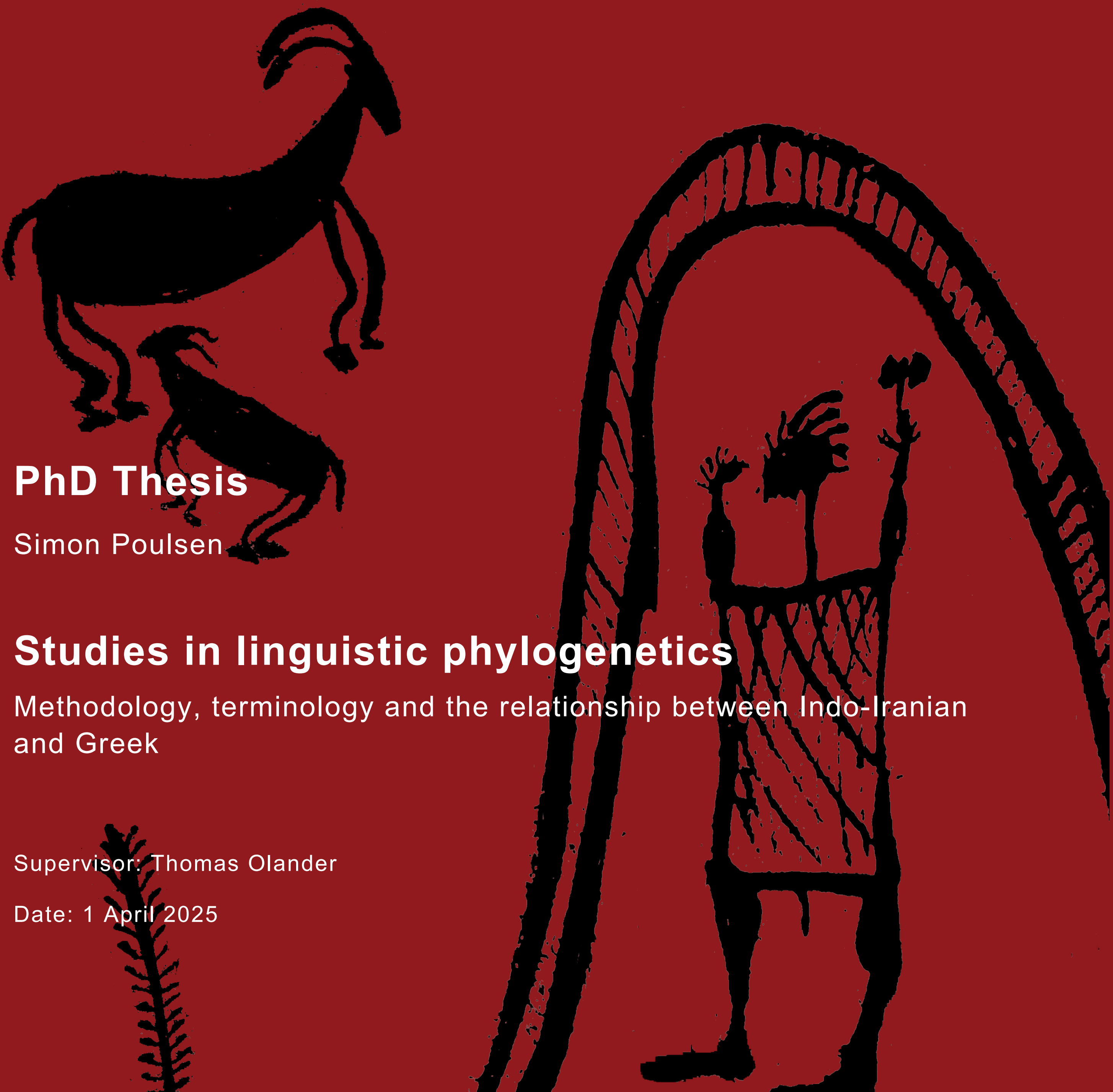
UNIVERSITY OF COPENHAGEN
DEPARTMENT OF NORDIC STUDIES AND LINGUISTICS



PhD Thesis

Simon Poulsen

Studies in linguistic phylogenetics

Methodology, terminology and the relationship between Indo-Iranian and Greek

Supervisor: Thomas Olander

Date: 1 April 2025

Abstract

This thesis examines the relationship between Indo-Iranian and Greek, two branches of Indo-European that are on the one hand frequently suspected of forming a clade or having been spoken close to each other in prehistory, but on the other hand preserve most of the linguistic material on which Proto-Indo-European is reconstructed. It consists of three studies in article form and an introductory section dealing with the methodology of higher-order subgrouping and its application to the hypothesised latest common ancestor of Indo-Iranian and Greek. First, it assesses the strengths and weaknesses of traditional phylogenetics, computational phylogenetics and prehistoric dialectology. This section reveals that traditional methods have a weakness when comparing very conservative and abundantly attested languages with much more scarcely attested languages, but that computational methods rely on the least salient evidence to avoid the data bias. Then it assesses the phonological isoglosses shared by Indo-Iranian and Greek and reconstructs the phonological system of the latest common ancestor. This reveals that it is impossible to reconstruct a younger ancestor than Proto-Indo-European, and that the proposed innovated phonemes are either inherited (**a, *b)* or parallel (voiceless aspirates). It assesses the morphological isoglosses shared by Indo-Iranian and Greek which reveals that the two branches are – unsurprisingly – conservative representatives of non-Anatolian Indo-European. Their unique isoglosses are either unidentifiable in other branches or rather parallel and arose form system-internal pressure than contact. Finally, it applies an archaeolinguistic approach to narrow down when and where the alleged proto-language or ancestral dialect continuum was spoken: 2000 bce north of the Black Sea.

Article 1 presents a survey of the term "Indo-Greek" and of published tree topologies. It concludes that the "Indo-Greek hypothesis" does not exist in a narrow phylogenetic sense, and that no (available) family tree places the two branches as sisters. Article 2 discusses the role of loanwords in linguistic phylogenetics under different methodological approaches and supplies two case studies. It concludes that **peleḱu-* 'axe' cannot be a borrowing from semitic, and **(H)a(i̯)ĝ-* 'goat' is unlikely to be Caucasian. Article 3 is a Maximum Parsimony analysis of East Iranian; a group of languages often considered a Sprachbund. The results are inconclusive, but the paper offers new insights into the practical applications on computational analysis on grammatical data.

Danske resumé

Denne afhandling undersøger forholdet mellem indoiransk og græsk; to sproggrene som på den ene side ofte mistænkes at have udgjort en undergruppe eller at have været talt i nærheden af hinanden i forhistorien, men som på den anden side er ophav til meget af det lingvistiske materiale, som urindoeuropæisk kan rekonstrueres på baggrund af. Den består af tre artikelformede studier og en indledende del, som behandler forskellige metodiske tilgange til lingvistisk undergruppering og deres anvendelsesmuligheder på indoiransk og græsk. Til at begynde med opvejer den styrker og svagheder ved traditionelle fylogenetiske metoder, computer-drevne kvantitative metoder og forhistorisk dialektologi. Det viser sig, at de traditionelle metoder møder en faldgrube, når de skal appliceres på sprog med meget forskellige grader af attestering, men at kvantitative metoder i deres forsøg på at undgå dette skævhed i data i stedet baserer sig på den mindst signifikante evidens. Dernæst undersøger afhandlingen de foreslåede fonologiske isoglosser og rekonstruerer det seneste grundsprog, som indoiransk og græsk kan gp tilbage det. Det er umuligt at rekonstruere et senere fonologisk stadie end urindoerupæisk. De delte isoglosser er enten arkaismer (_*a, *b_) eller opstået parallelt (ustemte aspirater). Efterfølgende undersøger afhandlingen de morfologiske isoglosser. Det konkluderes at indoiransk og græsk, ikke overraskende, er konservative ikke-anatoliske sproggrene. Deres unikke isoglosser er enten usammenlignelige i andre grene eller snarere parallelle dannelser, som i højere frad skyldes systeminternt pres end forhistorisk kontakt. Til sidst anlægger afhandlingen en arkæolingvistisk metode for at indsnævre hvor og hvornår de to grene sidste gang var i nærheden af hinanden: 2000 f.v.t. nord for Sortehavet.

Artikel 1 er en undersøgelse af termen ”indogræsk” og af publicerede stamtræer. Den konkluderer, at den indogræske hypotese aldrig har eksisteret i snæver forstand, og at intet (tilgængeligt) stamtræ viser de to grene som søstre. Artikel 2 diskuterer låneords rolle i lingvistisk fylogenetik. Den konkluderer, at _*pelek̂u-_ 'økse' ikke kan være fra semitisk, og af _*(H)a(i̯)ĝ-_ 'ged' næppe er kaukasisk. Artikel 3 er en Maximum Parsimony analyse af de østiranske sprog, som ofte bliver anset som et Sprachbund. Resultaterne er inkonklusive, men artiklen giver ny viden om de praktiske anvendelsesmuligheder af kvantitative metoder til analyse af grammatisk data.

Acknowledgements

TAK. This little word is written way too early in the morning after a very long night, day, week, month, year of working on this study – and *countless* side quests.

I wish to thank my colleagues at *Roots of Europe* and *Connecting the Dots* (CtD) – including Nicolas Jansens, Marc Canby, Rasmus Bjørn – and the Swedish satellite Oscar Billing and Erik Elgh. I am also grateful to Don Ringe for the fruitful discussions we had during the *CtD*-research stay, where I also properly met and befriended Dr. Scarborough during an exclusive viewing of the Helvetica documentary. I am also grateful to Lucien van Beek for hosting my own research stay in Leiden – a big thanks to the wonderful community of PhD candidates, including (but probably not limited to) Louise Friis, Axel Palmér, Andrew Wigman, Abel Warries, Niels Schoubben, Rasmus Thorsø.

I owe thanks to Martin Kümmel, Agnes Korn, Thomas Jügel and Stefan Schaffner for sharing their versions of Klingenschmitt's dissertation with me, and to Sarah Mentzler for compiling the Appendix of East Iranian trees.

On a more personal level, I thank all of my fellow PhD candidates at *NorS*. Above all, of course, my office-mate und Bundesmutti, Antje Schöps. Zänk you for keeping the office gelüftet. I am also incredibly grateful for all of my wonderful friends who kept me from going clinically insane: The Ædegilde-gang and your offspring, the ex-Egmont theatre crew, Anders Solitander Bohlbro, Lærke Olsen, Amina Aïssaoui, Kali Stuart, Benjamin Schrott, Celina Digebjerg, Regan Pierce and Sif Dam Sonniks (without whom I would be hungry). Thanks for all the love and support to Martin, my parents, Lea, Nick, August, mormor og morfar – I wish I had the confidence you have in me.

But most of all, thank you, Thomas!

Contents

Part 1

1. Introduction

Since the dawn of Indo-European comparative linguistics, the relationship between Greek and Indo-Iranian has sparked curiosity and excitement. Before the discovery and decipherment of the Anatolian and Tocharian languages, most of Proto-Indo-European was indeed reconstructed on material from these branches, that are attested relative early and in abundance.

With the rise of archaeogenetic data and methods, the dispersal of the Indo-European languages and the prehistoric migrations of the peoples that brought them are key areas of interest. It is, however, increasingly clear that there are several issues with the chronology of the archaeolinguistic and archaeogenetic findings. A crucial method for improving the connection between the real world and its artefacts and the language that mediated their creation, is a phylogenetically informed reconstruction of the cultural vocabulary. Not everything that shines like Indo-European is in fact Proto-Indo-European. This has become increasingly clear from the insights into the prehistory of the Anatolian languages that have led to new evaluations of the Indo-European proto-language and the innovations that led to its attested daughters. The dispersal of the Proto-Language and the nature of the branchings are still largely unknown, but they play a crucial role in the connection of prehistoric languages and cultures. In order to make any credible claims on the connection of archaeolinguistics and archaegenetics, we must have a reasonably clear view of the diversification of Proto-Indo-European. Otherwise, we risk searching for contacts and culture that never existed, or we risk placing the Homeland of a part of the family in an area where its speakers never was.

It is that Proto-Indo-European proper cannot be reconstructed without Anatolian, but it is much less clear what the alternative branches on the other side are. In fact, the "Indo-Greek" hypothesis has never been substantiated much. The term "Indo-Greek" is frequently used to describe some sort of ill-defined non-Anatolian Indo-European, but it is also mentioned as a clear-cut hypothesis. In a recent preprint of an archaegenetic paper, the researchers summarize:

> The so-called Indo-Greek hypothesis groups Greek as well as the closely related Phrygian with Indo-Iranian, while the competing Graeco-Armenian hypothesis posits that Greek forms a subclade

> with Armenian, possibly also including Albanian "Balkan Indo-European"
>
> (Yediay et al. 2024: 3)

Without realising it, this is quite different from the original "Indo-Greek" hypothesis:

> Das Griechische ist mit mehreren idg. Sprachen durch charakteristische Gemeinsamkeiten verbunden: im Wortschatz vor allem mit dem Armenischen, im Lautstand mit dem Iranischen und im Verbalparadigma mit diesem und dem Indischen. Diese Übereinstimmungen (Isoglossen) beruhen teils auf gemeinsamer Bewahrung von Bestandteilen der idg. Grundsprache, teils auf Sprachberührungen in vorhistorischer Nachbarschaft, die dem späteren Sprachaustausch zwischen Griechisch und Latein vergleichbar sind.
>
> (Rix 1976: 8)

In order to further the stratified reconstruction of Indo-European, we need to settle the relationship between Indo-Iranian and Greek. If their grammatical similarities can be ascribed to innovations in a (late) common ancestor, then the prehistoric culture we lean from their shared vocabulary might be equally young and exclusive to the other branches. On the other hand, if the similarities are ancient, all differences in other branches must be innovations on their behalf. Finally, if Indo-Iranian and Greek formed a dialect continuum in prehistory, this raises questions of when, where and with how many other branches.

This thesis will thus examine the nature of the relationship between Indo-Iranian and Greek. It will do so from a phylogenetic angle and test if can reconstruct a common ancestor that is later than Proto-Indo-European for them. In order to do so, the mythical nature of the "Indo-Greek" comparisons of the past must be mapped out. To what extend can their similarities be formulated as shared innovation is a phylogenetic sense, and if they can – how many other branches also descend from this latest shared ancestor of Indo-Iranian and Greek.

The thesis is structured as follows. It consists of two parts. Part 1 contains the methodological framework and a survey of the proposed Indo-Greek isoglosses. Chapter 2 is a thorough investigation into the methodology of linguistic subgrouping, comparing traditional phylogenetic methodology to classical dialectology and computational methods. Chapter 3 gives a brief summary of the shared isoglosses proposed in the literature. Chapter 4 is a reconstruction of the phonological system of the latest shared ancestor of Indo-Iranian and Greek. Chapter 5 is a survey of the morphological isoglosses shared between Indo-Iranian and Greek in light of the diverging "Indo-Greek" hypotheses just quoted. Chapter 6 applies the archaeolinguistic method to the latest common ancestor of Indo-Iranian and Greek to substantiate the claims of their areal affinities or shared prehistory.

Part 2 contains three articles:

Article 1 is a survey into the research history of Indo-Greek comparison though the lens of terminology and topologies.

Article 2 assesses the role of loanwords in linguistic phylogenetics under different methodological approaches. Culture-specific vocabulary is crucial evidence for linguistic prehistory, but it is inherently tied to phylogenetics and circular assumptions of prehistory. It is accompanied by Addenda & Corrigenda from fruitful discussions after its publication.

Article 3, co-written with Agnes Korn, follows the phylogenetic and contact-linguistic track, but shifts the scope. It presents a Maximum Phylogeny analysis of the East Iranian languages, widely held to be a dialect continuum. It does so to disentangle unsubstantiated claims on linguistic prehistory, and to present a case study on a material that more confined than "Indo-Greek". It is preposed by a preamble elaborating the methodological approach.

2. Methodology of (higher-order) linguistic subgrouping

2.1. Introduction

To explore the relationship between Greek and Indo-Iranian, or any related languages for that matter, it is necessary to lay the methodological grounds firmly first. In the following, I will address the methodology of linguistic subgrouping – with an emphasis on higher-order subgrouping. Much has been written on the discipline, and I rely heavily on previous works (Hoenigswald 1966; Clackson 1994: 11–27; Ringe, Warnow & Taylor 2002; Nakhleh, Ringe & Warnow 2005; Nichols & Warnow 2008; Ringe 2017a; Olander 2018; 2019a; 2019b; Jacques & List 2019; Campbell 2020: 219–253; Thorsø 2020; Scarborough 2020; 2022a; Clackson 2022; Ringe 2022; Greenhill & Gray 2012; Pereltsvaig & Lewis 2015; Greenhill, Heggarty & Gray 2021; Heggarty et al. 2023a; Pellard, Ryder & Jacques forthc.).

I cannot offer a full and detailed research history. In Article 1, I give a survey of the research history of the relationship between Indo-Iranian and Greek viewed through and exploration of the terminology "Indo-Greek" and "Græco-Aryan" and published family trees. For the development of linguistic phylogenetics from the Neogrammarians to 1994, I refer to Clackson (1994: 1–11) who claims not to be exhaustive, but is at any rate very rigorous. Coincidentally, Ringe and colleagues began working on their phylogenetic data in 1994 [1] which led to a series of publications (Warnow, Ringe & Taylor 1996; Ringe et al. 1998; Ringe 1998; Taylor, Warnow & Ringe 2000; Nakhleh et al. 2005) – and most prominently Ringe, Warnow & Taylor 2002 and Nakhleh, Ringe & Warnow 2005, both of which contain thorough methodological sections. For the research history after "the quantitative turn", see Jacques & List (2019: 135–8).

The revolution of DNA sequencing paved the way for the rise Bayesian and other statistical methods in biological phylogenetics where they are now predominant and by far superior methods. Inspired by this, lexicon-based statistic methods (not to be confused with lexicostatistics) swept its way into the linguistic debate (Gray & Atkinson 2003; Bouckaert et al. 2012; 2013; Chang et al. 2015; Kassian et al. 2021a; Heggarty et al. 2023a). While the results of these studies are generally not held in very high regard by experts within the field of Indo-European linguistics, they are

---

[1] Don Ringe, p.c., Philadelphia, March 2022.

certainly prominent and have caused newspaper headlines all over the world. It continues to fascinate that ancient and irrecoverable truths can be recovered my modern scientific methods – but while the results are easily represented in beautiful graphs, the magic of the machinery is opaque, and the calculations are much less exciting than the claims.

2.2. The family tree model

2.2.1. Linguistic relationship models

Linguistic relationships can be described in different ways and through different models and metaphors (Drinka 2013a: 393–7, with refs.). Among these are the wave model (Schmidt), rivers and tributaries (Mufwene), (river) banks (Terell,) nets (McMahon & McMahon), networks (Nakhleh et al.) and, most famously, the tree model (Schleicher's *Stammbaum*). The "starburst" phylogeny where all ten branches disperse individually from the parent language is hardly a model at all; all it shows is that the languages are related but not the same.

While tree model is simplistic, it offers a great tool for the exploration of linguistic prehistory. Notably, it produces falsifiable hypotheses and can be applied as a heuristic devise to optimise the reconstruction of proto-languages as realistic linguistic entities, not mere repositories of etymological projections.

2.2.2. Significant shared innovations

The prevailing hypothesis is that linguistic genetic relationship (once relatedness has been established) can only be based on **shared non-trivial** or **significant innovations**. Additionally, it holds that it is more likely that two (or more) languages who share the same non-trivial innovated feature underwent this innovation only once when they still formed a coherent speech community, and that they form a subgroup, rather than that they innovated identically, but independent of each other. This is truly a powerful tool, but it is also very restrictive, which becomes apparent when we pick apart the individual elements.

2.2.3. Innovations, not archaisms

What a linguistic innovation is self-explanatory – in theory. In practise, it is a much more complicated matter to determine what is the innovation, what is the archaism, or if or if no attested language attest the state that should be reconstructed for their parent language:

> It is worth noting here that there is a danger of **circularity in the arguments** for sub-groups, particularly if sub-groups are assumed during the reconstruction of the parent language
>
> (Clackson 1994: 14)

The family tree is deeply embedded in linguistic reconstruction: A feature only found only in on language (a leaf in the tree metaphor), say, Modern West Jutish Danish is rarely reconstructed all the way back to the root, Proto-Indo-European. But since West Jutish descends from Proto-Indo-European, it might as well preserve features (which it does, e.g. **u̯* which is kept [w] as in English). A lot of factors are taken into consideration when reconstructing. Among these are relative age and chronology (which the West Jutish example hints at), directionality (which will be discussed extensively below and in Article 3) and geographic distribution (it is clear that languages do not exist in a vacuum, and horizontal transfer (borrowings and other contact phenomena) is very real).

Assuming for a moment that the ancestral state is clear, an innovation can constitute a new creation a feature (e.g. a lexical root, a meaning, a derivation, a derivational suffix, a morphological category, an inflectional ending, a phoneme, a word order etc.) but also the generalisation of one of two existing features, the coalescence of two or even the loss of a feature.

It is important to stress that shared archaisms do not make for arguments of linguistic subgrouping. The model takes for granted that all languages descending from the same parent language possessed the same ancestral state at some point in their prehistory. Unless they have undergone the innovation of loss, the feature must be preserved.

However, that obviously does not mean that archaisms, contact phenomena and structural (superficial) similarities are irrelevant. They are among the features that give languages their unique flavour. To distinguish between significant features in this sense and significant features in the linguistic phylogenetic sense (see below), I find the term “branch-defining” feature as employed by Peyrot (2022: 83–4) useful.

2.2.4. Shared, not identical

Linguistic phylogenetics is not concerned with superficial similarities. The term "group" is often applied very vaguely to any languages or dialects that share features for a multitude of reasons, but the term subgroup is applied in a narrower sense.

For an innovation to be **shared**, it does not suffice that it is identical. We could add to the definition that **shared** does not simply mean identical, but ideally **the same**, that it something that happened once in the shared prehistory of the languages in question. In character-based cladistics (see Article 2, Article 3), this is formulated as follows:

> States must be assigned on the principle that each state should have arisen **only once in the evolutionary history of the family**
>
> (Ringe, Warnow & Taylor 2002: 71, my emphasis)

Notice already here the difference between my definition and the one quoted. In an ideal world, we would only base groups on those innovations that certainly only took place once in the evolutionary history, but in practise, it is often impossible to rule out independent, parallel and convergent evolutions.

