem of the 'Ahuric' and 'Daevic' Vocabularies Revisited." In Michele Bianconi, Marta Capano, Domenica Romagno, and Francesco Rovai (eds.), *Ancient Indo-European Languages between Linguistics and Philology: Contact, Variation, and Reconstruction*, 254–275. Leiden: Brill.

Untermann, Jürgen. 2000. *Wörterbuch des Oskisch-Umbrischen.* Heidelberg: C. Winter universitätsverlag.

de Vaan, Michiel. 2019. "On the Homonymy of 'put' and 'suck' in Proto-Indo-European. *Indo-European Linguistics* 7.176–93.

Vine, Brent. 2019. "Greek στωμύλος 'chatty': An anomalous *ō*-grade (and some anomalous *o*-grades)." *Indo-European Linguistics* 7.222–40.

Wennerberg, Claes. 1972. Indogermanisch **stomen*- 'Mund'. *Die Sprache* 18.24–33.

Wichmann, Søren, and Arpiar Saunders. 2007. "How to Use Typological Databases in Historical Linguistic Research." *Diachronica* 24 (2).373–404. https://doi.org/10.1075/dia.24.2.06wic.