This definition which in some way is used by any serious phylogeneticists means that it is a complete strawman when Heggarty and colleagues argue that French could be considered a satəm language (2023b: 85–6). The centum/satəm and the *ruki*-rule are classic examples of isoglosses or innovations that can fall under this definition, but also spike caution and lead to discussion since they are not completely identical in the languages that exhibit them. Accordingly, it is not self-explanatory that they occurred only once in the prehistory of the languages that exhibit them. It is, however, excluded that French could have taken part in the "satəm"-merger. While PIE **k̂* is fronted to *s* before a front vowel, and **kʷ* ends as /k/, historical data show that this is *much* later, and that the assibilation of **k̂* applies to both **k* and **k̂* - but only in front of an Old French front vowel, e.g. *cœr* 'heart', *croire* 'believe', not ***sœr,* ***sroire* (< Lat. *cord-, crēdere* < PIt. **kord-,* **krezðe-*) < ultimately PIE **k̂erd-*, **k̂r̥d-* (cf. Lith. *širdìs*, Ved *hr̥d*[2], *śraddháti*).

[2] On the unexpected reflex *h* (also Av. *zərəd*), not ***ś* (and s), see Jacques (2019 with refs.)

Sometimes, relative chronology can offer a helping hand in distinguishing between what can and cannot have taken place only once in the evolutionary history. Once something has been lost, it cannot reemerge in the same distribution. For example, while both Greek and Indo-Iranian lose the laryngeals, the loss of them **H* > *Ø* cannot be shared since they leave irreconcilable traits (prothetic vowels in Greek, intervocalic hiatus in Indo-Iranian).[3] More importantly, intervening changes can sometimes settle if a change is **shared** or not. For example, it could be argued that Indo-Iranian shares the coalescence of **(H)a, o* with Baltic. However, since Brugmann's Law applies only to PIE **o* not **a*, and Winter's Law lengthens **o, a* to **ō*, *ā* that do not coalesce in Baltic, "since we have Li. *núogas* (a.p. 3), Lv. *nuôgs* from PBS *ˌ*nō̰gas* (not **ˌ*nā̰gas*) < PIE **nogʷos*" (Olander 2015: 51), the change cannot have taken place once in the joint prehistory of Baltic and Indo-Iranian.

Innovations that are identical, but did not occur only once, are called parallel. As mentioned, it often difficult to distinguish between parallel and shared innovations with any certainty (Ringe, Warnow & Taylor 2002: 66–8). To account for these scenarios, traditionalists distinguish between different types of innovations (see next section), and computationalists can work with mathematical likelihood, probabilistic approaches (Scarborough 2023), quantitative cumulative evidence or weights and costs. It has, however, been argued that it is impossible to single out shared innovations in our sense, because any innovation is in some way motivated by the system in which it occurs (Harrison 1986).

While we can sometimes say that changes, such as the merger of **a* and **o* are parallel, it may also be necessary to distinguish between **shared** and convergent innovation. The latter did, in some way, take place only once – and indeed at the same time – but not once in the shared evolutionary history of the languages. Confusing at this is, it is possible for languages or dialects to innovate together – due to borrowings, pressure from areal Sprachbünde and other contact phenomena when they no longer formed the same node in a phylogenetic sense. This gives rise to two important questions: What is a node in a phylogenetic tree? And is the phylogenetic model of language change realistic? These will be discussed below.

---

[3] At least not in terms of phonemic mergers, see "Phonology" below and Chapter 4.

2.2.5. Significance

The greatest issue is that most linguistic changes are readily repeatable and not locked in place by relative chronologies as explained above. As mentioned, this can be combatted by a weighting scheme or by counting cumulative evidence, but it still leaves the fact that all languages are surprisingly similar. It is the same phonetic, auditive and cognitive processes that drive linguistic change, and while the social motivating factors vary greatly, the ultimate results are still very similar.

This is why **shared innovation** is accompanied by **non-trivial** or **significant** in traditional phylogenetics. To rule out chance resemblance and parallel innovations, it is necessary to base the subgrouping arguments only on material that is unlikely to have occurred independently. However, it should be emphasized that there is a high degree of subjectivity in what is deemed significant (Clackson 2022: 25–6). Sadly, there is no quantitative measure for how normal a linguistic change is. There is no universal number (a lower-bound heuristic) to prove a relationship, nor does it make sense to reconstruct the proto-language based on the number of s continuing a trait (majority-rule heuristic) (Goldstein 2022).

Luckily, we do not need a number to estimate the likelihood of linguistic change. While most innovations are "natural", some are not. Cross-linguistic typological studies of linguistic change, be it in phonology, morphology, syntax, convergence or a combination of it all help distinguish between likely and rare innovations. This is easier for phonology where the number of phonemes is ultimately relative small, and all humans produce the speech sounds with the same organs, than it is for morphology where the internal pressure from the system and the external factors can be multifaceted. Harrison's objection should, however, be recalled. If even an unlikely innovation could happen once, why not twice, independently?

2.2.6. Positive evidence and absence of evidence

At this point, it important to introduce a pitfall which many traditional phylogenetic works have fallen into, and which I will discuss multiple times in this thesis and the articles: The danger of judging from positive evidence alone. While it is tempting to base a subgrouping argument on striking similarities or innovations between two languages or branches, it cannot be done convincingly without considering all other potential members of this subgroup. This becomes evident when comparing languages that are attested over the course of millennia and whose degree of innovation vary greatly. If two branches, to take an example, make up 46 percent of the compared material, it is hardly surprising that there are more striking innovations among these languages. Especially not if another potential candidate for a close relative continues only 2 or 3 percent.

This is the reason I have no sought to solve the mystery of the nature of the relationship between Indo-Iranian and Greek only by seeking out unique isoglosses in, for example, the *LIV²* or the *NIL*, as a modern-day reiteration of Birwé (1956) and Euler (1979). Such a study would, however, not be as tedious as it used to be thanks to recent publications (Piwowarczyk 2022a). While these reference works are immensely important sources, counting positive evidence in them will lead us astray.

The *LIV* assumes that it is a gross catalogue of the roots and stem formation patterns in Proto-Indo-European. This was criticised by Seebold (1999) in his review of the first edition (*LIV¹*) who argued that much of the material need not be Proto-Indo-European, but rather belongs to younger strata or is geographically confined. However, the *LIV²* doubles down:

> [...] die Behauptung, eine einzelsprachlich abstrahierte Wurzel sei erst nachurindogermanisch entstanden [ist] nicht weniger willkürlich als ihr Ansatz für das Urindogermanische. Sicher, je mehr Einzelsprachen eine Wurzel bezeugen, desto weniger wahrscheinlich ist erst einzelsprachliche Entstehung. Man darf aber [...]den Schluß nicht umkehren und annehmen, daß eine nur in einer einzigen Einzelsprache belegte [...] Wurzel nachurindogermanisch sei; das hieße, den Zufall der Überlieferung zum Maßstab für das rekonstruierte Lexikon zu machen.
>
> (*LIV²*: 34–5)

This directly violates the more cautious phylogenetic approach to reconstruction outlined below. It also has the consequence that branches that are well-attested through large-corpus languages (like Greek and Indic), and branches with a more conservative vocabulary because they did not pass through a period of extensive language contact (like Armenian and Albanian did) – or because they died out beforehand will have greater potential for shared features. However, as the LIV considers itself a gross catalogue of PIE, everything shared is simply archaisms and therefore phylogenetically irrelevant (*pace* Holm 2008).

In terms of stem formations and not lexical roots, it will also give an advantage to languages with a conservative morphology (like Greek and Indo-Iranian). This is where the numbers above come from; they are not purely theoretical: Indic (20%), Iranian (12%) and Greek (14%) make up 46% of the stem formations occurring in more than one branch in the *LIV*^{2}. Albanian and Armenian only continues 2% and 3%, respectively. It is essentially more informative to examine with which branch(es) Armenian and Albanian share stem formations when any survive, rather than searching for isoglosses in the majority-languages. These isoglosses could equally well be lost in Armenian and Albanian.

This ties into a final criterion for significance, known to Peyrot (2022: 90) as **identifiability** and overlapping with the computational concept of **Parsimony Informativity** (see Article 3) (Ringe, Warnow & Taylor 2002: 71 esp. n. 8). To recognise a significant innovation, the domain should be identifiable in the lower as well as the higher node. Applying this to the current example, counting instances of, say, uniquely shared reduplicated aorists in Indo-Iranian and Greek is pointless, as the category is lost a "productive" category in all other branches, and only survives (through mergers with the reduplicated perfect) in Armenian, Tocharian and Italic[4] (Bendahman 1993: 1). For the curious reader, the number is three: **u̯eu̯kʷe-* 'said' survives in all three (Gr. *εἶπε* < **u̯ei̯kʷe-*; Ved. *ávocat*, Av. *-avaocat̰* < PIIr. **(H)a-u̯au̯ča-*), and two survive in Greek and Iranian, exclusively: **gʷʰé-gʷʰn-e-* 'slew' (Gr. *ἔπεφνον*, YAv. *-yaγnat̰*) and **h₁é-h₁nk̂-e* 'reached, took' (Gr. Att. *ἐνεγκεῖν*, OAv. *ną̇sat̰*).

If the category is lost, so are the endings associated with it, and if a formation type is lost, so are all occurrences of it, which is why it is difficult to find informative morphological data (Nakhleh et al. 2005: 385). It also holds true for many intriguing innovations in morphology. In fact, the rich and extensive morphologies of the ancient Indo-European languages is somewhat of a luxury. Imagine for a moment that the Tocharians had walked further south, and instead of the heavy substrate influence from Samoyedic (Peyrot 2019), had acquired a Tibetan phonetic system, there would be nothing left to secure a subgrouping with other Indo-European

---

[4] Maybe also Celtic if OIr. *-fúar* < **u̯é-u̯r(h₁)-e-* (*LIV*^{2}: 698).

languages than root cognates and phonology of the first (and only) syllable (cf Pellard, Ryder & Jacques forthc.: 18–22).

exclusive shared roots (Phrygian, Messapic omitted)

	Anatolian	Tocharian	Italic	Celtic	Germanic	Greek	Armenian	Albanian	Indo-Iranian	Balto-Slavic
Anatolian	169	5	3	0	1	9	0	0	13	4
Tocharian		140	5	0	0	5	1	0	6	3
Italic			317	10	10	12	0	1	11	13
Celtic				184	4	5	1	0	7	6
Germanic					369	10	1	2	17	50
Greek						467	9	4	44	23
Armenian							108	1	2	2
Albanian								81	3	2
Indo-Iranian									560	28
Balto-Slavic										439

Figure 1 Table showing how many roots each pair of branches share exclusively with each other. The table is taken from (Olander 2022b).

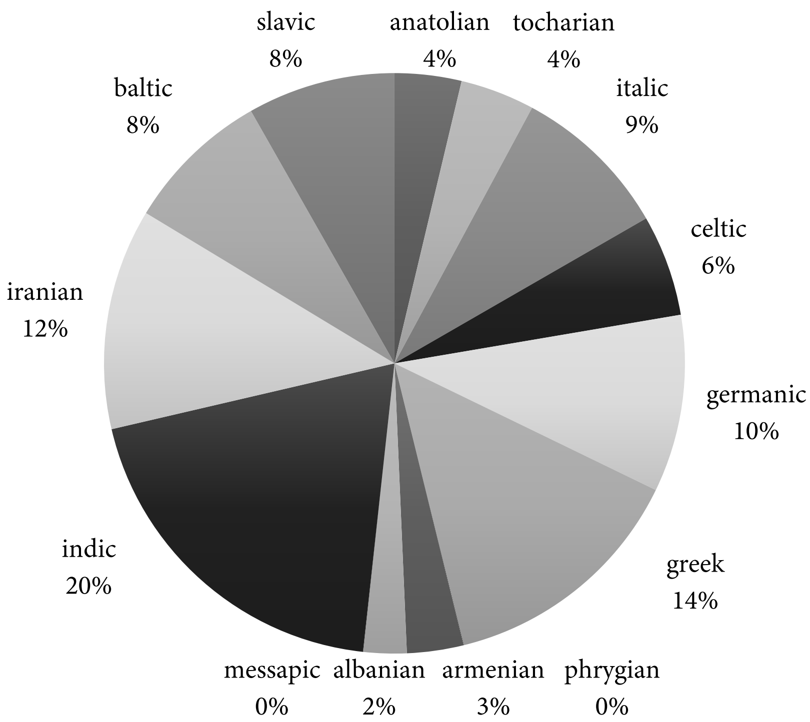


Figure 2 Share of stem-formations in *LIV*² appearing in more than one branch. This chart should serve only as a visualisation.[5]

[5] The chart was originally compiled for a presentation (Olander & Poulsen 2022b) using a preliminary version of a dataset of the *LIV*², and sorted using the script of (Bacovcin & Wilson 2018)

2.2.7. Unrealistic assumptions?

Most of the critique of the family tree model revolves around misunderstandings of the different scopes and aims of the model and its competitor, the wave model (Geisler & List 2013). This will be discussed in the next section. However, the model does apply some gross oversimplifications which does mean that it is not an accurate description of the entirety of linguistic history, only of linguistic descent, defined as

> A language (or dialect) Y at a given time is said to be descended from language (or dialect) X of an earlier time if and only if X developed into Y by an unbroken sequence of instances of native-language acquisition by children
>
> (Ringe, Warnow & Taylor 2002: 63)

When we view linguistic ancestry from far away, this assumption does make sense, but the closer we get to the actual innovation, the more absurd it becomes. innovations do not occur over-night in entire speech communities, and speech communities do not split in half once some of the speakers pick up a new word or palatalise their velars. In fact, the clean splits with no subsequent contact are almost impossible. Even across the most hostile distances and terrains, humans stay in contact. The bifurcating model is, however, practical – and more realistic than automatically assuming a trifurcating split; if one clean split is impossible to imagine, two *simultaneous* clean splits are too (Jacques & List 2019: 143; Pellard, Ryder & Jacques forthc.: 4; Hale 2007: 238–9).

When it became clear from studies in synchronic dialectology, and later sociolinguistics, how innovations spread across the community like wings in water, the wave theory revolutionised the thinking of linguistic innovations (Campbell 2020: 242–51). Innovations begin, as all language, among people and spread along social lines, "but an innovation becomes part of the history of the language only when it spreads through the network to become a stable feature in the speech of a group of speakers" (Ross 1997: 215).

Attempting to combine that realisation with the tree model gives rise to new complications. If all language change begins as dialectal diffusion, and all language contains variation, what is it we reconstruct as a node in a family tree, and how realistic is it to distinguish between shared, parallel and converging innovations

(Jacques & List 2019: 144–8). Since most linguistic variation is traced back to an earlier state anyway, it is not unreasonable to want to model this relationship topographically. With many synchronic, closely related dialects, the picture becomes muddy in the family tree model, because it is only a model: a simplified representation of reality (See also Kalyan & François 2019).

What makes up a node in a tree and what constitutes a language are also important assumptions, but there are rarely satisfying answers. If the reader finds this vague, I will stress that something similar applies to every unit in linguistics: defining the boundaries of a phoneme, of a morpheme, a word, a sentence, an utterance, a dialect or a language is not easy and self-explanatory. The fuzzy boundaries of the elements that we are describing do not invalidate studying them, but it does require openness about the limitations and should urge us to consider our implicit assumptions and be explicit about our stance and starting point.

Extrapolating from the following quotes, it will suffice for now to instate a working definition:

> **a proto-language or a node in a family tree represents a speech community so close-knit that innovations can spread across it**

Ringe et al. define it as follows:

> So long as the phonology and morphosyntax that have to be posited for any node are internally consistent (or nearly so), that node can be taken to represent a group of closely related and mutually intelligible dialects – a genuine linguistic unity, in fact a speech community in the broad sense.
>
> (Ringe, Warnow & Taylor 2002: 107)

For Heggarty et al., who do not work with morphosyntax and phonology, but rather use shared cognacy as a proxy for linguistic descent, a node (the point at which the split of lineages occur) as follows:

> A split does not just correspond to the major difference between discrete, mutually unintelligible "languages." Rather, lineages must in principle already be split from each other for them to be free to start developing differently.
>
> (Heggarty et al. 2023a: 3)

Accordingly, the do not examine or find the innovations or splits themselves, but use the fact that a lexical innovation did *not* sweep across the entire set of languages as evidence for the languages no longer innovating together. At this point, it could perhaps also be helpful to supply Heggarty et al. with the very pragmatic definition of Clackson (1994: 17): “The difference between dialect and sub-group is therefore one of time and degree”. Similarly, in a paper on dialect networks, Ross points out that the differences between the tree and the wave approach are not as grave as it may seem at first glance:

> the circumstances in which an innovation with diagnostic substance is copied to *all* other lects of a linkage are limited. The linkage needs to be close-knit. By this I mean that speakers of each lect are in regular contact with speakers of at least one other lect and the lects are perceived by their speakers as ‘the same language’. These circumstances are likely to occur only rather early in a process of differentiation, so that the linguist’s inference of a common ancestor is likely to be correct
>
> (Ross 1997: 225)

Importantly, Ross remarks that these circumstances in which language diverges gradually are more likely to occur early in the process of differentiation. While it may be impossible “[i]n practice in some cases […] to decide with certainty whether a shared innovation made by two languages is due to an earlier subgroup or dialect relationship” (Clackson 1994: 17), over the course of time, these differences are levelled. Perhaps because the intermediate languages of the dialect continuum die out (Garrett 1999), or because the dialects diverge so much that the speech communities are no longer “close-knit” enough or “in regular contact” to innovate together. The distinction between a Proto-Language and a Common Language (Olander 2022c: 8; 2015: 18–21) is also relevant at this point, the difference being whether the innovations swept across the entire speech community.

Finally, the role of contact – especially among still-diverging dialects – should be addressed, which causes us to return to the question of when an innovation can be considered the same. Borrowings are a linguistic reality, especially in the lexicon, which is one of the reasons it is difficult to view vocabulary removed from culture and geography (Clackson 1994: 12). Borrowings of morphosyntax do occur (Thomason & Kaufman 1988) are, however, much rarer, and require large-scale community-wide bilingualism (Ringe, Warnow & Taylor 2002: 68). Among still-diverging dialects, this is present, as elaborated in the quotes by Ross above, but the role of contact has often been overemphasised (Ringe, Warnow & Taylor 2002: 107, esp. n. 34). At least, creolisation, hybridisation and dialect admixture should not be invoked for prehistoric languages without considering code-switching and large-scale influx of L2-speakers in a speech community.

Accordingly, we might do the model justice by imagining the family-tree metaphor more less in terms of the genealogy of a single person (the EGO of the tree), and returning to biology where there is genetic variation in the ancestral population, but it still makes sense to talk about genetic ancestry between species – at a distance. Humans and whales share a common ancestor, albeit not a single individual by a species.

2.3. Waves, not trees?

2.3.1. The wave model and the Meid model

I cannot repeat the centuries-old and often too acidic debate between two apparently irreconcilable approaches to recovering linguistic prehistory and language relationships in entirety, but it is necessary to discuss the dialect geography-based wave-model and Meid's (1975) time-and-space model (see Article 1) in relation to the family tree model.

There is no doubt that the tree-model is a gross oversimplification of linguistic history. While it is simplistic, it has the great advantage that it can produce or test concrete hypotheses that can be compared to other types of evidence (Pellard, Ryder & Jacques forthc.: 27; Olander 2018). On the other hand, accepting for a fact that linguistic descent is ill-described as a tree tends to leave a methodically uncomfortable space for idiosyncratic and unfalsifiable hypotheses. It is not per se a problem for the scientific community that competing hypotheses exist, in fact it is a key motor in driving the field forwards. However, when these competing hypotheses are mutually exclusive, and they are unfalsifiable in nature, the discussion of them becomes a matter of faith.

Much of the critique of the tree-model revolves around it not depicting a full and realistic picture of all aspects of linguistic history – but this is not what the model intends to do. Drinka's (2013a: 386–7) critique of the tree model not incorporating "horizontal" language change, i.e. contact and synchronic variation, is mistaken. While it is true that the model cannot stand alone as a historically accurate representation of the facts, it still shows part of linguistic reality: the unbroken chain of language descent (Ringe, Warnow & Taylor 2002: 63). Another, less striking, weakness is that most models end up being virtually tree-like anyway. Ironically, all but one graphic depiction in Drinka 2013 could equally well be represented as a tree.

The drawback of the horizontal models of language is the lack of criteria for falsification. Whether an isogloss is interpreted as parallel but independent change or as the result of prehistoric contact or dialectal variation may largely depend on the taste of the researcher (see Article 2, Article 3). What is more, even such a mixed model need not exclude that some part of the linguistic area simply cannot be described as a tree-like diversification in any credible manner; yet such a clear dialect continuum may very well descend from a node in the tree (Ringe 2017b: 7). Just because some languages do not fit a tree very well or present as a tight-knit dialect group, it may be very probable that an underlying tree structure is still recoverable.

2.3.2. Dialect continua in prehistory

The circularity is especially grave for followers of Meid's (1975) time-space model which rests on the theory that languages can be part of different innovating dialect areas at different times. While this may of course be true, it will be impossible to disprove it – also within the same framework. How can we distinguish between competing scenarios? If language A innovated trait X with language B and trait Y with language C, how can we know which innovation occurred first? More importantly, how can we know where the speakers were at the time these innovations happened? Nothing but the very similarities or isoglosses speak for the geographic proximity. Just like the tree-model is geographically underspecified, the wave-model is topologically and chronologically underspecified which leads to speculative conclusions (Clackson 2022: 28).

2.3.3. The nature of *ruki* and *satəm*

For example, Drinka (1993; 2009; 2013a) criticises the interpretation of the *RUKI*-rule and the *satəm*-phenomenon as common innovations advanced by Ringe et al. (Taylor, Warnow & Ringe 2000: 395, 397; Ringe, Warnow & Taylor 2002: 109;

Nakhleh et al. 2005: 176; Nakhleh, Ringe & Warnow 2005: 395). She goes on to argue that Ringe, Warnow & Taylor "mistakenly continue to view the satem and ruki rules as indicative of genetic rather than areal relationship" and claims that "ironically […] the 'satəm' development and the 'ruki' rule are often considered as prime examples of language contact, showing broad areal distribution" (Drinka 2013a: 381–2).

To justify the claim, Drinka refers to Hock (1986: 442–4; 1999: 15), but Hock does not give any evidence or argument to back Drinka's claim. In fact, Hock argues that the *ruki* change is too specific to be independent and must be a shared innovation but that the "incomplete spread" and geographical proximity justify an interpretation as an areal phenomenon. In Hock 1999, the isoglosses are simply presented on an anachronistic schematic map of the geographic attestation or presumed location in prehistory. He does add that there is a difference between the core area and the transitional areas of the satəm change, and that:

> a few of these languages however, including Armenian, seem to have retained some distinctions. In other languages, including Hittite (but not Luwian), Greek, Italic, Celtic, and Germanic, as well as Tocharian, the velars and palatals merge instead
>
> (Hock 1999: 15)

Interestingly, this does not make Luwian a part of the *satəm* dialect continuum which shows that more than geography and distribution of the isoglosses has gone into the "dialect map". It is a tool for visualisation of the degree of overlap of isoglosses, but it proves nothing about the prehistoric whereabouts of the ancestral dialects. He further argues that changes like *ruki* and *satəm*

> cannot be attributed to common tendencies which could very well have **independent** manifestations in different dialects. Rather, they can only be attributed to **common innovation**. That is, we must assume that a <u>single change</u> is responsible for the fact that *s* turns into *š* in this environment in Indo-Iranian and Balto-Slavic
>
> (Hock 1986: 442–4, bold as in original, my underlining)

Here, the argument against a genetic unity of this single common innovation is that the "generality and regularity" of the changes "are not evenly distributed". The only arguments in favour of this distribution being assigned to geography (areal spread) rather than phylogenetics and chronology (differences in productivity over time) is the relative geographic distribution on a topological model. Hock, unlike Drinka, states the facts and does not stretch the argument of geographical proximity out of proportion.

Notably, Hock specifically uses the label “common innovation” form linguistic phylogenetics and even describes *ruki* as “a single change”. While “single change” is less strictly defines and could perhaps be understood as a convergent innovation occurring once in time and space, but not in a genetic entity, the use of “common innovation” and the description of the nature of the isoglosses is not incompatible with the innovations having taken place in an ancestral close-knit speech community.

While the *ruki*-branches are likely to have been spoken geographically close in prehistory, a reconstructed and anachronistic dialect map of all Indo-European branches does act as evidence for it. On the connection of linguistic innovations with real-world time and space, see Ch. 6). It is also puzzling to me that Hock considers the fact that there is no total overlap between the fullest extent and outcome of the sound changes as evidence of areal spread. The obvious alternative, which Drinka argues against, namely that the phenomenon is continuing change after the break-up of a dialect in which the “common innovation” occurred, is not disproven.

On this matter, I must side with Ringe et al. (2000: 176, n. 3): Interpreting the isogloss as one single innovation in the same speech community is preferable to the claim that we are dealing with areal diffusion. Of course, if something clearly spoke against the possibility of interpreting these isoglosses as shared innovations, we would have to view them accordingly. However, it is unproblematic to argue that the cradle of the innovations began as one genetic innovation but continued to gain productivity after the break-up of the (intermediate) protolanguage.

That said, I do agree with the critique that the opposite coding strategies of the two phenomena employed by Ringe et al. seem forced (Heggarty et al. 2023b: 69–71): In order for a language to be coded as having *ruki*, it needs only to exhibit traits of it: Latvian, which presumably later obscured the change by reverting **š* > *s* is assigned a unique state in order to be compatible with the rest of Baltic; whereas

Lithuanian and Old Prussian are assigned the same state as the Indo-Iranian languages even though the domains of the change are not identical (see also Ch. 4).

Conversely, in order for a language to be considered *satəm*, it needs to exhibit both the merger of the PIE plain velars and labiovelars (**K, Kʷ* > **K*) as well as the further affrication or assibilation of the PIE palatovelars (**K̂* > **Ć*). This excludes Armenian and Albanian from the *satəm* core since they only partially merge the velars (by keeping the unvoiced **kʷ* distinct from **k* in palatalising contexts: **k* > Alb. *q*, Arm. *kʿ* before **e i* (Arm. also **ki̯* > *čʿ* (Schmitt 1981)), but **kʷ* > Alb. *s*, Arm. *čʿ* before **e ĭ̄ i̯*).

One could easily deem this inconsistent, but I strongly disagree with Drinka (2013: 385) who thinks that Ringe et al. "fail to notice" that *satəm* and *RUKI* are areally distributed and occur at different times. Although the domains are not identical, it is attractive to point to a common focal point in the languages from where the innovations gained further productivity after the break-up of this latest common ancestor. Claiming the opposite, that the *ruki* and *satəm* developments have no genetic origin but spread only through language contact, suffers the major drawback that it concludes that the speech communities were in geographical proximity in prehistory based solely on the fact that they share isoglosses too vague to have been completed at the same time.

2.4. Computational phylogenetics

2.4.1. Computational approaches to phylogenetics

It is beyond the scope of the present section to give an overview of different computational methodologies. I refer to handbooks and great overviews for this (Nichols & Warnow 2008; Campbell 2020: chap. 16; Pellard, Ryder & Jacques forthc.; Goldstein 2020).

However, as the present work constantly compares "traditional" phylogenetics to computational studies, a brief introduction to the data preparation will be given below. Article 2 also deals with the "matrix format" and its relevance as a heuristic device also for traditional phylogenetics. Article 3 is a phylogenetic experiment using Maximum Parsimony – see this article, and the Preamble preceding it for the choice of this method over Maximum Compatibility or a Bayesian study.

2.4.2. Computational use of lexical data: the matrix-approach

A well-established alternative to the qualitative approaches to lexical data is computational studies where the data input in a matrix format in which comparable cognates derived from the same ancestor are coded according to semantic slots. This means that each token is the answer to the question of what etymon the standard or most "unmarked" word for a given meaning stems from. This format is employed in most computational studies of linguistic phylogeny. Most datasets try to eliminate synchronic synonyms (polymorphic characters), even though it is a very common in natural languages to have synonyms or near synonyms for "basic" concepts.

Organizing the data by etymon as done in lexicostatistical surveys (Tischler 1973; Bird 1982; 1993) instead of by semantic slot would mean that all characters would become uninformative in a maximum parsimony analysis. All languages sharing the cognate-set should be assigned the same state, but since cognates can easily be lost, all languages *not* sharing the cognate-set would have to be assigned unique states (Ringe, Warnow & Taylor 2002: 71 n. 8). Assigning just one, common state for lacking the etymon would force upon the data that the absence is either the original state, or that it is a shared loss in the prehistory of these languages (Taylor, Warnow & Ringe 2000: 396). It is a problem for parsimony analysis that a large group of characters will end up useless because they are uninformative. It is obviously impossible to draw a tree if a character has the same state or unique etyma everywhere; but the real problem lies in the fact that each character must have more than one non-unique state. Unique states can be places anywhere in the tree, so in order for the character to be informative, it needs two competing group-forming etyma (Taylor, Warnow & Ringe 2000: 398).

When two or more branches share a cognate coding in a matrix, it only means that they have inherited an etymon in a given meaning from a common ancestor – if parallel innovations have been ruled out. Further data is needed to conclude whether that ancestor is intermediate, and the cognate set therefore represents an innovation of a subgroup, or if the ancestor is the oldest reconstructible state, and the token simply represents an archaism. On the contrary, when one or more branches exhibit a different coding in the matrix, it may correspond to a wide range of innovative processes in the real spoken language: for instance, the etymon could be lost, a semantic development could have shifted the meaning, or the etymon might never have existed in that branch at all.

2.4.3. "The matrix does not discriminate"

While it seems unnecessarily rigid to most traditional comparativists, the matrix does have some embedded strengths. The comparable language data will be equal. The matrix format will not favour branches that have more attested material or are more thoroughly treated. It will also combat confirmation bias for the selected tokens, since all the analysed languages or branches will be included. This approach is usually employed to lexical data only, but there are attempts to combine the bias-combatting matrix and data which is deemed more reliable, like phonological innovations and morphological isoglosses (Ringe, Warnow & Taylor 2002; 2007a; 2007b). This dataset is also unique in that it combines different types of evidence. However, because of the computational method (maximum parsimony) employed by the Ringe group, they rely more on the quality of the phylogenetic signal of the input data than on the number of characters. However, especially higher-order subgroups risk resting on very narrow grounds if the dataset is too small (Greenhill, Heggarty & Gray 2021: 236). This is what is discussed and evaluated under "the robustness of the tree" by Ringe, Warnow & Taylor (2002: 98–104) in a section that is quite unique in comparing the computational result with the linguistic reality it requires. Technically, adding more languages (taxa) to the matrix will increase the number of informative characters – as long as they are related, they will share a state with someone somewhere (Ringe et al. 1998: 402 n. 9). In practise this mostly only solidifies lower-order subgroupings further.

It is important to keep in mind that a data visualisation (like a phylogenetic tree) or a data presentation (like a coded matrix) is never the full picture of a complex linguistic reality. One point that cannot be iterated too many times is that lexical data in a matrix organized according to semantic slots do not have directionality. This means that when we observe more than one coding for a given token, they are most likely not formed at the same stage. Accepting this type of polymorphism as synchronic and linguistically real would translate to projecting all synonyms back into the parent language (Warnow, Ringe & Taylor 1996: 317). The original state could very well be lost everywhere and thus not represented, or it could be preserved as an isolated state anywhere in the tree. The distribution – or even amounts, as Bird

(1982: 11) claims – of cognates do not *per se* reveal the original state, no matter if it is wide or narrow.

2.5. Types of data

2.5.1. A scale of significance

It is possible to place the different types of linguistic data on a scale of their value as arguments when drawing phylogenetic conclusions. This is of course a gross simplification, and there is a great caveat in this presentation because many types of data can of course be used if weighted and substantiated accordingly. However, the scale is a handy visualisation of the underlying paradox, namely that there seems to be an almost perfect inverse ratio between the quantity and the quality of the data. On the one hand, there is an abundance of comparable material in the open-ended categories like the lexicon. On the other hand, with the increase in qualitative potential, the potential amount of available data decreases.

Figure 3 The scale of the significance of different types of linguistic data for linguistic phylogenetics. Originally made for a presentation (Poulsen & Olander 2021).

2.5.2. Syntax

It is clear that Syntax is very prone to synchronic variation and contact-induced change (Hock 1988: 561). It is also a category that is inherently tied to pragmatics, and in the case of the material that survives from the ancient Indo-European languages, be it inscriptions, metric hymns or glosses, it is very far from natural speech. There is no apparent directionality in change. In terms of concord, the rules that govern verbal rection and case functions, much relies on what is attested (Clackson 1994: 23–3). Finally, much what is reconstructed for Proto-Indo-European, which is arguably necessary to point out innovations, is reconstructed on the basis of Greek and Indo-Iranian anyway (Keydana 2018). While all languages to some extend have word-order, it is rarely informative for phylogenetic analysis.

2.5.3. Lexicon

2.5.3.1. Lexical innovations

Traditionalist approaches collect lexical *innovations* including semantic shifts and derivations, whereas computational approaches based on cognacy only code lexical *replacement*, which is arguably a form of lexical innovation, but not the only possibility. Traditional comparativists do not rely much on lexical data because it is prone to prone to change due to contact. This is the golden inheritance from the Neogrammarian principles: if we want to establish sound laws and reconstruct linguistic prehistory, it goes without saying that we can only rely on inherited material: Armenian does not become an Indo-Iranian language by obtaining Iranian loanwords.

Innovations in the lexicon are often lumped together, but they represent quite a range of linguistic realities. Nevertheless, many of these end up as similar changes of state in the matrices of computational studies on cognacy. In the following, I rely on the types and examples given by Clackson (1994: 23–4), and the material compiled for Article 1. There is, therefore, a certain overlap between the two.

It should clear by now that linguistic subgrouping is to be argued on the basis of shared innovations. Unfortunately, it is surprisingly often difficult to point to the innovation when examining the lexicon (Ringe, Warnow & Taylor 2002: 104–6). An additional problem is that the category is somewhat open-ended. There is not a fixed amount of lexemes in a language, and there are few linguistic factors – other than replacements and loss – that limit what survives to the historical records. Accordingly, we easily run the risk of letting oneself be hypnotized by beautiful etymological connections between the large-corpus languages. However, such connections tell us very little when the feature (be it the meaning or the lexeme) is not identifiable in the other languages.

On the other hand, if we limit the exploration to a fixed set of meanings (the hotly debated universal vocabulary), we might have identifiable features, but there is rarely an inherently archaic or innovative nature in simply sharing a cognate in a basic meaning (Ringe, Warnow & Taylor 2002: 104–6).

Many such unique isoglosses can be found in Porzig (1954). One example could, in Porzig's (1954: 159) formulation, be the distribution of the lexemes meaning 'word' where Greek and Indo-Iranian share the etymon **u̯ekʷ-es-* as opposed to the derivatives of **(h₂)u̯erdʰ-* shared by Germanic, Italic and Baltic. Such an observation cannot stand alone, especially because the original state is opaque: we cannot know when the polymorphy occurred. Polymorphy is the computational equivalent of synonymy: that more than one lexeme fit the same semantic slot at a synchronic level – either in an attested language state or at a reconstructed level. In the case of the word for 'word', the polymorphy could have occurred because of an innovation in one branch, or it is possible that we should trace both preforms back to the parent language with slight semantic variations (Warnow, Ringe & Taylor 1996: 317). For some linguistics, the assumption is that the bulk of such distributions will mimic the split-up of the family; but others recognise that the lack of directionality in lexical data. This fact that a distribution does not equal cladistics is neatly summed up by Ringe, Warnow & Taylor (2002: 104): "That is because it does not *force* either subgroup; it is possible to accommodate this distribution of states in a tree in which one or the other of those larger subgroups is not posited".

Since the Indo-European languages all descend from the same proto-language, and it is quite unlikely that this language exhibited an unusual amount of synonyms, the distribution of lexemes must, in some way, reflect innovations. But since we often cannot pin-point when, how or even who replaced what, we therefore need to ask ourselves if replacement of a lexeme in a particular semantic slot truly is a meaningful proxy for linguistic change.

The trickiest type of lexical innovation is the Pandora's Box, which Clackson labels "creation of 'new' roots". Clackson (1994: 24) exemplifies this by Gr. *ἔχθ-* 'hatred' and the contrast between its many native-looking derivatives and the lack of attestations of this root in other branches. It gets much less transparent if the innovation belongs to the debatable domains: onomatopoetic creations, iconic creations and borrowings.

Onomatopoeic creations should in theory be recognisable, but the judgement and the analysis as such is ultimately subjective. This is a known fact, and word lists of semantic concepts avoid domains that are likely to be onomatopoeic. However, in the Swadesh 207-list, the token "to laugh" occurs; and while the concept might be of human in nature and not culture-specific, the etyma might not be as robust as one could hope for: While all Germanic forms go back to PGmc. **χlaχjan-*, the origin of this form seems suspiciously onomatopoeic (Seebold 1970: 258). In a very different semantic domain, a derivate form an interjection has also been used to explain the word for ‘wild cat’ in Lithuanian: Smoczyński (2018: 1667) argues that *vilpišỹs* is rather an agent noun formed to *vìlpš* used of sudden and quick movements than being derived from *$h_2$lou̯p- ‘fox’ or **lup*- ‘fox, marten (?)’. For now, it is not important what the correct etymology is, but to note that such wildly different analyses impact the cognate set.

Iconic creations might be the most difficult of all. How do words spontaneously come into being? The consequence of assigning a new number – corresponding to a new and unique cognate class – to a language in a matrix without knowing the directionality is technically proposing this etymon as an equally good proto-form for the entire family. For those of us more interested in language relationships than in mathematical possibilities, such isoglosses of opaque semantic or morphological developments present new headaches. The handbook treatment of the possible types of formation of new words (e.g. Campbell 2020: 103–13) gives many opportunities of how words can arise *ex nihilo*. Usually, the examples come from a historical or modern context, where we can trace exactly when a term was coined and how it spread from a creative author or the name of a person, a location or even a brand. Imagine a prehistoric *Sandwich*, *Watergate* or *Tempo-tuch* and the problems they would cause for cognate analysis, especially how we would distinguish them from a borrowing from unknown sources.

2.5.3.2. Lexical semantics (innovation in meaning)

This type of innovation occurs when an inherited lexemes changes its meaning and replaces another one in a semantic slot, like Gr. *θῡμός* ‘soul’ being formally identical to Ved. *dhūmá*- ‘smoke’, but having undergone a semantic shift. We know that the lexeme is inherited and reconstructible, and descendants of the semantic shift would share the replacement. Even when we understand the directionality of a change, that is what is the innovation, and what is the archaism, there is a high risk of parallelism in semantic shifts:

> Lexical characters (vocabulary) are actually the least reliable, because parallel semantic development is rampant – words meaning ‘person’ often come to mean ‘man’ and then ‘husband’, for instance
>
> (Ringe 2022: 54).

Consider for the sake of the argument the fact that both Gr. *ὀστέον* 'bone' and Ved. *ásthi* 'bone' secondarily acquired the meaning 'kernel of a fruit' long after the split-up of the clades and the earliest attestations. The likelihood of most semantic shifts and the high risk of parallelisms is the major drawback of such lexical arguments.

Like with phonological mergers, some semantic innovations are impossible to reverse. One famous example of such a claim is Schmidt's (1992: 101) analysis of the semantic development of the root **i̯ebʰ-*. This verb presumably preserves a most likely older meaning in Tocharian where *yäp-* means 'to enter', whereas the meaning is purely sexual 'to penetrate' (occasionally also obscene) in Sanskrit, Greek and Slavic (Winter 1997: 185; 1998). It is impossible to imagine this shift in reverse, and we can therefore be certain that it happened after the break-off of a clade. However, we cannot rule out that it happened multiple times independently. In many and most other cases, a semantic development is too trivial or too easily reversible to base a phylogenetic argument on.

Even this argument is not flawless, however, because it ignores the potentiality of original polysemy. lexical innovation does take place a neat replacement of a number in a spreadsheet even though they are coded as such in studies of wordlists. As a reviewer pointed out to me (Poulsen 2025a: 119), a semantic shift is not an immediate replacement of one meaning with another, but it goes through a period of polysemy. In the case of **i̯ebʰ-*, the final shift in meaning whereby the decent and puritanical connotation was finally lost, is the only irreversible semantic change. The root could have been polysemous for millennia.

2.5.3.3. Lexical derivation (innovation in form)

New words – and new cognates – also arise through word formation (compounding and derivation). Since innovations tied to derivational morphology are inherently linked to the lexemes they derive new forms from, it should be considered a type of lexical innovation, not a type of morphological one (*pace* Olsen & Thorsø 2022: 211). However, this only goes for the specific combinations of lexemes and suffixes, not for the suffixes and the categories they represent themselves. These are, of course, a type of morphology – and even one that is often "under-exploited" in linguistic phylogenetics (Weiss 2022a: 121). Naturally, this distinction is not easy to make – consider, for instance, the discussion of the East Iranian adverbs derived from deictic pronouns in Article 3 (2.3.10):

On the one hand, the innovation of the category of deictic local adverbs and the productivity of the combination of the suffix *-*θra* (and others) with the three deictic pronouns **ima-*, **aita-*, **au̯a-* in Sogdian, Bactrian and Khotanese, is a morphological one, but on the other hand, individual examples could easily be lexicalised and fossilized forms. When Khotanese shares *vara* 'there' with Sogd. *ʾwrδ* and Bactr. *οαρο* < **au̯a-θra-* despite the pronoun **au̯a-* otherwise being lost, is it then evidence for a lexical Gleichung or is it a relic of a morphological innovation?

Along the same lines, methodologically speaking, Clackson concludes on the shared innovation of the *e*-grade athematic *nu*-present **u̯es-nu-* 'clothe' that it is rather "the continuation of a morphological process of Proto-Indo-European" (and therefore a parallel, not shared lexical innovation), since also Greek and Indic uniquely share cognate pairs like Ved. *dāśnóti*, Hom. **δηκνύμενος* with the same morphological pattern (Clackson 1994: 180). Innovations by means of productive suffixes stand the strongest not just when form, function, meaning and the relative chronologies align, but when they form conglomerates that are unlikely to occur anywhere else.

This raises the biggest problem for using derivational lexical evidence: The languages examined ultimately go back to the same state, and it is therefore very difficult to rule out parallel innovations by means of the same inherited productive suffixes. If there is no accompanying change in meaning, the motivation might be difficult to grasp, as in Clackson's example of a purely derivational innovation: Gr. *καρδίᾱ* 'heart' is derived from the root noun found in Hom. *κῆρ* and has kept its meaning, but has been extended with a productive derivational suffix.

It may also be the case that there is no change in meaning, but that the suffix is no longer productive. While Ved. *hásta-*, Av. *zasta-* and OP *dasta-* obviously go back to the same root morpheme as Hitt. *keššar*, Gr. *χείρ* and many others, only the Indo-Iranian languages share the extension of PIE *$\hat{g}^h$*es-r̥-* 'hand' into *$\hat{g}^h$*es-to-* > *$\acute{j}^h$*asta-* Ringe, Warnow & Taylor (2002: 82–3). As the suffix is somewhat opaque, it may be easier to posit a single change rather than multiple.

Occasionally, the innovation is both derivational and semantic, such as Lat. *filius* 'son' shows both innovative morphology (forming the noun *d^hh_1*ilio-* 'suckling' to the root *d^heh_1*i̯-*) and a semantic change from *'suckling' to the attested 'son'.

It is always possible to find such unique shared features by languages that descend from the same ancestor, which in cases where we cannot determine the directionality of the lexical replacement leads to the framing of random common archaisms as diagnostic for a subgroup. One such example is again Porzig (1954: 159), bringing to light that the *s*-stem **u̯et-es-* replaces an ancient root noun **u̯et-* in the same meaning, but only preserves the meaning 'year' in Indo-Iranian (*tri-vatsá-* 'three-year-old ox or calf' (VS+), *vatsará-* 'year' (VS+)) and Greek (*ἔτος*), elsewhere it only means 'one-year-old animal', cf. Ved. *vatsá-* 'calf' (RV+) and Lat. *vetus* 'old'. However, this tells us nothing more than that these two languages are related. The old root noun is, as Porzig notes himself, preserved in the Greek adverb *πέρυσι* 'last year, a year ago', and the semantic shift of the *s*-stem can be seen even within Indic. Finally, the Vedic forms listed are derived from the *s*-stem, but this goes for most other clades as well. Especially in Anatolian where we find forms that must be derived from or build upon the "Graeco-Aryan" *s*-stem: HLuw. *usa/i-* 'year' and Hitt. *šāu̯itišt-* 'a cow younger than one year' (Kloekhorst 2008: 739, 1015). Summing up, this isogloss is probably either just an archaism or a parallel semantic innovation, and therefore not interesting for the subgrouping of the family.

Another case could be the *no*-extension of the adjective **su̯eh$_2$du-* 'sweet' (**su̯ādu-* in his notation) meaning 'lust' underlying Gr. *ἡδονή* and Skt. *svādanam* (Porzig 1954: 159). Porzig claims that neither the formation nor the semantic shift is found elsewhere, but the evidence is problematic. First of all, the forms do not go back to the earliest stages of the languages. Within Indic, there is even the issue that the neuter noun *svadanam* 'the act of tasting, licking, eating or enjoying' is only attested in lexicographers; whereas Vedic only has *svā́danam* 'seasoning; making [food] savoury' – used only once, of Agni (R̥V 5.7.6). The difference in vowel length in the first syllable of the Indic forms might be explainable (Lubotsky 1980: 133), and the lack of Brugmann's law in the second syllable might reveal a Hoffmann-formation. That is, if the forms can rightly be equated and not explained as younger, independent formations. These problems show how such arguments are difficult at best and misleading at worst.

2.5.3.4. Lexical (root) cognacy and the distribution of lexemes

Traditionalist phylogeneticists have been sceptical towards this choice of data, going all the way back to Leskien (1876) and only reinforced by the scepticism towards the works of Greenberg and lexically based discipline of glottochronology. Modern computational phylogenetic methods are built upon statistical biology and designed to work with large quantities of streamlined and standardised data that can be represented in relatively simple manners, like proteins in DNA sequences.

The lexicon is, however, the preferred data choice for the prominent Bayesian studies (Heggarty 2021). While many computational studies do not dwell on the choice of data, Greenhill, Heggarty & Gray (2021: 232–242) have spent quite some energy defending the use of the lexicon. A prime argument is that the term is somewhat a "misnomer", because "[c]ognacy assignment, when properly performed, integrates and rests on all of the data, methodology, and findings of orthodox comparative-historical linguistics, not least in phonology and morphology" (Greenhill, Heggarty & Gray 2021: 235). Chang et al. (2015: 200–1), on the other hand, explicitly adhere to the "lexicon" to avoid the interdependence of phonology and morphology.

The most prominent argument for Greenhill, Heggarty & Gray (2021: 236–7) is that lexical data offer "not cherry-picking but 'safety in numbers'". Their methods require more tokens than what is possible with phonology and morphology. Obviously, there are more lexemes than morphemes or possible phonological developments available; but in order to achieve compatible and streamlined data, the datasets have to be limited to somewhat cherry-picked tokens within the basic vocabulary. These are the least prone to be borrowed – within one of the most borrowable linguistic categories.

Much has been written on word lists and whether "basic", culturally universal concepts actually exist; and that it is extremely difficult to come up with a word list of cross-linguistically or anthropologically universal semantic concepts (Pereltsvaig & Lewis 2015: 71–2, with refs. Kassian et al. 2010; Tadmor, Haspelmath & Taylor 2010). Especially one that is long enough for the statistical methods to be reliable (Greenhill, Heggarty & Gray 2021: 236). The Ringe-dataset differs significantly from the others in that it not only acknowledges its language family biased approach, but also embraces it by adding specifically Indo-European cultural terms to it. At first glance, it seems like a great idea to engage with the context in which the languages exist, but it does lead to the emergence of a possible bias towards the traditionally more dominant languages and cultures behind the reconstruction of a non-stratified Proto-Indo-European.

This idealist presentation of the prevalence of the lexicon does hide some pragmatic facts. The choice of lexicon is also a simple question of availability. As Chang et al.

(2015: 201) put it: "[…] because large lexical data sets are available, most statistical work on language relationships analyzes lexical traits". This quote reveals that it is not just because lexical data offer "safety in numbers" instead of "cherry-picking", or that the rise of computational power requires a completely different approach to the data itself, and not just its organisation. Studies rest upon each other, and once the enormous task of compiling and completing a dataset has been executed, it finds its way into the mainstream.

However, choosing an available datatype does not mean that one does dot run into problems similar to assessing whether an innovation is identical or shared. The organisation of the data in a matrix format offers a lot of problems in itself. Scarborough (2020) has discussed many of the practical issues – a prime example is Mod. Gr. _δεν_ 'not', which descends directly from an unbroken functional continuity of negations, where the lost the very morpheme that used to carry the meaning of the semantic slot it covers has paradoxically been lost continuously (Scarborough 2020: 195–6): Mod. Gr. _δεν_ 'not' < Att. _οὐδέν_ 'not' < Gr. _οὐ δ'ἕν_ '(and) no-one' with the negation Gr. _οὐ(κ)_, Myk. _o-u-_ < _*ou̯-ki-_ < _*ou̯k^wid_ < _*né $h_2$oi̯u k^wid_ *'not in a lifetime'.[6]

Much of Greenhill's defence of the choice of data is very controversial to most traditionalist linguists. While it is important for him to distinguish the works of Bayesian phylogenetics from the lexicostatistics and glottochronology of the past, one major premise is inherited from these, and that is the assumption that lexical replacement is constant enough over time to calibrate dates on (Greenhill, Heggarty & Gray 2021: 240–2). There is not such claimed constant for phonological, morphological or syntactic innovations, which is why lexical data is of great importance to the second half of the bayesian project: absolute dating of the speciation events and proto-languages of the established tree. This even means that some authors find it less relevant to isolate loanwords – a loan is after all lexical

[6] The exact etymology and the connection with Arm. _očʿ_ is not crucial for this argument; but it is indeed for Graeco-Armenian as the innovation involves lexical semantics, syntax and pragmatics (Cowgill 1960; Clackson 1994: 155–6; van Beek 2022a: 195; Fortson 2020; Ringe, Warnow & Taylor 2002: 103). Cf further the more or less speculative suggestions along the same line: Alb. _as-_ < _*(ne)_ _*$h_2$ói̯u k^wid_ (Hyllested & Joseph 2022: 226; Joseph 2005; Schumacher & Matzinger 2013: 241), Lat. _haud_ < *(_ne-)χe ái̯ud_ < _*ne=g^he $h_2$ái̯ud_ (Garnier 2014: 104), TA _mā ok_ and TB _mā wkä, māwk, māk(ä)_ < _*meh$_1$ $h_2$ói̯u k^wid_ (Fellner 2021).

replacement (Chang et al. 2015: 205). If we do not buy into this premise of lexical replacement like radioactive decay, or if we as phylogeneticists chose to remain agnostic about the dating of the branches on the basis of shared cognates (Pereltsvaig & Lewis 2015: 95ff.; Ringe 2022: 59), the choice of lexical data becomes much less self-explanatory.

2.5.4. Phonology

Despite it being the most intuitive example of linguistic change, sound change is surprisingly uninformative; phonetic changes are very often universal and trivial, and true irreversible phonemic mergers are surprisingly rare (Ringe, Warnow & Taylor 2002: 66–8; Hoenigswald 1966: 12). While phonological and phonetic changes are well-studies, and the relative likelihood of sound changes can be explained by or backed up by (experimental) phonetics, and acoustics, typological surveys (Kümmel 2007; 2012), phonological theories and “natural classes” (Vennemann 1974), this known directionality comes with a cost: It is rarely unidirectional. As long as there is no phonemic merger, it is possible, but not always likely, to obtain the original distribution again.

A classic example of this problem is the assimilation of **p…k*^{w} > **k*^{w}*….k*^{w} in Italo-Celtic. Only the root constraint against to equal stops in PIE make it certain that the innovation is in fact on the Italo-Celtic and not Indo-European side, but it cannot be ruled out that other languages took part as well but undid the assimilation by a dissimilation later, since there was no contrast between **p…k*^{w} and ***k*^{w}*…k*^{w} (Hoenigswald 1966) – or that this is the explanation for Lat. *quercus* ‘oak’ < **k*^{w}*erk*^{w}*us* < **perk*^{w}*us* rather that the *βουκόλος*-rule bled the assimilation (Weiss 2022b: 103; Kölligan 2023: 324–5).

Similarly, it is impossible to describe the prothetic vowels of Armenian and Greek as a shared innovation in terms of phonemic mergers. Since **h̥*$_1$ *h̥*$_2$ *h̥*$_3$ all become **a* in Armenian (Olsen 1999: 762–4), but **e a o* in Greek, there is no structural difference in the reconstructible shared prestage. /h$_1$ h$_2$ h$_3$/ could have been realised [ə$_1$ ə$_2$ ə$_3$] (whatever that means, perhaps [ɘ~ə ɐ ɵ]) or with an epenthetic vowel [h$_1$ə h$_2$ə h$_3$ə], but it is not before the loss of the laryngeal and the merger of the (coloured) epenthetic vowels the innovation is complete and irreversible, and backformation can be ruled out (see Chapter 3).

This is, perhaps, an extreme example. It may be clearer in the classic complex of palatalisations and the vowel-merger in Indo-Iranian. It is only when the vowels **e a o* merge in **a* that the distribution of **k* and **č* is no longer automatic. Importantly, another linguistic change can *feed* or *bleed* an innovation (Wandl & Thelitz 2024: 2–5), that is produce input that undergoes the innovation or block it from happening. Therefore, while it is often difficult to rule out backformation, phonemic changes are powerful for establishing the relative chronology and the directionality of other changes.

2.5.5. Inflectional morphology and morphosyntax

Most emphasis is usually given to the realm of morphology, and especially to inflectional morphology for the following reasons. First, it is the least prone to borrowing, thus eliminating the risk of drawing cladistics conclusions from horizontal language change (Nakhleh, Ringe & Warnow 2005: 387). Second, the number of categories and their formants are in principle unlimited, and the risk of chance resemblance is therefore low. In related languages, unstable paradigms or other types of systemic pressure might lead to the same levellings, but parallel innovations in form and function on a grammatical level are not likely to be frequent. Therefore, shared innovations in morphology are more likely to be significant. Hock's (1988: 561) caveat that the general shortness of morphemes should increase the possibly of chance resemblance is rather weak: It focuses too much on the formal and not on the functional aspect. While there is a relatively high chance of inheriting a string of, for example, two phonemes, the innovation of assigning a specific meaning and a specific morphosyntactic function to exactly this phonetic phonemic combination is low.

Morphological data does face some major issues. One is that "the preference for morphological paradigms belongs only to a best-case wish-list. In reality, extensive paradigms are simply not available in all language types. With isolating languages, the comparative method has to make do without them" (Greenhill, Heggarty & Gray 2021: 235). Another issue is the smaller amount of comparable items, not only if the purpose of their counting is a traditional search for isoglosses or adaptation into characters for computational studying, in itself, but also the internal complexity of each item make the results depend a lot more on each individual analysis and each researchers reconstruction of the Proto-Language. This leaves room for subjectivity sneaking into the results (Clackson 2022: 23). Every analysis will to some extend be subjective. The greater risk does not come from the data analysis, but rather from the data selection itself. The core of the critique is that researchers induce relationships from the factors that spring to mind or they happen to know of find valuable themselves – "by inspection" (Clackson 2022: 21–2).

This risk is what is often referred to as cherry-picking (e.g. Greenhill, Heggarty & Gray 2021: 236–7). While this label is unnecessarily derogatory and undermines the general high quality of the grammatical material, it does place some main issues in the spotlight. First, the agreement in form and morphosyntactic function is rarely exact, and that gives some unfortunate wiggle-room. Second, we might be comparing languages that have so radically different histories that the items once shared by them could be lost in one altogether. Third, when this is the case, some forms might be analysed as fossilised forms – and how should they then be treated? Can relics, isolated adverbs or even hapax legomena truly be taken as evidence for or "Überreste" from lost innovations an earlier stage?

The fate of the locative plural ending can serve as an illustration. The Albanian ablative plural ending *-sh* and the Greek dative plural in *-σι* (restored after the consonant stems in the vocalic stems of alphabetic Greek, but Myc. *-hi* after vowels (Sihler 1995: 263–4; Bernabé & Luján 2020: 217–8)) are obviously related to the Vedic locative plural in *-su*. Greek and Albanian might share the introduction of the vowel **-i* into the plural from the locative singular (van Beek 2022a: 193; Hyllested & Joseph 2022: 237 n. 18). Therefore, this morpheme would be relevant for an investigation of their relationships within the Indo-European language family. However, because of the case synchretisms of the two Balkan languages, the functions can no longer be equated: The semantics of the suffix has been broadened to other functions, no matter the label assigned (the Greek dative is much more than a dative). The massive phonetic reduction of Albanian also complicate the situation. Such a comparison will automatically contain several assumptions and preconceived ideas.

2.6. The phylogenetic approach to reconstruction

The previous preliminaries might seem banal to many comparativists, but accepting the reality of a binary-branching tree has consequences for all aspects of linguistics reconstruction. In order not to project features all the way back into the earliest proto-language, it is necessary to evaluate carefully whether each item is truly a feature of the oldest stage, or if it could be an innovation at a later node in the tree. It has been highlighted in recent research which has made it increasingly clear that we need to rethink reconstruction based on unresolved trees or unstratified data (Olander 2018; 2019a; Jacques & List 2019; 2022a; Søborg 2020; Goldstein 2022). For each linguistic unit, it is necessary to take explicit stance on how far back it can be reconstructed. This means that the nodes within the family tree should be treated as linguistically real intermediate proto-languages. They may not have been completely uniform, but they could be treated as tightly knit speech communities capable of undergoing the same innovations to such an extent that we probably would not be able to recognise the distinction.

For the archaeolinguistic endeavours (Chapter 5) this has immense consequences (Gąsiorowski 1999: 55). Simply projecting whatever is inherited at branch-level back to the ultimate proto-language skews the analysis of the shared culture associated with this ancestral state. Such data has to be sifted not just chronologically and geographically (as often done, e.g. Mallory & Adams 1997; 2006), but also topologically (Ehret 2011; 2015; Olander 2019c; Kroonen 2021; Kroonen et al. 2022).

3. “Indo-Greek” isoglosses

3.1. Lexical isoglosses

In the spirit of Clackson (1994), one could perhaps expect this thesis to continue with a list of proposed isoglosses between Greek and Indo-Iranian. However, as these languages make up almost half of the attested material, we would run the risk of judging from positive evidence alone.

Much work has been done to collect such material – but, as I show in Article 1 – almost always under a dialect-geographical framework lacking a strict methodology of what can truly be extrapolated from these comparisons (Dehò 1957; Bonfante 1976: 71–93; Porzig 1954: 157–61; Birwé 1956; Zgusta 1957; Schindler 1972; Euler 1979)

3.2. Porzig 1954

Porzig (1954: 157–61) lists only 16 isoglosses, much of which are not tenable, and none of which are significant:

1. **n̥* > **a* also /_C (except **i̯)*
2. *-*i̯es* >> *-*tero*- in younger formations
3. *-*ai̯* >> *-*tai̯* in MP.3sg
4. **u̯eĝʰ*- ‘drive’ builds *s*-stem ‘wagon’
5. **ĝʰolo*- m. ‘bile’ thematic, no -*n*-
6. **pelito-*, **pelitnī*- ‘grey’
7. **u̯et-(e)s*- ‘year’, not ‘yearling’
8. **su̯ād-Vn-V*- ‘lust’ from ‘sweet’
9. **u̯ekʷes*- ‘word’
10. **kani̯o*- ‘young’
11. **selo*- ‘swamp’
12. *-*ĝʰesl*- ‘thousand’
13. **kerbero*- ‘spotted(?)’, not **perk̂*-
14. **ₑresi̯ā*- ‘feindseliges verhalten’
15. **pelek̂u*- ‘double axe’
16. **pāusōn* ‘der (das Vieh) fett macht’

Of these isoglosses, **1** is treated in Ch. 4, **2** and **3** in Ch. 5, **7** and **8** are given as examples above and **15** is treated extensively in Article 2. The remaining are insignificant (lexical roots: **5**, **6**, **10**, **11**, **12**, lexical derivation **4**, **9**) or so speculative that they should be dismissed out of hand (**13**, **14**, **16**).

3.3. *LIV²*

There are 44 lexical roots in the *LIV²* shared exclusively by Greek and Indic and/or Iranian, but as these languages make up the plurality of the material (Indic account for 15% of the total root-attestations in the *LIV²*, Iranian 10% and Greek 14%), I have not pursued them further. Similarly, there are unique lexical derivation in the *LIV²* (see **2.5.1.5.**): Indo-Iranian and Greek share 22 to the exclusion of others, Indic and Greek share 52 and Iranian and Greek share 12. While they may indeed be interesting, it is unlikely evidence for linguistic subgrouping. A quick assessment of these reveal that these unique shared formations most often belong to categories that have disappeared or merged (23% root aorists, 15% perfects, 8% *s*-desideratives) in most other branches; it is thus impossible to distinguish archaism from Indo-Greek innovation, or it is uninformative because other branches have innovated further.

Instead, I have examined another kind of lexicon, namely borrowings into common prestages, albeit from a methodological perspective (Article 2).

3.4. Ringe, Warnow & Taylor 2002

From "word-list linguistics", attempting to combat the data-bias by setting up identifiable semantic parameters, there is also little luck. In the datasets of Ringe, Warnow & Taylor (Ringe & Taylor 2003; 2004; 2007a; 2007b; 2007c; 2012) there are very five few unique cognations shared between Greek and Vedic, Young Avestan and/or Old Persian, and no exclusive morphological or phonological isoglosses. The following are from the screened dataset (2004):

1. **denk̂*- 'bite' (Gr. *δάκνει*; Ved. *dáśati*; Av. *dąs*-)
2. *h_1*geh*$_1$*gór*- 'be awake' (All pf.: Gr. *ἐγρήγορε*; Ved. *jāgā́ra*; Av. *jaγāra*)
3. **ĝʰéslo* 'thousand' (like Porzig) (Gr. *χῑ́λιοι*; Ved. *sahásram*; Av. *hazaŋrəm*)
4. **pi-sed*- 'squeeze' (Gr. *πιέζει*; Ved. *pīḍáyati*)
5. **bʰénĝʰu*- 'thick' (Gr. *βαχύς*; Ved. *bahulá*-)

These are not good candidates for innovations. **denk̂*- 'bite' seems to be an archaism, it is also attested in Germanic, albeit with a semantic shift (*LIV²*: 118). The use of the perfect of *h_1*ger*- for 'to be awake' is striking on the surface, but the aspectual system only exists as such in these languages. The root could very well be an archaism – but it is also attested in Albanian *ngrihet* < 'raises (oneself)' (*LIV*^{3}: 36).

The word for thousand is a simple distributional lexical isogloss. Indo-Iranian *_sa-ȷ́ʰasra-_ goes back to *_sm̥-ĝʰeslo-_, found in another derivation in Greek: "PIE" *_ĝʰésli(H)o-_ > PGr. *_kʰéslii̯o-_ > Ion. gen. pl. *χειλίων*, Thess. acc. pl. fem. χέλλιας (Porzig 1954: 159; Chantraine 2009: 1290; Ringe 2024: 126).[7] The isogloss is uninformative at best and an archaism at worst: Since a phonetic form is unattested in Anatolian, Tocharian has its own form (TB *yaltse*, TA *wälts* < PT *_ẃəltse_ < PIE *_u̯él-i̯o-_, cf. OCS *velьjь* or *velikъ* 'great' also 'thousand' (Adams 2013: 532)), Armenian, Albanian and Celtic have loan words, only the "North European" contestant *_túHsn̥t-_ remains as another candidate for PIE term. If PIE indeed had one. However, if Lat. *mīlle* is to be derived from *_smih$_2$-ĝʰeslih$_2$-_ > PIt. *_smīχeslī_, this too is a derivative of the inherited root (de Vaan 2008: 380; but cf. Ringe & Taylor 2012: 113; Mayrhofer 1996: 489–91, 539–40).

If Gr. *πιέζω* is indeed somehow from *_πίζω_ < *_pi-sd-e-_ and connected to Ved. *pīḍayati* < *_piẓda-i̯a-_ < *_pi-sd-e-i̯e-_ (transposition), this is a potential candidate for a unique innovation: It is an unpaired derivation filling a semantic slot in the basic vocabulary. The etymology is, however, uncertain. There is no easy way to get from the expected *_πίζω_ to the attested forms; and what is the first element? The formation is parallel to the noun *_ni-sd-o-_ 'nest' (Ved. *nīḍá-*, *nīḷá-*) consisting of a preverb and the rood *_sed-_ 'sit', but there is no preverb *_pi-_, and it is impossible to lose the *_h$_1$-_ of *_h$_1$(e)pi-_ in Greek. The forms may ultimately not be connected at all (Mayrhofer 1996: 136–7; Robert S. P. Beekes 2010: 1189).[8]

Finally, Gr. *βαχύς* 'thick' and Ved. *bahulá-* 'id.' continue *_bʰénĝʰ-us_/*_bʰn̥ĝʰ-éu̯-_. It is probably also continued in Lat. *bìežs/bìezs* in the same meaning (*NIL*: 114; *IE-CoR*: s.v. Cognate Set 1585), making it not unique. The Latvian form is included by Ringe et al., but not etymologically connected to Indic and Greek. However, *_bʰenĝʰ-_ is

---

[7] The vowel is irregular in Greek. For one suggestion, see Rix (Rix 1985: 341–2)

[8] The seemingly exact nominal parallel *_pi-zd-áh$_2$_ 'Vulva, weibliche Scham' seen in Ashkun *pəṛ'ī* < *_pīḍikā-_), Lith. *pyzdà*, OPr. *peisda*, Russ. *pizdá,* Pol. *pizda* and Alb. *pidh/pith* (*NIL*: 591, 595–6) is probably a mirage of wrong etymologies and polish borrowings in Baltic (Smoczyński 2018: 115, n. 57; Vasmer 1955: 355).

attested as an adjective slightly different meanings: Hitt. *panku-* 'complete, whole, all', Lat. *breuis* 'short', Av. *mərəzu-* 'tight' (*NIL*: 113–5; Ringe 2024: 126).

3.5. Kassian et al. 2021a

In Kassian et al. (2021b)[9], there are similarly three unique lexical roots shared between Indic (Atharva-Veda), Iranian (Proto-Iranian) and Greek (Attic):

1. **denḱ-* 'bite' (like Ringe et al.)
2. **u̯ekʷ-* 'say' (Gr. *εἶπα*, Ved. *vac-*, PIr. **wayn-* (their notation)
3. **h₃ógʷʰi-* 'snake' (Gr. *ὄφις*, Ved. *áhi-*, PIr. **aǯ-i-* 'dragon' (violating the semantic criterion))

Proving just how very small changes make a difference for studies like these, we can compare these two very short lists. The three meanings from Kassian et al. are also found in Ringe et al., but they code *λέγει* as the most basic term for 'say', and 'snake' did not make the cut to the screened dataset (because or polymorphism and parallel innovations). While **u̯ekʷ-* only functions a verbal root in Indo-Iranian and Greek, the root noun **u̯ōkʷs/u̯okʷ-* 'voice' has a wider attestation (Ved. *vā́k, vācaḥ*; Av. *vāxš*, *vaco*, Gr. Hom. hapax acc. sg. *ὄπα*), but also Lat. *uōx, uōcis* and TB *wek* < **u̯ŏks* (Schindler 1972; de Vaan 2008: 691–2; Adams 2013: 660). The word for 'snake' is also by no means isolated to Greek and Indo-Iranian, but the root vowel and especially semantics is not identical across the board (Euler 1979: 131).

3.6. Heggarty et al. 2023

In Heggarty et al. (*IE-CoR*), counting such unique cognations specific meanings make less sense since the database includes many more languages. In the examples above, it is somewhat reasonable to use Vedic as a proxy for (Proto-)Indic and an attestation in Avestan and/or Old Persian as proxy for (Proto-)Iranian, but surface match between etymon in meaning in a Modern Greek dialect and New Iranian or Indic language tells nothing about the latest common ancestor. Nevertheless:[10]

[9] Only Proto-Iranian cognations are argued there, Greek (Plato's Attic) and Atharva-Vedic elsewhere (Kassian 2011; Trofimov 2016)

[10] Indic, Iranian & Greek: iecor.clld.org/cognatesets?cladefilter=Hellenic%2CIranic%2CIndic, Indic & Greek: iecor.clld.org/cognatesets?cladefilter=Hellenic%2CIndic, Iranian & Greek: iecor.clld.org/cognatesets?cladefilter=Hellenic%2CIranic

1. *$\underset{\frown}{u}e\underset{\frown}{i}\hat{k}$*- ‘house’ (Gr. *οἶκος, οἰκίᾱ*, Myc. *wo-i-ko-*; Pashai *we*; Khot. *bäsa,* OP *viθ*)
2. **meh₁*- ‘count’ (Mod. Gr. *μετρώ*; Marathi *mojaṇe*; Oss. *nymajyn*)
3. **pet*- ‘fall’ (Gr. *πίπτω*; Pali *patati*; Sogd. *ʾmptt*, Khot. *pīttä*)
4. **denk̂*- ‘bite’ (like Ringe and Kassian)
5. **deh₃*- ‘hit’ (Cappadocian *δίνω*, Tsakonian *ντίου*, Gawarbati (Indic) *łik*)
6. **gʷʰreh₁*- ‘smell’ (Gr. *ὀσφραίνομαι*, Ved. *jíghrati*)
7. **tk̂en*- ‘kill’ (Gr. *ἀποκτείνω*, Wakhi *šit*; but cf. Ved. *kṣaṇóti* ‘to injure’)

1, **2**, **3**, **5** and **7** are obviously parallel semantic developments. **4** is treated above. **6**, **gʷreh₁*- ‘smell (something)’ is apparently a unique lexical isogloss (*LIV*[2]: 221), unless MHG *bræhen* < OHG **brāen* < PGmc. **brǣ-ja-* < **gʷʰreh₁-i̯e-* should be connected (*LIV*[3]: 31–2). In that case, it is no longer a unique lexical isogloss.

4. The phonology of Indo-Iranian and Greek

4.1. The latest shared phonological system of Indo-Iranian and Greek

The task of linguistic subgrouping does not end with the graphic image of a family tree. Each node in the tree represents the accumulation of shared developments, and researchers of this field should really take the consequence of this and be clear about their reconstructions of the proposed proto-language of each node. For many of the higher-order subgroups, our limited knowledge will inevitably lead to reconstructions of *very* similar tentative language states. I for one think this endeavour has great value - even if the outcome is only pointing to the absurdities of the gaps in our knowledge or the fragility of the house of cards. For too long, too much of the Indo-Greek relations have been done in the dark and only hinted at in vague phrases (See Article 1).

If we are to take the task of assessing what the latest ancestor that could be Proto-Indo-Greek could be, we must look at the phonological systems and reconstruct a common one. Elsewhere, I will discuss suggested innovative isoglosses in more detail, but in this section, I will address the latest shared phonological system that the two branches can go back to. It is not meant to be exclusive to those two per se, but it is interesting to set the scene and see in context what system we arrive at when applying a bottom-up reconstruction approach to just the two branches.

A very striking issue in Indo-Greek comparison within the Indo-European language family is the relative conservatism of both branches – or perhaps, alternatively, the great role they have played in the reconstruction of the protolanguage, especially before the discovery and deciphering of the Anatolian languages. If indeed there is no closer genetic relationship between the two branches, it should come as no surprise if the ancestral language of what later turned into - at least - Proto-Indo-Iranian and Proto-Greek turned out to be identical to Proto-Indo-European itself. If the two branches did in fact innovate together, it will soon be very clear that they at least remained relatively conservative in their phonology but instead evolved in its morphology and lexicon.

If we compare the reconstructed Proto-Greek that we arrive at from the relatively homogenous Greek varieties with the Proto-Indo-Iranian based on the much more heterogeneous Old Indic, Old and Middle Iranian and the additional relatively conservative but contemporarily attested Nuristani languages, we will soon realise that the common ancestral tongue that these must go back to does in fact resemble traditionally reconstructed Proto-Indo-European strikingly.

4.2. Vowels

4.2.1. Short vowels

At the surface level, the only system that can account for the distribution of the five short vowels *a e i o u* of Greek and their Indo-Iranian counterparts *a i u* is the traditionally reconstructed Proto-Indo-European five-vowel system usually assumed in a laryngeal colouring context.

It has been known for centuries that the Indo-Iranian palatalization of velars and labiovelars must precede the merger of the vowels **e a o* in **a*, and that the vowel system is therefore younger than any potentially shared state.

The phonemic status of **i u* and their relation to **i̯ u̯* was probably just as in the parent language, whatever that was.

However, there are instances of inherited **a*, whatever its ultimate origin and status in Proto-Indo-European, that do seem to point to a phonemic status of **a* in the latest shared state of Indo-Iranian and Greek. Rarely, the introduction of phonemic **a* is seen as a post-Anatolian innovation (Trager & Smith 1950: 66). At least judging from some direct comparanda of roots with **a* that cannot be of laryngeal origin, i.e.

- **bʰag-* 'to receive one's share' in Gr. aor. *ἔφαγον* 'ate' and Ved. prs. act. *bhájati* 'allots', prs. med. *bhájate* 'receives one's share' (*LIV*[2]: 65; *LIV*[3]: 11)
- **Hi̯aĝ-* 'to honour' in Gr. *ἅζομαι* 'honour', *ἅγιος* 'holy', Ved. *yájate* 'sacrifices, honours', OAv. *yazaitē* 'honours' (*LIV*[2]: 224–5)

For these roots, original **e* is possible if the Greek vowel is a "schwa secundum" from the zero grades **bʰg-* and **Hi̯g-* (LIV[2]: 65, 1 ; 224–5, n1.).

On the basis of Old Persian onomastic material (the participle *-bigna* 'bestowed' in the personal name *Bagabigna*), Cheung (2007: 2) has suggested that a reconstruction **bʰeh₂g-* fits Iranian better. However, this is far from certain – and it requires accepting the glottalic dissimilation rule of loss of laryngeals before mediae: **bʰeh₂g-*, **bʰh̥₂g-* (under the assumption that this is **beʕg-*, **bʕg-*) > **bʰag-*, *bʰəg-* (Lubotsky 1980). TB *pāke*, TA *pāk* 'part, portion, share, piece' do not "speak for old *a*" as argued in LIV; if the word is not a borrowing from Sanskrit or Iranian, Tocharian does not attest forms that would allow us to distinguish between *b^hag*- and *b^hh̥$_2$g*-, as both would yield PT **pak*- > TB /a/, TA *ā* (Adams 2013: 389).

For the second root, Cheung (2007: 219–20) somewhat automatically reconstructs **Hi̯eh$_2$ĝ*- without commenting on the consistent short *a* of Indo-Iranian. Once again, he must rely on the loss of laryngeal before **ĝ*. This root may not be unique to Indo-Iranian and Greek after all. Rieken (Rieken 2007: 273) has suggested that it survives as HLuw. *izii̯a*- 'to do, make' < **Hi̯aĝ-i̯e*- and that this was the original meaning of the root. It seems to have been well-received (Yakubovich 2020: 469; *eDiAna* s.v. lemma 3731; Sasseville 2021: 324–5; *LIV*^{3}: 61), though it does meet criticism on the formal and semantic side – **ĝi̯* > *z* is not uncontroversial, the placement of the accent and the vowel grade of the root are all debated (Kloekhorst 2019).

4.2.2. Laryngeal colouring and Brugmann's law

An additional problem is the ambiguity of the evidence of Brugmann's Law in Indo-Iranian. It is well-known that **o* is lengthened to **ā* in open syllables, and that laryngeals were still in place to block this development when it happened.

It is, however, unclear if **o* from *h_3o* actually undergoes Brugmann's Law as well. It is traditionally assumed that it should be expected because laryngeal colouring is assumed already for Proto-Indo-European. This is based on the economic assumption that since all branches undergo the same colourings of **e* adjacent to *h_2* and *h_3*, it is most likely a common trait inherited from the parent language rather than individual parallel innovations occurring multiple times independently over the history of the language family. As Rasmussen (1993) expressed it, this solution "which is *theoretically* possible, is hardly credible since it would make the rounding of **e* to **o* when contiguous with *h_3* a separate event which just happened to occur in all Indo-European languages we have come to know about".[11]

The problem is that there is no positive evidence in favour of **ā* from *h_3e* in an open syllable. On the contrary, there are examples of positive evidence against it,

---

[11] I am grateful to Thomas Olander for sharing this quote and his notes for the presentation *Laryngeal colouring in Indo-European* planned for the Covid-cancelled conference *The Sound of Indo-European IV*, Universität zu Köln, 1–3 April 2020.

e.g. Ved- *ávi-* 'sheep' < *h_3*éu̯i-* (or *h_2*óu̯i-*, unclear from all (other) branches, e.g. Gr. *ὄϊς*< **óu̯is*, Hitt. *ḫāu̯i-* < **$h_{2/3}$óu̯i-*) and never ***āvi-* (Lubotsky 1990).

Rasmussen (1989a: 167) adduced the example Ved. *styā́yati* 'coagulates, hardens' which to him could hardly be anything but a Brugmann-coloured **o* from **h_3o* < **h_3e* in the iterative **stih$_3$-ói̯e-ti* < **stih$_3$-éi̯e-*. This was wholeheartedly accepted by Hajnal (1995: 309), though explaining the formation as an intransitive, not an iterative. This form has several problems. On the semantic side of things, I do not see what is inherently iterative about the act of "becoming coagulated". On the philological side of things, any present of the verb is attested late – Whitney (1885: 194, s.v. styā, stī "stiffen") quotes the Brāhmaṇas and Classical Sanskrit. The form *styā́yati* 'stiffens' itself is troubled by the fact that it is not attested – at least with the accent and in the active. Insler (1987: 61–2) inferred this "underlying" active of the verb, otherwise only attested in the middle. According to this analysis, the middle would have been modelled on the preceding *ā́pyāyatāṃ* with which it always occurs. The verb occurs as the middle imperative *níṣṭyāyatāṃ* once in Vedic (Atharvaveda-Pāippalāda 2.39.1, repeated four times in the Vājasaneyi-Saṃhitā) in what seems to be a fixed collocation:

> *tát* (or *sá*) *ta ā́pyāyatāṃ nístyāyatām soma rājan*
>
> "Let it swell up and become solid for thee" (Insler 1987: 61) or "let it (scil. soma-juice) swell up and become coagulated [or become quiet?] for thee, o king Soma" (Lubotsky 1997a: 58)

Because of the preceding verb and the fixed phrase, Lubotsky considers it a nonce form. Concerning the etymological analysis, Lubotsky (1997a: 57) "strongly disagree[s] with the development *Cih$_3$C* > Gr. *Ci̯ōC*" on which the etymological connection between *styā́yate* and *σῶμα* 'corpse' (Homer, otherwise 'body') < *(*s)tih$_3$-mn̥* hinges. I do not find the connection "semantically far-fetched" – a dead body is, after all, stiff; though the semantic shift from 'corpse' to 'living body' in post-Homeric Greek might be more difficult to imagine than a semantic specification of 'body' > 'corpse' in the Epics. I am much more sympathetic towards "laryngeal breaking" in Greek than Lubotsky, especially in light of the emendations to the early theories proposed over the last couple of decades (esp. Olsen 2009, cf. below). Lubotsky (1988: 104), rejecting the laryngeal breaking, instead preferred connecting *styā* with Gr. *στέαρ* 'hard fat, dough' from **steh$_2$-i̯-*, and *i*-extension of **steh$_2$-* 'stand'. Technically, Rasmussen also made this connection, though positing an ablauting root vowel *a/ā/Ø* (Volkart 1994: 26 n. 97). As to the proposed nonse-form explanation of *-styāyate*, it is not exactly easy to pinpoint where and how this *ā* arose; even in the suspected trigger-word *pyāya-*, the stem formation is unclear and schwebeablaut might be necessary (*LIV*[2]: 465; Kümmel 2000: 317; Mayrhofer 1996: 172). Kulikov notes, although accepting the influence of *-pyāya-* 'swell, prosper' on *-styāya-* 'become coagulated', that there is a "group" of semantically and phonologcally similar verbs synchronically forming a group denoting "spontaneous, non-controllable processes" – the third member of the group being *sphāyate* (late *sphīyate*) 'swell' (Kulikov 2012: 511, 515).

Finally, I find it suspicious that this "ancient" development of the Brugmann-lengthened vowel of the iterative would have gone completely under the radar and only be preserved in this quite obscure instance.

Returning to the more serious bulk of evidence against the operation of Brugmann's law on **o* < **h_3e*, there is no doubt that these instances are counterexamples to an early laryngeal colouring. However, none of them are particularly strong and very easily explainable as paradigmatic levelling. Further, there are additional counterexamples to Brugmann's Law, like *pati-* < **poti-* 'master' and the preverb *prati* < **proti* which require other explanations (Grestenberger 2024). What is more, it is worth noting that while Lubotsky's solution has provided an explanation for the lack of Brugmann's law in word where short *a* can be analysed as **h_3e* rather than **o*, the very same explanation requires other paradigmatic levellings to explain why there is no palatalisation of velars in front of non-coloured **eh_3*. To explain the short vowel in the paradigm of Ved. *gáuḥ,* acc. *gā́m*, gen.-abl. *goḥ*, dat. *gavé*, loc. *gaví*, Lubotsky (1990: 134) reconstructs **g^wé$h_3$us, *g^wé$h_3$um, $g^w$$h_3$éu̯s, $g^w$$h_3$éu̯ei̯, $g^w$$h_3$éu̯i* rather than, as traditionally assumed, **g^wṓu̯s*, **g^wṓm* (with Stang's Law), **g^wéu̯s, *g^wéu̯ei̯* etc (Nielsen Whitehead 2018). According to Lubotsky's own relative chronology of the sound changes on the same page, the attested forms of the nominative and accusative must be analogical:

1. Brugmann's Law (**o* > *ā*)
2. palatalisation of velars in front of **i, e, ē*
3. the vowel merger (**e a o* > **a*)
4. the coalescence of the three laryngeals (**h_1 h_2 h_3* > **H*).

Since there was no original **o* in Lubotsky's paradigm, it would not be affected by Brugmann's law, but of the three subsequent developments:

1. (pre-PIIr.) nom. *$g(^w)\bar{\acute{e}}h_3us$*, loc. *$g(^w)h_3\acute{e}\underset{\frown}{u}i$*
2. nom. *$\check{j}\acute{e}h_3us$*, loc. *$gh_3\acute{e}\underset{\frown}{u}i$*
3. nom. *$\check{j}\acute{a}h_3us$*, loc. *$gh_3\acute{a}\underset{\frown}{u}i$*
4. nom. **jáHus*, loc. **gaHí*

This should yield a Vedic paradigm nom. ***jauḥ* or ***joḥ,* loc. *gaví*. Thus, while the analysis has solved the problem of the analogical short *a* from **o* in an open syllable, it requires extensive analogy to explain why *eh_3 does not cause palatalisation if the colouring to **o* is so late. A similar critique, instead of the lack of palatalisation in the genitive *$g^w\acute{e}\underset{\frown}{u}s$ applies to the traditional reconstruction (De Decker 2011a: 52). Lubotsky's (1997a: 58) own defence of the chronology of changes 1 and 2 raised by Volkart (Volkart 1994: 27, n. 98) fails to convince me: "we cannot know when the phonetic palatalization started, but it became phonemic at the moment when the conditioning factor, i.e. the difference between **e* and **o*/**a*, disappeared". It is absolutely true that the palatalisation was only phonemicised with the vowel merger, but this does not explain away the fact that the pair **Ke* : **Ko* remained in systematic opposition, only as **Ca : Ka*, and did not merge and were not redistributed at a toss-up.

4.2.3. Long vowels

Most long vowels (especially **ī, ū*) were probably still sequences of short vowel plus laryngeal (see below). Original **ē* and **ō* were simply preserved, again proven by the Indo-Iranian palatalisation rules. The unsolvable question of when the laryngeals were lost with compensatory lengthening (**VH* > *V̄*) will be treated below.

Proto-Greek and Proto-Indo-Iranian definitely had a vowel **ā*. Indo-Iranian, obviously, merged **ē ō* with **eH aH oH* and had productive morphological long grades, also of roots with original *a*-vocalism, e.g. in the root aorist of *yaj-*, among other forms attested in aor.ind.2.sg.act *ayās* (R̥V 3.29.16.c) < *$h_1e\text{-}\underset{\frown}{i}\bar{a}\hat{g}\text{-}s$*.

Whether the two branches had a vowel **ā* is debateable. It is difficult enough to find secure examples of PIE long vowels of non-laryngeal origin as well as the root vowel **a* in itself (Sihler 1995: 49–50). Examples of **ā*, if they existed, could be expected to be found in ablauting root nouns (Rasmussen 1989a: 260), but the usual examples invoked for this type are coincidentally not preserved in both branches, making it difficult to argue for the existence of an even marginal phoneme (e.g. **să̄l-* 'salt', **nă̄s-* 'nose').

It is unclear whether PIE had *ī and *ū in the first place. In most languages, the outcome would be identical to **iH, *uH* anyway (Ringe 2024: 102). In Greek, the accent-conditioning of laryngeal breaking (Olsen 2009) would make them indistinguishable (and if not, paradigmatic levelling could do the rest). In Balto-Slavic, where the presence of a laryngeal can usually be detected by the tone, there is no consensus on the treatment in monosyllables (Olander 2009: 106; cf Yamazaki 2014; 2016).

4.2.4. Vocalized laryngeals

The laryngeals probably still behaved as they did in Proto-Indo-European, that is they were primarily consonants but when necessary, they could fill the syllable nucleus (Byrd 2015: 27–34). How this vocalized allophone was realised is uncertain and up for debated. We can be certain that the vocalic versions of the laryngeals cannot yet have merged with any other vowel phonemes. In Greek, they vocalise with the famous triple reflex *h̥$_1$ h̥$_2$ h̥$_3$ > *e a o*, whereas they all merge in the high vowel **i* in Indo-Iranian (Kümmel 2022: 251).

Both developments are much easier derived from an archaic stage than with any intermediate development. Theoretically, we could postulate that the vocalized laryngeals had become their own vowel phonemes that took part in the ablaut system, but it would be quite uneconomical to invent three extra phonemes. The consonantic laryngeals cannot have disappeared entirely in a common prestage. It is possible that the laryngeals were pronounced with a prop-vowel **Hə*, which could in turn have been sub-phonemically coloured like the full vowels [h$_1$e h$_2$a h$_3$o] (Byrd 2015: 27–34). This certainly fits the Greek developments better than the Indo-Iranian where we can indeed see traces of **Hə* if one accepts that the laryngeal can aspirate a preceding stop and vocalise in cases like **dʰuĝh̥$_2$tér-* > **dʰuĝʰH̥tér* > **dujʰitár-* 'daughter'. However, the reconstruction of this exact word is one of the hottest topics in Indo-European linguistics and there are certainly other options as well (Lubotsky 2018: 1883; Kobayashi 2004: 131; 2017: 332; Kümmel 2018a: 169). Within Indo-Iranian, it also not entirely clear if the laryngeals failed to vocalise in some positions, or were vocalised but lost subsequently lost. Olsen (2018a) convincingly argues that they did develop but were lost outside the first syllable in Iranian. This accounts for the variation in the paradigm of the word for father where – if one assumes that the Avestan nominative is either analogical from compounded forms, or is simply the expected reflex because of the close connection to the preceding word:

PIE>PIIR	Ved.	OAv.	OP
Nom.sg. **ph̥$_2$tḗr* > **pitā́(r)*	*pitā́*	*(-)ptā* (**pitā* ?)	*pitā*
Gen.sg. **ph̥$_2$trós* > **pitrás*	*pitraḥ*	–	*piça* (gen)
Dat.sg. **ph̥tréi̯* > **pitrái̯*	*pitré*	*piθrē*	–

The hapax dat.sg. *fᵊδrōi* would be analogical (influence from the strong cases, i.e. **p(i)trōi* > **fθrōi*) by this account. In the word for 'daughter', Olsen's (2018a: 264–5) rule explains the Ved. and OAv. paradigms, does not mention the stem YAv. *duxt-* which is continued in many (and all Eastern) Middle and Modern Iranian languages (Wendtland 2009: 175; Rastorgueva & Ėdel'man 2003: 476–9), e.g. Sogd. *dwγtʿ* /δuγda/, Khot. *dutar-* /δudar/ < **duxtar-* (Skjærvø 2022: 123).[12] This particular word shows that Bartholomae's law (progressive voicing assimilation) was active longer than the vowel deletion in Iranian, but not in Indic, and apparently not in Nuristani, where Prasun *lüšt* < **dužitā* sides with Indic in vocalisation and secondary palatalisation (Mayrhofer 1984: 253; Degener 2002: 108), against Iranian (Bartholomae's law), i.e. Indic does not have ***dugdhār-*, and Avestan does not have ***dučitā* or ***duxtā* (or for that matter ***dušta/*duža,* as were it a **ĵ,* not a **ǰ*) – and Prasun does not have ***lüd* < **dugdā-* (Budruss 1977: 34, n. 16). At any rate, the "undoing" of Bartholomae in forms like OAv. *aogᵊdā*, YAv. *aoxta* < **(H)au̯gʰ-ta-* (Hoffmann & Forssman 1996: 95)

Kümmel (2018a: 169) has even suggested that the vocalisation of the laryngeals is a trait that points to an early split-off of the Indo-Iranian branch. These languages are the only ones to merge the vocalic laryngeals with a high vowel. Other than in Greek, this vowel is always low, and most often **a* (Italic, Celtic, Germanic, Balto-Slavic,

---

[12] This stem is difficult. It may be analogical from other stems in **-tar* (Rastorgueva & Ėdel'man 2003: 476), or it may be expected; Kümmel (2018: 169) suggests that it is either the old strong stem if **dʰug-Htar-* > *dukHtar-* > *duxtar-* or the old weak stem if **dʰugtr-* had already been simplified to **dʰuktr-* in PIE. Hoffmann & Forsmann (1996: 82) explain it in reverse; to them OAv. *dugᵊda* is the regular reflex of PIIr. **dʰugitar-*, and Ved. *duhitár-* came into being "[d]urch Kontamination der Nebenformen **dʰugʰter-* und **dʰugiter-*.

Tocharian (as PT *a*), but probably not in Anatolian[13]). If, however Indo-Iranian could be proven to be an early language to branch off (as also suggested by Hamp (2013)), the qualms of Brugmann's Law and laryngeal colouring would be smaller.

4.3. Resonants

4.3.1. Liquids

Another uncontroversial and archaic feature is the preservation of the liquids. However, it is rather unremarkable as these are preserved distinct in all branches but Indo-Iranian. Although the Mycenaean script does not distinguish /r l/ but write them both the *r*-series, we have no reason to believe that they had phonemically merged (Bernabé & Luján 2020: 78; Del Freo 2016: 133; Thompson & Meißner 2024: 100) – it is, on the other hand, also impossible to prove or disprove. Important is that the series corresponds to *r*- and *l*- of the other Greek dialects and Indo-European languages.

4.3.1.1. The merger of liquids in Indo-Iranian

The merger of the liquids seemingly happened in all Indo-Iranian languages. Still, it is often said that the merger cannot have been of Proto-Indo-Iranian date. I believe this to be wrong. To clarify, the standard handbook explanation of the matter is not entirely clear. It is always states that it is an Indo-Iranian trait to merge **r* and **l*, but whether there ever was a phonemic merger in Proto-Indo-Iranian is debated.

Ironically, opinions differ on which of the branches is the more archaic. Cantera (2017: 485), in his chapter on Iranian, states that Indic neutralises the difference, but that Iranian preserves traces of distinct phonemes; whereas the indologists Kobayashi and Lubotsky, in their chapters on Indic and Indo-Iranian, respectively, state that (Old-)Iranian shows a full merger, but that the distinction recoverable in Indic must be inherited (Lubotsky 2018: 1878; Kobayashi 2017: 330).

[13] The alleged Proto-Anatolian **a* from CH̥C is unfortunately tied to the reconstruction of the word for 'daughter' **dʰuĝ⁽ʰ⁾₍h̥₎₂-ter-* (Melchert 1994: 70; Kimball 2017: 253; Kloekhorst 2008: 903; 2011; Schürr 2023)

In Iranian, the alleged evidence for the preservation of two distinct phonemes come from Middle and Modern Iranian languages whereas it has been argued that already the Rig-Veda show evidence of dialect mixture. In some accounts, **r* and **l* merged into one phoneme which has realised as *r* in the West and as *l* in the east (Sihler 1995: 174). Later influence from the eastern Prakrits would then be to blame for the increasing number of *l*'s in place of inherited **l* > **r* in the attested history of Indic (Classical Sanskrit has more *l*'s than Atharva-Veda which again has more than the Rig-Veda).

To others, the cases of **l* in place of inherited **l* are evidence of a "Central" dialect preserving the distinction (Beekes 2011: 138):

Proto-Indo-European							
	Proto-Indo-Iranian						
		Proto-Iranian	Proto-Nuristani	Indic			
				Proto-Indic	Western	*Central	Eastern
*r	*r	*r	*r	*r	r	*r	l
*l	*l	*r	*r	*l	r	*l	l

By this logic, in order for the unattested central dialect of Indic to have preserved the *l* in place, there can never have been a full phonemic merger in Proto-Indic and Proto-Indo-Iranian.

Arguably, the best argument in favour of the preservation of two phonemes would be positive evidence for exactly that. There are a handful of examples in which *l* does seem to continue **l*:

- **l* > *l*: IE **pelit-* 'grey' > Ved. *palitá-* (cf. Gr. *πελιτνός* (Porzig 1954: 158), although Mitanni Indic shows *paritannu-/barittanu-*)
- **l* > *l*: IE **lei̯ĝ*^{h}*-* 'lick', Ved. *leh-/lih-*, ModPers. *lištan*, Kurd. *listin*

4.3.1.2. Indic

There is no direct evidence from any dialect or language faithfully preserving **r* and **l* as distinct phonemes, but there are varieties in which individual etymologies seem to preserve especially **l* in its proper place.

This has been substantiated in several ways. Burrow (1972) advocated for this dialect continuum in which **r* and **l* did not merge. Mayrhofer (2004), on the other hand, argued for a complete phonemic merger, with subsequent differing realisations across the descending dialects. Hock (1991) added some sociolinguistic tendencies. Not only are the majority of _l_'s in Vedic (and younger Indic) of Dravidian or unknown foreign origin, they are also associated with unexpected sound changes, low-register speech in the Vedas and names of (female) demons, all pointing to Substrate.

There is no doubt that the two phonemes merge in most cases in all attested languages. Further, the vocalic alternants _*r̥_ and _*l̥_ always merge in _*r̥_ (see below) – the latter point can of course only be circumstantial evidence for the preservation of the distinction of consonant phonemes.

- _*r_ > _r_: IE _*b^h**r**á$h_2$ter-_ 'brother' > Ved. _bh**r**ā́tar-_, Av. _b**r**ātar_, OP _b**r**ātar_
- _*l_ > _r_: IE _*k^wé**l**$h_1$eti_ 'turns' > Ved. _cá**r**ati_, Av. _ca**r**aiti_

Lubotsky (2018: 1878) sums the matter up very well:

> Nevertheless, Skt. /l/, which is relatively rare in the RV and becomes more prominent in later texts (e.g., RV _áram_, AV _**álam**_ adv. 'fittingly, accordingly, enough' < PIE _*$h_2$erom_; RV _reh-_, AVP+ _leh-_ 'lick' < PIE _*lei̯ĝʰ-_; RV+ _**pal**itá-_ 'grey' < PIE _*pelit-_; RV+ _prav-_ /_**pl**av-_ 'swim' < PIE _*pleu̯-_; RV+ _rep-_/_**lep**-_ 'smear' < PIE _*lei̯p-_, etc.), for the most part corresponds to PIE _*l_.)

Niels Schoubben (2019) has substantiated this "for the most part" in his MA-thesis. According to his count, there are 142 words containing /l/ in the Rig-Veda. 82 of them are not inherited. 35 have uncertain etymologies. 25 have secure etymologies, and of these 16 correspond to _*l_ and 9 to _*r_. Though it is certainly small numbers, I see why it is tempting to accept these at face value as evidence for some sort of retention of _*l_ in the original place.

However, other than the geographic and the sociolectal explanation, there are also some phonetic tendencies worth considering. Hock (1991) further adds that there seem to be some sort of dissimilation going on. First, we observe _álarti_ and _álarṣi_ rather than expected _*árar-_ < _*HárHar-_ < _*$h_3$órh$_3$or-_ (Mayrhofer 1986: 105–6, 247). Second, it is rather clear that a labial environment has an influence on the fate of the merger. To Schoubben, _*r_ and _*l_ began merging but were kept apart in labial surroundings; to Hock _*r_ and _*l_ merged completely and was subsequently realised as _l_ in labial surroundings.

In fact, in 13 of the 25 cases of secure etymologies in Schoubben's count, there is a labial consonant:

- *u**p**ala**p**rakṣín-, **p**alitá-, **p**ulú-, **p**lu, **p**lāśí-, **b**ála-, **m**ála-, **m**lā, la**bh**, li**p**, lu**bh**, ló**p**āmudrā-, ló**m**an-* (Schoubben 2019: 44)

If we were to take such examples at face value, they do indeed indicate that *l* can be preserved in its proper place in some varieties. However, making these few lexemes carry the burden of reconstructing contrasting phonemes might be a bit of a stretch. Especially because there are clear cases of inherited **r* also surfacing as *l*:

- **r* > *l*: Ved. *cal-* next to *cárati* < **k^welh$_1$-*
- **r* > *l*: Ved. *lohitá-* 'red' next to *rudhirá-* < **h$_1$r(e)ud^h-*
- **r* > *l*: Ved. *lóman-* 'body hair, fur, pelt' next to *róman-* < **Hreu̯men-*

These can only occur if there was indeed a phonemic merger of **r* and **l*. It might very well be that in some variety, be it a "central dialect" or any other entity, the outcome of this phoneme was realised as something so [l]-like that it could be borrowed as *l* in *r*-dialects. This is, in my opinion, reason to follow Mayrhofer's (2004) explanation and abandon any thought of preserved contrasting phonemes in Proto-Indic (*pace* Adams 2023: 227).

Of course, it is an uneconomical explanation to claim that the cases of *l* (re)surfacing in their original place are the result of dialect borrowings after the phonemic merger. However, the fact that even inherited **r* can undergo the same fate severely weakens the hypothesis that there existed dialects in which the two phonemes were kept apart. This becomes a rather unnecessary complication.

4.3.1.3. Iranian

There is reason to read Cantera's conclusion with a bit of scepticism:

> Although in these languages we find some cases of *l* going back to **r* (Parth. *dʾl* 'wood'), thus proving that at a certain time *r* and *l* no longer stood in functional opposition, evidence found in New Persian and Ossetic proves that in Proto-Iranian the opposition between *l* and *r* was still not neutralized.
>
> (Cantera 2017: 486)

First, I would offer the small amendment that “a certain time” is a bit misleading. In my book, phonemes cannot reemerge in their original position once they have been lost or have merged; and thus it would have to be “in a certain area” or in a certain dialect cluster. Second, I find it questionable that individual lexical correspondences that go against the general tendency are seen as archaisms without further consideration.

There are no Old Iranian languages that preserve **l* in its original place, but Avestan reintroduces the phoneme from loanwords. Old Persian is even a “productive *r*-dialect”, in the sense that even foreign words are adapted to /l/ (or, at least <r->): *Bābiruš* ‘Babylon’. Since the middle and modern Iranian languages no not descend directly from the attested varieties of Old Iranian, they could just as well preserve evidence of a more archaic layer than the languages whose written record happen to begin earlier. However, the few “clear” examples disproving a merger in Proto-Iranian are probably not to be taken very seriously. They might even be “a mirage” (Kümmel 2022: 249).

It is a few prominent lexical correspondences. The usual *paradebeispiel* is **lei̯ĝ*^{h}- ‘to lick’ which has *l* in Mod. Pers. *lištan*, Kurd. *listin*, Wakhi *lix̌*-, Parači *līs-/lušt-*, Ormuri *las*. Middle Persian *l’s* is ambiguous and could either spell /lis/ or /ris/. Although many Iranian, and especially Eastern Iranian languages agree in “preserving” *l* in this lexeme, Khotanese shows *r* in *rīśtä* < **raiza-* <**lei̯ĝ*^{h}- (*LIV*^{2}: 404). This lexeme would have been even more attractive since also Vedic has *leh-/lih-*. However, in this example, I fear that another obvious explanation has been avoided: Rather than being overly archaic, it may simply be onomatopoeic. It is, after all, a verb denoting an action of the (human) body using the tongue.

The examples from Ossetic, leaving aside that this language along with many other Iranian languages develop a new phoneme *l* as the outcome of various sound changes, e.g. **ri̯ i̯r* > *l*, is also dubious. Take the word *læsæg* ‘salmon’ which is famous for surviving in Ossetic – the etymon is otherwise only known from Centum languages. Exactly for this reason, it should be considered a loan word (Abaev 1973: 38). These scattered examples do not prove that the general tendency for *r* and *l* to merge in *r* was not a full phonemic merger.

4.3.1.4. A phonemic merger?

Although there are scattered examples of lexemes in which *l* at first glance corresponds to PIE **l* across the Indo-Iranian languages, these cannot be taken as evidence against a full phonemic merger of **l* and **r* in Proto-Indo-Iranian. The Middle and Modern Iranian individual lexemes easily find other explanations: They can be borrowings (Oss. *læsæg*), they can have alternative or uncertain etymologies, or they can be iconic or onomatopoeic (the root **laiǰʰ*).

There are additional – non-critical – arguments as well. We only find IIr. **r* in pronouns and endings (Lubotsky 2018) – but the number of original *l*-forms is too low for this to be salient. It is clear that vocalic **r̥* and **l̥* always merged in **r̥* (Cantera 2017: 486) – while it is attractive to generalise this fact to the consonantic allophones, it does not include preservation elsewhere.

In Indic, economy speaks against the preservation of a distinct phoneme **l* in any dialect: Since both inherited **r* and **l* can undergo the same processes, whatever their details are, to surface as **l*, it is much more likely that the two phonemes merged, before they split – in phonetic, dialectal or sociolectal environments.

Finally, the strongest argument in favour of a complete and universal merger in Proto-Indo-Iranian is the relative chronology. In all Indo-Iranian daughter languages, including the Nuristani branch, PIE **l* causes the *ruki*-retraction of **s* (*s > *ş > š, ṣ): **kʷles-* 'draw, plow' > **karš-* > Av. *karš-*, Ved. *karṣ-*, Nur. **kaṣ* (Hegedűs 2012; Cheung 2007: 241–2) Phonetically, this is easier to understand if **l* was no longer a dental (Kümmel 2022).

4.3.2. Vocalic liquids

The vocalic liquids **r̥* and **l̥* are partially reconstructed on the basis of the Indic evidence since they are preserved as such in this branch. That is, *r̥ is directly preserved and also continues **l̥*. In the R̥g-Veda, only forms of the lexeme *kalp*/*kl̥p* '(com)press', e.g. perf.mid.3pl *cākl̥pré*. In Iranian, there is also good evidence that they were preserved as [ṛ] relatively late. In Avestan, they are continued as such (in some circumstances), only written with anayptyctic schwas on both sides <*ərə*>.

- **r̥* > *r̥*: IE **bʰ**r̥**ĝʰént-* 'high, mighty' > Ved. *b**r**hánt*, Av. *b**ərə**zaṇt-*
- **l̥* > *r̥*: IE. **u̯ĺ̥kʷos* 'wolf' > Ved. *vŕ̥ka-*, Av. *vəhrka-*

In the other languages, **r̥* is continued by *Vr*, but the details differ in the later languages. Korn (2016: 409–13) has argued that is even an overlooked feature distinguishing the West Iranian branches from each other. In neutral – that is, not labial and not palatal – contexts, Middle Persian and Parthian show *ir* (Modern Per. *īr*), Baloči *ur* and Zazaki *ar* which cannot all go back to **ir*. Korn reconstructs **ər* for Proto-West-Iranian – this may well be, but it is a direct phonological continuation of **r̥* - in strict Hoenigswaldian (1966) terms, no systematic change has happened.

In Proto-Greek, the vocalic liquids were still in place – whatever their exact phonetic status was. This has long been recognised on the basis of the internal variation in the alphabetic Greek dialects and in Homer. Like in Iranian, they are assisted by a prop vowel, but the dialects differ on two axes: The placement of the vowel (*VR, RV*) and the choice of vowel (*ο, α*). The situation is complicated by Mycenean and especially by the fact that all dialects show some kind of secondary variation, be it analogical or phonetic. Recent thorough discussions have reevaluated the evidence. Attic-Ionic is certainly an *α*-dialect, but whether the regular vocalisation is *ρα* (as traditionally assumed) or *αρ* has been called into question (van Beek 2022b). There are clear an unambiguous examples of *ορ/ρο* in Aeolic (Lesbian, Thessalian, Boeotian) (Scarborough 2023: 92–105). The situation is complicated in Arcadian and Cypriot where there seems to be evidence for a standard treatment to *α* but *ο* in labial environments. Mycenaean is – as always – complicated by the lack of certain evidence and the unhelpful writing system. It is clear that the reflex was not *ro*, but *or* maybe in variation with *ar* or even preserved **r̥* are possible (van Beek 2022b: 100–2).

For our purposes, it is sufficient that all three branches continue can be traced back to vocalic reflexes. The merger of **r̥* and *l̥* is specific to Indo-Iranian, and the diverging results of **r̥* and **l̥* in Greek and the new **r̥* in Indo-Iranian happened even after the dissolution of the branch-specific protolanguages.

4.3.3. Nasals

The consonantal variant of the nasals are uncontroversially preserved from Proto-Indo-European to Indo-Iranian and Greek – the minor difference being the word-final shift of **-m* to **-n* in Greek.

A possible common innovation, if correct, would be that *m experienced a lower sonority in the immediate shared ancestor of Indo-Iranian and Greek, since it is less prone to vocalise (Zair 2018). This would allegedly be the explanation for the

> Word final *-*m* in the accusative singular of acrostatic and proterokinetic *i*-, *u* and *r*-stems unexpectedly remains consonantal, giving *-*im*, *-*um*, *-*r̥m* rather than expected *-*i̯m̥*, *-*u̯m̥*, *-*rm̥*
>
> (Zair 2018: 280)

Zair argues that the best evidence for this tendency comes from Indo-Iranian and Greek, and since they are often traced back to a common node in the Indo-European family tree (see Article 1 for this particular misunderstanding), it could be an innovation of this clade. However, it would be a very easy analogy to replace the inherited "backwards" accusatives. In fact, Indo-Iranian seems to point to this being the case: The relic forms (acc.pl.) Ved. *paśváḥ*, YAv. *passuō* < **paću̯as* go back to the expected **pek̂u̯m̥s*, not the regularised **pek̂ums* known from Lat., Goth. *-ūs* and Gr. Cret. *-υνς* (and the usual endings PIE *-*ums,* *-*uh₂s* > PIIr. masc. *-*unš,* fem. *-*ūš* > Ved. masc. *-ūn*, fem. *-ūṣ*; OAv. fem. *-ūš*) (Olander 2015: 245).

4.3.4. Vocalic nasals

The first – and only phonological – isogloss in Porzig's (1954: 158) lists of Greek and Indo-Iranian shared features is the continuation of the vocalic nasals **m̥, n̥* as **a* – except before **i̯* and word-finally. While it is true that all other branches preserve some consonant reflex *as well* as a vowel, i.e. *im, in* in Lithuanian, **um, un* in Germanic, **am, an* in Italic and **əm, ən* in Tokharian, it is not possible that this is a shared innovation of Indo-Iranian and Greek.

Already before the decipherment of Mycenaean, it was clear that the result of **m̥, n̥* cannot have merged with **a* in Proto-Greek, as also "Aeolic" dialects show vocalisation into *o* (Rix 1976: 66–7). The situation is once again complicated in Mycenaean where *-o* seems to be the regular reflex in labial surroundings, otherwise *-a* is the standard result for both **m̥* and **n̥* (van Beek 2022b: 27–30; Thompson 1996). Importantly, this "special reflex" only concerns **a* from **m̥* and **n̥*, no other *a*'s, meaning that **m̥* and **n̥* had merged in Proto-Greek, but had not yet merged with **a* and **o*. Whether we refer to this merged phoneme as **m̥ n̥ N̥ ą ã ə̃* or *å* is only a matter of notation.

Indo-Iranian consistently shows **a*. There was likely a middle step, which could have been a nasal vowel such as [ə̃] (Lubotsky 2018: 1876). The same [ə̃] or a similar [ɑ̃] middle step is assumed for Greek by van Beek (2022a: 177).

Lubotsky explains how this [ə̃] was "realized as oral occlusion if **n̥*, **m̥* were followed by a resonant or a laryngeal, i.e. PIE **n̥R*, **m̥R* > PIIr. *anR*, **amR* (where R = a resonant or a laryngeal)". Comparing his examples

- PIIr. 3sg. middle **mani̯atai* > Skt. *mányate* 'thinks, considers', OAv. *mańiietē* 'understands') from PIE **mn̥i̯e-*
- PIIr. Superlative suffix **-tamHa-* (Skt. *-tama-*, Av. *-təma-*) from PIE **-tm̥Ho-*

It is important to stress that this [ə̃] from **n̥* and **m̥* did not merge fully, otherwise PIE **mn̥i̯e-* and **-tm̥Ho-* would have exhibited the same phoneme and not PIIr. **-an-* and **-am-*. Unless, of course, it is analogical from the fullgrade.

Conversely, **m̥* and **n̥* merge completely in Greek, also in front of **i̯*, where also **m* and **n* become **n*:

- PIE **gʷm̥-i̯ō* > **bamjō* > **banjō* > *βαίνω* 'I go' (Schrijver 2006: 51) (but rather **gʷm̥-i̯ō* > **gʷami̯ō* > **gʷaɲō* > PGk. **gʷáɲō* > **báɲō* > **báɲɔ̄* > **bái̯nɔ̄* > *βαίνω*, since the labiovelars must have been preserved in Proto-Greek)

In Greek, also consonantic **m* and **n* merge in front of **i̯*; but this is not the case in Indo-Iranian. Compare **k̂om-ios* > **konios* > *κοινός* 'common, public' (Rix 2009: 76) with Ved. aor. opt. 2.sg. *gamyās*, YAv. *jamiią̊*; Ved. *somyá-* 'soma-(offering)', Av. *haomiia-* 'pertaining to soma'; Ved. *śámyā-* 'yoke pin', YAv. *simā-* < **k̂(e/o)m-i̯áh₂-* (Skjærvø 2007: 902–3, 918; Monier-Williams 1899: s.vv.; Kroonen 2013: 206).

We strictly cannot see how early the merger of **ni̯* and **mi̯* happened in the prehistory of Greek, but it is not shared with Indo-Iranian, just as we cannot know how the reflexes of **m̥* and **n̥* were realised in Proto-Indo-Iranian. However, the dissolution of *[ə̃] into etymologically expected */am/ and */an/ in (pre?-)Proto-Iranian makes it just as if not more likely that the processes were independent.

Compare, Kümmel's (2022: 247–8) circumstantial argument on Grassmann's Law- The introduction of the unexpected, palatalised onset in **ǰadʰí* 'slay!' to avoid homophony with **gadʰí* 'come!' following the neutralisation of the initial aspiration by Grassmann's law, is indeed helped greatly if the forms were in fact homophonous:

- PIIr **gadʰí* < **gadʰí* < **gə̃dʰí* < PIE **gʷm̥-dʰí* ‘come!’
- PIIr. **gadʰí* < *gʰadʰí* < **gʰə̃dʰí* < **gʷʰn̥-dʰí* ‘slay!’

However, if the quality of the syllabic nasal was still as immediately clear to the speakers in front of a stop as it was in front of **H* and **i̯* (**-tm̥Ho- (> *-tə̃Ha-)* > **-tamHa-*, cf. Lubotsky above), the motivation is even less clear.

Summing up, while **m̥* and *n̥* merge in **a* in the “standard” varieties of Indo-Iranian and Greek, internal Greek evidence shows that they cannot have merged with the vowel **a* by Proto-Greek. In Indo-Iranian, there is slender evidence for them having potentially merged in one phoneme, probably already **a* – at least in some contexts, at an earlier, shared state.

4.3.5. “Long” syllabic resonants

In combination with a laryngeal, the outcome of the syllabic resonant is different. Between two consonants, *CR̥HC the outcome is always a long vowel but the details vary across the branched as phonemes. The same general principles, however, apply, and it is difficult to reconstruct another system than that of PIE.

In Indo-Iranian, **Cn̥HC* and **Cm̥HC* merge in **CāC* (Hoffmann & Forssman 1996: 70), e.g. Av. *-zāta-*, Ved. *jāta-* < **ĝn̥h$_1$tó-* ‘born’ and Ved. *śāṃtá-* (for **śātá-*) ‘apeased’ < **-k̂m̥h$_2$tós-* (Hom. *ἄκμᾱτος* ‘tireless’). In Greek, the three laryngeals cause three different vocalic reflexes (**ĝn̥h$_1$tós* in *κασι-γνητός* ‘brother’, **dʰn̥h$_2$tós* > Dor. *θνᾱτός* ‘mortal’, **ĝn̥h$_3$-tós* > *γνωτός* ‘known’ (indistinguishable from the full-grade), Ved. *jñātá-*), and **m* and **n* do not merger before the laryngeal: *-γνητός* and Hom. *ἐΰτμητος* ‘well-cut’ < **h$_1$su-tm̥h$_1$tos* (Sihler 1995: 104–6; Ringe 2024: 103). Phrygian seemingly does so too, at least the Middle Phrygian form *γλουρεος* ’golden’ seems to agree with Gr. *χλωρός* ‘yellow-ish green, pale’. Whatever the explanation is for analogical or secondarily accented double-reflexes (the type *θάνατος*), it is not shared with Indo-Iranian.

The liquids are also lengthened, in Greek with the triple reflex, and in Indo-Iranian with a merger of **rH* and **lH*. In Indo-Iranian, the outcome of this **r̥H* (**r̥̄*) is *īr* or *ūr* in Vedic, but *arᵊ* in Avestan: Cf. Hom. *ἄκρᾱτος* ’unmixed (of wine)’, Ved. *ā́-śīrta-* < **k̂r̥h$_2$tó-* and Gr. *στρωτός*, Ved. *stīrṇá-*, OAv. *starᵊta-* < **strh$_3$tó-* ‘spread-out’.

It has been alleged – and recently defended (Clayton 2022) – that the reason for the reflexes *īr*/*ūr* are distributed so that *īr* is the “standard” reflex, and *ūr* occurs after a labiovelar. This would imply that Vedic and thereby Proto-Indic and Proto-Indo-Iranian is not a true satəm language (cf. Chapter 2, 2.3.3.). While it fits some examples, e.g. Skt. *udgūrṇa-* 'raised' < **gʷl̥h₁tó-* (Gr. *ἀπό-βλήτος*) 'disposable, worthless' and *gūrtá-* < **gʷr̥h₂tó-* 'praised' – it does not work in **gʷr̥h₃-nó-* 'swallowed' > Ved. Skt *gīrṇá-* (cf. Gr. *βρωτός* 'edible'). It is clear by now that initial labials, *u*-vowels and *v* also trigger this, e.g. *pūrṇá-* 'full' < **pl̥h₁-nó-* (cf. Lith. *pìlnas* 'full') and *urvárā* 'arable land' < **h₂r̥h₃-u̯ér-eh₂-* (root otherwise unattested). Therefore, Sihler (1995: 93) explains the different reflexes of **gʷr̥h₂-* > *gurú-* 'heavy', *girí-* 'mountain' as due to the following labial vowel – but this would fail to account for *gūrtá-*. Clayton (2022: 46–47), instead, limits the scope of the sound law so that the rounding only occurs in closed syllables (**Kʷl̥H.C* > Ved. *Kŭ̆rV*), but Ved. *KirV* is to be expected in open syllables. This explains *girí-* and *gūrtá-* but in turn makes *gurú-* irregular. While I cannot explain all the details without turning to analogy (Lubotsky 1997b), neither can Clayton's account by which *gīrṇá-* is irregular and explained away due to its late attestation (R̥V 10). The attention to detail and the attestations should be applauded, but the extraordinary evidence required to back the extraordinary concluding claim that Vedic is no longer a proper satəm language has not been met. Unsatisfactory as it is, we must accept a few oddly oscillating roots.

4.3.6. Prevocalic "long" syllabic sonorants

It is worth reiterating one of the environments in which a PIE syllabic resonant had been lost before Proto-Greek, and we therefor observe the same developments in all dialects (Scarborough 2023: 93, n. 153): **CR̥HV* > **CəRHV* > **CaRV*. In these prevocalic instances of the "long vocalic resonants", we observe no trace of the vocalic laryngeal as its own phoneme, but we do see this special and early vocalisation of **R̥* - including **N̥* - to **aR*:

- *ἔκαμον* < **h₁e-k̂m̥h₂-ont* << **h₁e-k̂m̥h₂-ént* (Rix 2009: 74)
- *τανaός* < **tn̥h₂-eu̯ó-* (van Beek 2022a: 11)

This highly specific environment (**CR̥HV*) coincidentally overlaps with the environment process where **N̥* is realised as **aN* (**CN̥{H,R}V*) in Indo-Iranian (Lubotsky 2018: 1876). It would require a great deal of speculation to posit the overlap as an early, potentially shared innovation of pre-Proto-Greek and pre-

Proto-Indo-Iranian **CN̥HV* > **CaNHV* – but recall **-tm̥Ho-* > **-tamHa-,* not **-*tanHa-*.

This forced bottom-up reconstruction unfortunately does not help us much forward. It is mostly interesting from a Greek perspective that there is no triple representation of the laryngeals and no dialectal differences of the “prevocalic long sonorants”. While the argument was speculative to begin with, it is obscured by the fact that all Indo-European branches insert the prop vowel of the vocalisation before the resonant when vocalising the “prevocalic long syllabic sonorants”.

Obviously, the development cannot be shared with the branches in which the development is never **a* – Germanic **uN*, Balto-Slavic circumflex **iN/uN*. It is likely that the syllabic nasals were preserved in Proto-Italic since the **-m̥(H)m-* yields *um* in Sabellic (in the cases where it can be distinguished from /om/) but *om* in Latin and Venetic (Weiss 2022a: 118–9, n. 21). Weiss gives three suggestions: First, it is possible – but not based on any other available evidence – that the outcome was Proto-Italic **om* which gave **um* in a medial syllable in Sabellic. Second, the outcome could have been something unique, contrasting with the regular prevocalic reflex of **m̥*, e.g. **ɵ* - or third, it could have been preserved as **m̥* and later rounded independently in the branches. While “one loses the generalisation that all Italic languages show a rounded vowel in this environment”, phonemically, there is no great difference between positing an extra phoneme **ɵ* and retaining an extra phoneme /m̥/ - which probably was pronounced labially or with lib rounding. The regular outcome of **N̥* is *eN* in Italic, but **aN* initially in Sabellic – thus the phonemic mergers can hardly be of Proto-Italic date. In Tocharian, if **RH* does not yield “breaking” (vocalisation of **RH̥* > **Ra*) when *R* is **i̯* or **u̯*, and the **H* is *h_2 or *h_3, the laryngeal is dropped, and **R̥H* develops just like **R̥* to **əR: kärweñe** ‘stone’ < cf. *g^w*r̥*h_2*u-*, cf Gr. *βαρύς* ‘heavy’).

Celtic and Armenian show no special development of *N̥* in *CN̥HV*, and since their reflexes are *aN* anyway, it is impossible to exclude that they, too, took part in an “Indo-Greek” early innovation, later masked by the subsequent outcome of all other *N̥(H)*’s. It would be quite absurd to posit a shared – and extremely conditioned – shared development with Anatolian. The outcome of *-N̥H-* varies greatly. Intervocalically, all laryngeals yield geminates (Kloekhorst 2008: 81). *CN̥HsV* yields *CnisV*, and *CN̥HsC* probably yields *CnisC* and *CassiC* (Kloekhorst 2008: 73). It is very hard to come by examples of *CN̥HV*, but judging from *sa-an-ḫa-an-zi* <

$s\mathring{n}h_2$énti, at least *some* laryngeals would be preserved in this context and trigger what is *written* as vocalisation to *aN*. However, Kloekhorst interprets this as /snHént͡si/ which would then simply be the archaic state of affairs (Kloekhorst 2008: 70).

Thus, even when forcing the bottom-up principle for reconstruction, we do not even find an exclusive isogloss of the early development of **Cn̥HV* > **CanHV* in Indo-Iranian and Greek.

4.3.7. Semivowels

The semivowels **i̯* and **u̯* were inherited into Proto-Indo-Iranian. In Proto-Greek, **u̯* was also untouched. Word-initial **i̯*, on the other hand, split into **dʒ*- and **h*- on the way to Proto-Greek, which is shown by the split already haven taken place in Mycenaean: **i̯eugos* 'yoke of oxen' > Myc. dat. pl. *ze-u-ke-si* /dʒeugesi/ (Gr. *ζεῦγον*, Ved. *yugám*), but *o*- /ho-/ < **Hi̯o*- (Gr. *ὅς*, Ved. *ya*-) (Bernabé & Luján 2020: 200).

This split is puzzling. There is a growing consensus that it is connected to the presence or absence of laryngeals, but both **Hi̯* > **h*- and **Hi̯*- > **dʒ* have been suggested (Sihler 1995: 187–8). It seems that *at least* *h_1i* gives **h*- (Ringe 2024: 107; Byrd 2015: 234–236). Other laryngeals probably follow the same pattern (Weiss 1995; Clackson 2007a: 187).

Armenian does something similar, but not in the exact distribution (Olsen 1999: 787; Schmitt 1981: 70). It has recently been suggested that Albanian shares this isogloss with Greek (Hyllested & Joseph 2022), but the examples look formally compelling, they are not semantically self-explanatory: **i̯euh$_3$*- becomes 'silt, mudbed', 'soup' and 'sourdough'; **i̯es*- becomes 'knead' and 'boil' and 'battle' is paired with 'fondle' (Adams 2023: 226–7; Kölligan 2023: 332).

4.4. Laryngeals

4.4.1. Consonantic laryngeals

The latest common ancestor of Indo-Iranian and Greek would have had to preserve all three laryngeals as they yield the famous triple reflex in Greek. It is usually said that the three laryngeals merged into just one, noted **H*, in Proto-Indo-Iranian (Lubotsky 2018: 1880).

Avestan faithfully preserves this laryngeal metrically between two vowels – or, rather, Avestan faithfully preserves a hiatus in its place. Vedic does so occasionally (Gippert 1986; 1997), but many instances have been levelled, showing that the consonant was no longer phonemic. However, a hiatus is not necessarily the same as a phonemic consonant, and preserving the hiatus does not equal preserving a consonant.

On the other hand, if Lubotsky's (2018: 1879) interpretation of the Iranian reflexes of the "voiceless aspirates" arising from sequences of a voiceless stop before a laryngeal are in fact just the regular reflex of a stop leniting to a voiceless fricative in front of another consonant (**TH* > **ÞH* > *Þ*) rather than proto-Indo-Iranian aspiration (**TH* > **T*^{h} > *Þ*), there would be an additional argument for interpreting the hiatus as a recently lost consonant. The idea is that since all voiceless stops (**p t k*) turn to fricatives (**f θ x*) before another consonants (unless preceded by **s*) (Cantera 2017: 471), there would be no need to reconstruct the intermediate **T*^{h}, and it would even be more economical to assume that the fricativisation is one instead of two distinct processes. In that way, the mechanism that worked in PIE **tréi̯es* 'three' > PIIr. **trái̯as* > PIr. **θrai̯ah-* > Av. *θraiiō*, Parth. *hry* and PIE **k*^{w}*ék*^{w}*lo-* 'wheel, circle' > PIIr. **čakrá-* > PIr. **čaxra-* > Av. *caxra-*, Manich. *chr* would be the same as in PIE gen.sg. **pn̥th*$_2$*és* 'road' > PIIr. **patHás* > PIr. **paθHah* > Av. *paθō* and PIE **róth*$_2$*o-* 'wheel' > PIIr. **rátHa-* 'chariot' > PIr. **raθHa-* > Av. *raθa-*, Khot. *rraha-*, Phl. *ls*. While the economic argument is certainly attractive, it is admittedly impossible to prove whether these instances of PIr. **f θ x* went through **p*^{h} *t*^{h} *k*^{h}.

Additionally, if Hegedűs' (2012) analysis of the conditions of the *RUKI*-rule in Nuritani (treated below) is correct, consonantal laryngeals must have been preserved relatively late.

Finally, if there is any truth in the bold claim that some instances of **h*$_2$ and only **h*$_2$ surface as h- in Modern Persian, this consonant would have had to have remained distinct from **h*$_1$ and **h*$_3$ at least in initial position all the way from Proto-Indo-European through Proto-Indo-Iranian and Proto-West-Iranian, getting dropped multiple times independently over the course of the millennia (Kümmel 2018a; Cantera 2017 with refs.).

In Greek, the Indo-European intervocalic laryngeals never make position metrically, and all Greek dialects merge the resulting hiatus (as long as it does not begin with a high vowel) into a single long syllable: **g*^{w}*ih*$_3$*ós* remained disyllabic *βιός* ‘life’, but *$h_1eh_1\acute{e}st$* ‘(s)he was’ and *$h_1eh_2\acute{a}\hat{g}et$* ‘(s)he drove’ became PGr. **ēs* and **āge* showing only later dialectal variation: remaining *ἦς*, *ἆγε* in Doric, but the latter shifting regularly to *ἦγε* in Attic-Ionic (Ringe 2024: 91–2).

4.4.2. Long vowels from compensatory lengthening

Very similar to the other issues discusses above, it is very difficult to assess when the long vowels /ā ī ū/ were phonemicised. All Indo-European languages share the trait that laryngeals were lost with compensatory lengthening and colouring before at least a stop, but often any consonant. The details differ, and the process this cannot have taken place in the parent language. Most noticeably, Hittite preserves consonantic reflexes in more contexts than the other languages, e.g. *h_2 is preserved as *ḫ* before **s* and resonants.

Another prominent feature is the rise of the acute tone in Balto-Slavic and the other laryngeal-induced sound changes of this branch (Olander 2019b: 367–8). Since these developments did not take place outside Baltic and Slavic, the laryngeals must have been preserved at a relatively late date in the ancestor of these branches. To my knowledge, it is impossible to distinguish the three laryngeals in Balto-Slavic (excluding laryngeal colouring of an adjacent **e*), and it is thus possible to assume that they could have merged into just one consonant before becoming suprasegmental tonal traits.

Because of the difference in Balto-Slavic accentuation being directly linked to compensatory lengthening from the loss of laryngeals (Olander 2015: 47), it is untenable to posit **aH* > **ā* as a non-Anatolian innovation (*pace* Trager & Smith 1950). However, I suppose only circumstantial evidence or structural arguments can give hints at how long postvocalic laryngeals were preserved in the evolutionary history of the branches:

> CL [...] cannot be reconstructed for PIE, since there is no indication the sequence *-*VH$* became *$\bar{V}$ in the proto-language. Nevertheless, the fact that laryngeal deletion in the sequence *-*VH$* invariably results in CL in the daughter languages makes it highly likely that laryngeals were syllabified [...] in this position in PIE.
>
> (Byrd 2015: 105–6, n. 76)

In Greek, secure examples of *ih_1* and *uh_1* are hard to come by, and there is a certain risk of circularity in the argumentation if an otherwise uncertain laryngeal is reconstructed as *h_1* because of the absence of breaking in Greek. Further, it is impossible to exclude if the lengthening of (at least preconsonantal, word-internal, unaccented and monosyllabic) *ih_1 uh_1 > *ī ū* could be shared already in the common ancestor of Greek and Indo-Iranian, although it is by no means economic to assume so. It is also unclear whether we should reconstruct **ī ū* for Proto-Indo-European. To Ringe, **u̯īs-* ~ **u̯is-* 'poison' is a clear example of **ī*: Gr. **u̯īhós* > *ī̓ός*, Lat. *uīrus* – but the short *ĭ* of Ved. *viṣá-* might actually point to **u̯iH-s-*. Similarly, it can only be determined by circumstantial evidence or structural arguments if we should reconstruct **muHs* or **mūs* for 'mouse'.

4.4.3. Laryngeal breaking

As hinted above, I am more sympathetic than Lubotsky towards the idea of laryngeals causing "breaking", that is the sequence of a semi-vowel followed by a laryngeal yielding a semi-vowel and a long vowel in Greek – and most likely also Armenian, perhaps even in Albanian. Since Greek "laryngeal breaking" and Indo-Iranian lengthening do not align, long **ī* and **ū* from **iH* and **uH* must have been phonemically separate.

It is uncontroversial that Greek shows breaking of *-ih_2* and *-uh_2* word-finally (Ringe 2024: 99–103). This might also be the case for Armenian, judging from forms like *sterǰ* 'sterile, barren' < **sterih₂* (Gr. *στεῖρα* < **steri̯ă*, Ved. *starī-*) (Olsen 1992; 1999: 771). Also Phrygian shows word-final laryngeal breaking, at least of *-ih_2* > -*iya* (Obrador-Cursach 2019: 235). Examples of *-Yh_3* are hard to come by, and examples of *-Yh_1* are ambiguous. The clearest example is Hom. *ὄσσε* 'eyes' < *$h_3ok^wih_1$* 'both eyes' which does indeed look like *$h_3\acute{o}k^w\underset{\frown}{i}\underset{\circ}{h}_1$*. However, it is quite easy to imagine that this form would have been analogically recharacterized with the productive dual ending. The situation is very similar in Armenian, cf. *ačʿkʿ* 'eyes' which is obviously remarked with the synchronic plural ending *-kʿ*, but may also contain the productive plural ending *-a* (< *-h_2*) after the decline of the dual, so we cannot be certain if it contains *-i̯h̥₁* or *-ih_1* > *-ī* > *-i̯+a* (Olsen 1999: 773).

A directly comparable process also took place in Tocharian (Normier 1977: 182–4; 1980: 273). Olsen and Thorsø (2022: 211) argue that since the development is found in both Tocharian, Greek and Armenian, it can hardly be considered an exclusive

Graeco-Armenian isogloss. However, I will argue that this at least depends on what "shared" means. There is no reason to assume that it is to be interpretated directly such that the innovation spread when the languages were still in direct contact or even formed the same speech community, no matter if this is seen as a node in a tree or neighbouring diverging dialects: Although it is similar, Tocharian regularly and consistently vocalises laryngeals over more sonorous neighbouring sounds. Additionally, there is not necessarily any sign of length in Tocharian where the "broken" vowel appears only *graphically* long: TB /a/ (written <ā> under the accent), TA *ā*.

Although the development has been proposed for Greek for a long time (Francis 1970: 276–84; Normier 1977: 182, n. 26), it has not gained mainstream acceptance yet. At first glance it does seem random when Greek shows the "expected" outcome, i.e. a long vowel like the other branches (e.g. *θῡμός* 'soul' < $*d^huh_2$-*mó-* 'smoke', cf. Ved. *dhūmá-*), and when it shows the alleged "breaking" (e.g. *ζωός* 'living' < $*g^wih_3$*u̯ó-*, cf. Ved. *jīvá-*, Lit. *gývas*, Lat. *uīuus* where all other languages show a zero grade in the root (Klein 1988)). However, the "traditional" extra assumptions required to explain the "broken" forms are indeed very inelegant. Why would Greek suddenly introduce unexpected full-grades (as in *δηρός* 'long' < **du̯(e)*h_2*ró-*) or Schwebeablaut. I believe that these difficulties have now been convincingly explained by Olsen (2009) who argues that the Greek reflex is dependent on the accent: An accentuated **í*h_2 *í*h_3 *ú*h_2 *ú*h_3 yield **ī ū*, whereas unaccented **i*h_2 *i*h_3 *u*h_2 *u*h_3 yield **i̯ā i̯ō u̯ā u̯ō*.[14] This rule perfectly explains the distribution – expect for the form $*d^huh_2$*mó-* where Olsen should expect PGr. $*t^h$*wārós* and therefore had to assume that this root contained $*h_1$ and that the *ḫ* of the Hittite cognate *tuḫḫae-*^{zi} 'cough, be in need of breath' (a denominative verb derived from $*d^huh_2$*-o-i̯e/o-*) was onomatopoeic. Since it is now clear that word does in fact mean 'produce smoke', at least something a mountain or a volcano does, it is quite difficult to see what would be so inherently noisy about smoke (Kloekhorst 2008: 886–9). This "counterexample" has recently found an alternative explanation: There might be a consistent monophthongisation of *u̯*-diphthongs before labial consonants in Greek, which would mean that Gr. *θῡμός* would agree with Germanic and Italic where Far. *deymur* 'strong smell', MoDu. *doom* 'mist' and Lat. *fūmus* can go back to the *o*-grade

[14] Very similarly, seemingly independently, at least without reference to Olsen (2009) also noticed by Woodhouse (2015: 265–7).

d^ho̯uh$_2$mó- with loss of the laryngeal due to the "Saussure effect" (Kristoffersen 2019: 47; forthc. a). Lithuanian *dū́mai* and Ved. *dhūmá-* would then continue the usually reconstructed zero-grade. Olsen's rule has gained acceptance and paved the way for new etymological suggestions (Hyllested 2004; Batisti 2020; Kristoffersen forthc. b).

Armenian most likely also shows word-internal laryngeal breaking (Olsen 1992; 1999: 770–3), and it has even been argued that the development was accent-dependent (Woodhouse 2015: 268). But most importantly, Armenian does not show the same triple (or at least double) representation that Greek does: the vowel of the "broken" sequence is always *a*, no matter if the input contained *h_2 or *h_3 (Olsen 1999: 770–3). Thus, although the initial steps may to some extent be shared between Greek and Armenian, the phonemisation must be branch-independent and does not link these two branches any closer to each other than to Indo-Iranian – see below on the prothetic vowels for an elaboration of this argument.

4.4.4. Prothetic vowels

The prothetic vowels arising from vocalisation of laryngeals in absolute initial preconsonantic position are often said to be a shared trait of Armenian, Phrygian and Greek – but obviously excluding Indo-Iranian and all other branches in which they presumably stayed consonants until they were lost. Phrygian does seem to agree with Greek in the distinction of three vowels if the few examples can be accepted: OPhr. *ev-* 'good' : Gr. *εὐ* (Hom. also *ἐϋ-*) < **$h_1$su-* (cf. Ved. *su-*, Av. *hu-*), NPhr. *αναρ* 'husband' : Gr. ἀνήρ 'man' < **h_2nēr* (cf. Ved. *nar-* 'man') and OPhr. *onoman*, NPhr. *ονομαν* 'name' : Gr. *ὄνομα* 'name' < *h_3n$h_3$mn̥ (cf. Ved. *nāman-*) (Obrador-Cursach 2019: 234; Kim 2018a: 253). The matter is more complicated in Armenian where there are conflicting analyses. Martirosyan (2010: 716) interprets the evidence as relics of a triple reflex as in Greek, whereas Olsen (1999: 762–3) sees a full merger of all three laryngeals into a single prothetic vowel **a-*.

Everyone agrees that – at least – **h_2-* yields *a*, as in *ayr* < **anīr* < **h_2nēr* 'man'. However, unlike in Greek, there is *no* prothetic vowel in front of **-u̯-*: Arm. *goy* 'is' : Gr. *ἄεσα* '(s)he spent (the night)' < **h_2u̯es-* (cf. Hitt. *ḫu̯iš-^{zi}* 'live') which makes it difficult to imagine the merger with **a-* at an early date; since in such a scenario only **a* < **H-* would be lost in front of **-u̯-* (Clackson 1994: 36). Interestingly, Olsen (1999: 763) also cites Arm. *gełmn* 'fleece, wool' < **h_2u̯elh$_1$mn̥(t)-* (cf. Hitt. *ḫulana-*

'wool') but does not give the Greek cognate *λῆνος* 'wool' < **h_2u̯l̥h_1-n-* which shows an expected "early loss" of the initial laryngeal (Olsen 2023: 18).[15] There is also clear evidence for **h_3- > a-*: *akn* 'eye' < **h_3k^w-* (Martirosyan 2010: 23; Olsen 1999: 763). However, whether **h_3-* behaves differently in other contexts is less clear. Especially in front of **-m-* where Martirosyan (Martirosyan 2010: 715) proposes a special development **h_3mV́- > *omV́- > *umV́- > *mV́-* to account for *mēz* 'urine' (Gr. *ὀμείχω*) and *mēg* 'mist' (Gr. *ὀμίχλη*), whereas Olsen keeps more options open: they could be non-indigenous or they could have lost the laryngeal by morphophonemic means (Olsen 1999: 763).

4.4.4.1. Nameless teeth without a bite?

**h_1-* is notoriously difficult. Again, in favour of a triple reflex, Martirosyan (2013: 89) cites *inn* 'nine' (Gr. *ἐννέα*) < **h_1ńeu̯n* and *erekʿ* 'evening' (Gr. *ἔρεβος*, Go. *riqis* 'darkness') < **$h_1$reg^{w}os*. The first is problematic because of the geminate *-n-* in Greek which seems to point to a metathesis or even a compound with another element (Olsen 1999: 764; Clackson 1994: 124–6), the second revolves around an automatic reconstruction of **Hr-* rather than **r-*, but it is difficult to assess *when* the prohibition of initial **r-* should be dated and thus if it is of Proto-Indo-European age (Olsen 1999: 764; 2011: 22; Clackson 1994: 33). Martirosyan (2010: 714) further adds a specific rule, namely that the prothetic vowel did not develop or was lost in the sequence **$h_1$le$_/$a-*. This would account for *lanǰkʿ* 'breast' (< *'lungs'), which Olsen (1999: 65–6, 763 n. 5) also cannot explain. I would, however, doubt the etymology altogether – I do not see a forcing semantic connection between *ἐλαχύς* 'small, short', *ἐλαφρός* 'light (in weight)' and *lanǰkʿ* 'breast' (*pace* Martirosyan 2010: 304; Olsen 1999: 65).

Olsen's (1999: 762–3) examples of **h_1- > a-* are not unproblematic either. She provides 5 – the last two are copounds with **$h_1$su-* for which other explanations may be possible (as indicated by her own added question marks. The first is *aloǰ* 'she-kid' < **$h_1$lm̥b^{h}ih_2* (Gr. *ἔλαφος* 'deer') which is a semantically attractive solution to an

---

[15] The lack of an initial **ἀ-* in this word is indeed a problem for the etymology of the entire Wortsippe and its connection with Anatolian (Kloekhorst 2023). However, the connection with the verbal root **h_2u̯el-* 'pluck' (Lat. *u̯ellō*) as well as the understandable derivation of the attested forms still make the connection attractive in my opinion (Viñas-Caron et al. 2024: See further; Olsen 2018b; Pinault 2016; Adams & Mallory 1997).

otherwise unexplained word, but it is formally difficult (Olsen 1999: 196; Martirosyan 2010: 16 (rejected with no comment)). The following two are based on problematic reconstructions – I do not doubt that they are evidence of a prothetic vowel, but I find it unlikely that this prothetic vowel can be examples of *h_1-.

4.4.4.2. anown : ὄνομα < ενυμα : nāman-

For *anown* 'name' the reconstruction with *h_1- in *$h_1noh_3mn̥$- is allegedly secured by the Laconian personal names *Ενυμακρατιδας*.[16] Some scholars see this as the *forma difficilior* and thus project the *e* < **h_1*- back into Proto-Indo-European, whereas other consider it unrelated, an "(incidental) vowel assimilation" or a dissimilation of *h_3neh_3mn- to *h_1neh_3mn- (Robert S. P. Beekes 2010: 1085). The suggestion of a vowel assimilation is quite strange; one could ask where the second *e* is? Surely, the *ē* of the final syllable is too different in quality and too many syllables away to play a role. The dissimilation of laryngeals is technically possible but completely *ad hoc*.

If, on the other hand, *h_1noh_3mn- is the original stem, the initial *o*- of all other dialects of Greek and Phrygian (and possibly even Armenian, if Martirosyan is right) would have to be independent late vowel assimilations (**enomn*- > **onomn*-). The existence of *νώνυμ(ν)ος* < *n̥-$h_3nh̥_3m(n)$- 'nameless, inglorious' (only attested with Cowgill's law (Vine 1999)) point to original *h_3 or some unlikely analogy by which the speakers recognised an alternative long vowel privative prefix *νω*- and extended it beyond words beginning in etymological *h_3.

While no language preserves the no doubt underlying ablauting paradigm, most variation concerns generalisations of the full grade or zero grade of the "root". Both Tokharian languages point in a different direction: TB *ñem**, TA *ñom* < PT **ńemə* whose palatalising /'e/ looks like a Proto-Indo-European long vowel *$h_3nēmn̥$ (Adams 2013: 288). Alternatively, the long vowel can be explained as a long grade *$h_3nēh_3mn̥$ (Ringe 2024: 110). Similarly, but not identically no doubt because he does not accept Eichner's Law, Beekes also mentions the possibility of a dissimilation of *$h_3neh_3mn̥$ > *$h_3neh_1mn̥$ (the reverse of the suggestion for Greek if *h_1- is original). Unfortunately, Tocharian only allows us to recognise the existence

[16] "Lacon. ***ἔνυμα*** prob. in pr. nn. Ἐνυμακρατίδας *IG* 5(1).213.45" (*LSJ*: s.v. ὄνομα). See also IG 5.1.213 itself.

of a long *$\bar{e}$*, but if this was morphological and original (and in combination with h_1 og h_3) is impossible to tell (Eichner 1973).

The etymon is attested in Anatolian, among others as Hitt. *lāman* 'name, reputation' and HLuw. /alaman/. The latter is attested multiple times, and it is always spelled with initial *á-* (never *a-*), e.g. *á-la/i-ma-za* and *á-lá/í-ma-za* /alaman⸗za/ (*eDiAna*: s.v. lemma 1304).[17] It has been suggested that the lack of *ḫ-* in Hittite *lāman* – even although it appears to be dissimilated from expected **nāman* – would speak against *h_3-. According to Kloekhorst (2006; 2008: 518), however, initial preconsonantal *h_1 and *h_3 merge in Proto-Anatolian and surface as *h_1. Sadly, even if this is correct, the absence of ***ḫ* in Hittite and the consistent spelling with <á> in Hieroglyphic Luwian, which has been suggested to represent /ʔa/ < *$h_1(V)$- (Kloekhorst 2004; Simon 2013), would not allow us to distinguish between original *h_1 and "neutralised" *h_3. If the connection with the verb Hitt. *ḫanna^i/ḫann-* 'to sue, judge' < *'to call into court' and the Greek verbal stem *ὄνομαι* 'to scorn, call names', aor. *ὦνατο*, is not secondary to *ὄνομα* but goes back directly to a root *h_3neh_3- 'to call (by name)' the laryngeal can securely be reconstructed as *h_3 (Kloekhorst 2008: 284).

4.4.4.3. atamn : ὀδούς < ἔδοντ- : dánt-

Similarly, in the word for 'tooth', Olsen reconstructs *h_1d-$\mathring{n}t\mathring{m}$ for Arm. *atamn* and compares Gr. *ἔδοντες*. Olsen is by no means alone in this connection, going back to the wish to connect 'tooth' with the root 'eat' (Gr. *ἔδω*) supposedly originally meaning *'bite'. I will readily admit that the semantic connection is attractive: *h_1d-*ónt-* does indeed look like the preform of the Hittitite participle *adant-* 'eating, eaten' and e.g. Ved. *dán-*, gen. sg. *dantáḥ* 'tooth' (*NIL*: 208–20). However, I find it a bit speculative to reconstruct the original – unattested – root aorist with the punctual meaning 'bite' on the basis of nominal derivations (Schindler 1975: 62). This aorist – and the original meaning – should have been lost already in Proto-Indo-European (*LIV*^{2}: 230).

Just as in the word for 'name', an assimilation must be assumed for Greek to arrive at Gr. *ὀδούς*, gen. sg. *ὀδόντος* (non-Homeric Ion., e.g. Herodote, also has nom. *ὀδών*) < **odont-* from **edont-* (Weiss 2020: 55). The reconstructed stem *h_1*dónt-* should

---

[17] These forms occur on KARATEPE 1 Hu §75, §74. On the reading of the sign *319 as <la/i> (and not <ta_4>) and *172 as <lá/í> (not <ta_5>), see Rieken & Yakubovich (2010).

be continued in the form *ἔδοντες* 'teeth' which is claimed to be Aeolic (*LSJ*: s.v. *ὀδούς*). However, just like *ενυμα*, the *ἔδοντες* has its own origin story and problems associated with it. The text securing the "Aeolic" origins in the *LSJ* is *in Platonis Cratylum commentaria* by the 5th century CE scholar Proclus Philosophus (Procl. *in Cra.* 85) reads:

> Ὅτι τὸν ἐτυμολογήσειν μέλλοντα δεῖ τὰς τῶν διαλέχτων εἰδέναι διαφοράς, τοὺς γὰρ ὀδόντας ἔδοντας καλοῦσιν οἱ Αἰολεῖς. δεύτερον καὶ τὴν χρῆσιν τῶν ποιητῶν
>
> (Pasquali 1994: 39)

> One intending to perform etymological studies should know the differences between the dialects. The Aeolians, for example, call teeth (*odontes*) *edontes*. Second, he should know the usage of the poets as well
>
> (Duvick 2007: 47)

Matthew Scarborough, who spent a great deal of time discussing this with me, has also noted: "[A]s for Aeolic ἔδοντ-, it only occurs in late grammatical texts and could have itself been created on the basis of an analogy to Greek *ἔδ-* 'eat'." (Scarborough, *IE-CoR*: s.v. "Cognate Set 328"). This "analogy" with *ἔδω* is similar to the original derivation of the noun from the root **$h_1$ed-* 'eat' (< *'bite') already in (pre-)Proto-Indo-European, and I may add that it could in principle have happened at any time – be it in the mind of Proclus, in an actual dialect spoken in the common era, in BCE Aeolic, in (Pre-)Proto-Greek or in Proto-Indo-European. The problem with basing the Proto-Greek and consequently also the Proto-Indo-European reconstruction of the word for 'tooth' on this form is that the most archaic form would have survived until the 5th century CE with no traces in the ancient dialects and apparently parallel assimilations of **edont-* > **odont-* in all other dialects – including Mycenaean, where *odakwenta* and *odatwenta* 'toothed' (of chariot wheels, possibly i.e. 'with hobnails') are attested in various spellings (Thompson & Meißner 2024: 103).

Completely analogue to *νώνυμος* 'nameless, inglorious', there is also a privative compound **n̥-h_3d-* > *νωδός* 'toothless'. It is slightly younger, *νώνυμος* and for metric convenience also *νώνυμνος* occur already in Homer and Pindar, whereas *νωδός* is found in Aristotle and Aristophanes. Again, it is not impossible to think that these forms were analogical after etymologically expected negated words in *νω*- from *$*\mathring{n}$-h_3-*.

Summing up, it is striking that two of the best instances of prothetic *h_1 in Armenian only show evidence of *h_1 in scarce relics in Greek (a personal name in a Laconic inscription and a 5th century commentary on etymology), but show positive evidence for *h_3 in the "standard" language, in Homer and in privative compounds.

4.4.5. Are the Græco-Armenian prothetic a hindrance to "Indo-Greek"?

Returning to the question of the alleged shared Greco-Armenian innovation of the prothetic vowels form laryngeals – and to its relevance for the comparison of Indo-Iranian and Greek, it is safe to say that the Greek and Armenian – unlike the Greek and Phrygian – developments can hardly have happened when these languages formed a subgroup: Only of the phonemic merger of the vocalised laryngeals and the inherited vowels **e a o* can be demonstrated for a common prestage of both branches with no intervening sound changes can it be considered a true shared innovation.

If Armenian indeed had a full triple-reflex (and the change of **o*- > *a*- in open syllables (Frederik Herman Henri Kortlandt 1985: 9; Frederik Kortlandt 1983) is correct), this would be the case. However, the fact that Armenian does not vocalise a laryngeal before *-*u̯*- tells that the entire merger cannot have been shared. If the special treatments of *h_1 in front of **l* and *h_3 in front of **m* added by Martirosyan (2010: 714–5) hold true, the triple reflex looks even less shared, and the mergers must be einzelsprachlich: if **e* from *h_1- and **o* < *h_3 can behave differently from *e and *o, they cannot have merged phonemically, and it is a matter of notation if we prefer *e_2 o_2 or *$ə_1$ $ə_3$ or *$\mathring{h}_1$ $\mathring{h}_3$. Now, if Olsen (1999: 762–4) is correct, and Armenian does not preserve a triple reflex of the laryngeals, any phonemic merger becomes impossible. Although I have been very critical towards here examples above, I will add that *aloǰ* 'female kid' < *$h_1l\mathring{m}b^hih_2$ seems plausible, and that the examples of *h_1 > **e* (*inn* and *erek*) are indecisive at best.

Summing up, it is not possible to formulate the Greek and Armenian prothetic vowels as "a true, homonymy-producing, irreversible merger" in the words of Hoenigswald (1966: 7, n. 1)[18]. This means that the isogloss links Armenian no closer

[18] though on the Italo-Celtic assimilation of **p…k^w* > **k^w…k^w*, not on Greek and Armenian.

to Greek than to Indo-Iranian, strictly speaking. However, it is striking that it is only these few languages who undergo so similar innovations so different from all other Indo-European languages. The conclusion reached by Clackson (1994: 34–6) that it might be an areal feature is definitely an option, but I will object to Martirosyan's (2013: 23) addition of the Anatolian preservation of the laryngeals in this areal feature – that is simply an archaism, albeit the same archaism as necessary for the latest shared ancestor of Armenian and Greek. Further, I dare not make any claims of when and where the pre-proto-Greeks, pre-proto-Armenians and pre-proto-Anatolians last were in a contact situation intense enough to warrant the preservation of fricatives in front of consonants.

Although Hoenigswald's methodology is strict and worthy of praise, I will stress that I am open for the possibility that Armenian and Greek shared some initial tendencies of the vocalisation of the laryngeals when – if – they formed a subgroup. This excursus is only to elaborate that it can neither be demonstrated that Indo-Iranian did not take part in this "initial vocalisation" as well, nor that it was exclusive to Graeco-Armenian since the phonemic mergers are demonstrably later than any shared pre-stage.

4.5. The stop system

4.5.1. Place of articulation

To get the most trivial points out of the way: Both branches continue labials (**p b b^h*), dentals (**t d d^h*) – and some velars (**k g g^h* plus **k̂ ĝ $\hat{g}^h$* and **k^w g^w g^{wh}* respectively). Indic, Nuristani and far-easter Iranian has acquired retroflexes as well.

Much has been written on the rarity of the Proto-Indo-European phoneme */b/. It remains questionable if it should be reconstructed for the oldest stages and what the reasons for its peculiar rarity are. According to Clackson's (2007b: 41) count, it the rarest of the 25 reconstructed consonants in the *LIV*[2]. There are various interesting solutions to this curiosity. Perhaps the rarity is to be seen in connection with the relative frequency of **m* and **u̯* and traced to some pre-Proto-Indo-European morphophonemic rule (Ringe 2024: 16–7) or perhaps some cases of **b* have been obscured by later sound laws, like Kluge's Law in Germanic (Kroonen 2011: 253–5).

For the so-called "clash of dentals", *tt > *tst and *dzd, where Old Hittite preserves the Indo-European state of affairs, e.g. *az-za-aš-te-ni* /atsteni/ 'you eat' < *h_1d-th_1e-, Greek and Iranian (and Balto-Slavic, for that matter) reduce the cluster to *st (e.g. **u̯oid-th₂a* > Gr. **u̯oitʰstʰa* > *οἶσθα* (see below for this form), Av. *vōista*), but Indic and Nuristani to *tt (cf Ved. *véttha*) (Ringe 2024: 121–2; Kümmel 2022: 255). This is a late and insignificant development and not to be assigned to any common dialect areas (*pace* Anttila 1989: 395; Kümmel 2022: 263).

Even under a "radical" phylogenetic approach, **b* is certainly to be reconstructed for the latest shared stage of Indo-Iranian and Greek, both of which continue word-initial and medial **b* (Olander 2022a): **bel-* 'strong' > Gr. *βέλτερος* 'better', Ved. *bála-* ''strength'.

4.5.1.1. Centum and Satem

Solving the centuries old problem of the centum-satəm isogloss is well beyond the scope of this chapter. In the search for a later common ancestor than PIE itself, the following quote is applicable to Greek and Indo-Iranian as well:

> The *centum/satem* division is a matter of embarrassment to those who have tried to see a close genetic relationship between Armenian and Greek.
>
> (Clackson 1994: 54)

Modern-day handbooks tend to reconstruct three "velar" rows (Clackson 2007b: 51–3; Beekes 2011: 122–3; Weiss 2020: 93–102; Ringe 2024: 5). However, the communis opinio has shifted over the years: In the very beginning, only two rows were reconstructed, until Bezzenberger's three row-analysis became mainstream (Bezzenberger 1890). About 50 years ago, the rise typological awareness made it shift again (Steensland 1973; Miller 1976; Kortlandt 1978). And now, the proposed distinction of plain and palatovelars in Luwic has made the opinion shift and settle again (Melchert 1987; 1989; 2012; Kloekhorst 2008: 17–9; 2022: 68).

While the three-way distinction is typologically unlikely, poorly supported by the material, and the relative frequencies suspicious,[19] attempts to solve this problem by

---

[19] A corpus-based approach finds that the plain velar are overrepresented before and after **u* and **u̯*, after **s*, **n* and **n̥* and before **r* and **r̥* (Olander 2024a).

reconstructing only two rows have been unsuccessful. While they may be theoretically attractive, the required rules of phonemisation are unlikely and opaque (Kümmel 2022: 258), and such scenarios are difficult to reconcile with the Armenian and Albanian material (Schumacher & Matzinger 2013: 237–43; Schmitt 1981: 63–5; Olsen 1999: 805–11). Another approach, to reconstruct three typologically likely rows requires every single branch to innovate (Kümmel 2007: 310–327). Reconstructing uvular stop instead of plain velars further clashes with the theory that the "Proto-Indo-Anatolian" laryngeals were uvular(?) stops (Kloekhorst 2016; 2018).

I am sympathetic towards the bold statement of Sihler (1995: 154)

> this is an artifact of the method, not a picture of the early history of PIE: there never was a variety of PIE with three dorsal stops

but as the matter is unsolved, there is no way around reconstructing *k g g^h, $\hat{k}$ $\hat{g}$ $\hat{g}^h$ and *k^w, g^w, g^{wh} for the latest common ancestor of Indo-Iranian and Greek. By Proto-Greek, these had merged in the centum system *k g g^h > k^h, k^w g^w g^{wh} > k^{wh}, and by Proto-Indo-Iranian they had undergone the satəm-merger of plain- and labiovelars, and the palatovelars had, presumably, become fully-fledged palatal stops (or palatal affricates; at any rate with preserved occlusion): *k g g^h, **ć*, *ȷ́*, $\acute{ȷ}^h$. The former palatalised to **č ǰ* $ǰ^h$ and were phonemisised with the vowel merger. In Old Indic, **ć* has become *ś* which has the synchronic stop counterpart c^h < **sč/sć*, **č* is preserved (written *c*), as are k g g^h, **ȷ́ ǰ* had merged in *j*, and **ȷ́* $ǰ^h$ in j^h > *h* (Kobayashi 2004: 13). In Proto-Iranian, primary and secondary palatals are kept distinct, and the system is **k g (x, γ), ć ȷ́, č ǰ*. **č* and *ǰ* are preserved in Avestan (transcribed *c/č* and *j/ǰ*), but yield /ts/ <tc> in Khotanese (palatalised to /č/ <c, ky>; see also Article 3) and /dz/ <js> (pal. /ǰ/ <j, gy>). In all descendants, **ć* and **ȷ́* yield dental affricates, fricatives or sibilants: Av. *s, z*; OP *θ*, **ð* > *d*; Khot. *s-*, *z-* <ys> (but -*ś*- <śś>, -*ź*- <ś> in clusters (Gercenberg 1981: 242–9; Skjærvø 2022: 123); it is therefore debated if PIr. **ć* was already *ts or still *tɕ or *tʃ (Cantera 2017: 492–3). Similarly, Proto-Nuristani had **k*, *g*; *ts* (written *ċ*; and notably **ćr* > retroflex *c̣*), *dz* (*j*) and *č*, *ǰ* (Budruss 1977; Nelson 1986).

4.5.2. Manner of articulation

It comes as no surprise that the inventory of stops reconstructed by the for Proto-Indo-European by the Neogrammarians and later scholars who worked before the decipherment of Hittite and discovery of Tocharian has to be reconstructed for the latest common ancestor of Indo-Iranian and Greek.

Notably, these two branches are the only ones to directly continue or show aspiration of the mediae aspiratae – though in a modern understanding, the aspiration in Indic is technically breathy voice rather than aspiration proper. Although they are distinguished in other branches, aspiration as such is not attested, but reflexes of it – such as fricatives in Italic or "sound shifted" outputs in Germanic and Armenian – assert the phonemic contrast.

The traditionally-reconstructed (post-Brugmannian) system of tenues, mediae and mediae aspiratae, *T D D^h is asymmetrical and probably unattested in the languages of the world – although Kelabit which has *T D $^DT^h$ comes close (Kümmel 2012; Weiss 2009). In Greek, the system has been normalised by devoicing the voiced aspirates yielding the much more common system of *T T^h D, in Indic, sound changes and (Dravidian) loanwords have given rise to a four-way contrast *T T^h D D^h, and in Iranian the voiced aspirates haver merged with voiced stops yielding *T D.

The "Indo-Greek" consonant system is very different from the Proto-Anatolian stop system, which has only two rows that are primarily distinguished by strength or length than by voicing and/or aspiration – the Anatolian fortis series corresponding to the voiceless and the lenis corresponding to the voiced and voiced aspirates. The "simple" merger of the mediae and mediae aspiratae is mirrored in many other branches, and we can be certain that the three-way distinction recoverable from the non-Anatolian branches is the older situation. It would be interesting to entertain the thought that Anatolian preserved the older stop system, and the three-way distinction was a later innovation. However, the complete lack of distribution makes it incredibly difficult to suggest that the third row of the non-Anatolian languages could have emerged out of an earlier stage.

Under a glottalic framework, the Proto-Anatolian system is somewhat closer to the original, since voice and aspiration play only subphonemic roles.[20] Proto-Anatolian fortis stops **t:* correspond to "Indo-Greek" tenues **t*, lenis stops **t* correspond to

[20] The glottalic theory has been presented in several different flavours. I here follow Kloekhorst (2016), since I find the preglottalized voicing [ˀd] phonetically implausible and incompatible with the material.

mediae *d and are reconstructed as preglottalized lenis stops *^{ʔ}t, and the "Indo-Greek" mediae aspiratae are reconstructed as plain invoiced stops *t – possibly with "facultative" aspiration.

I will not dive into various glottalic reinterpretations of this inventory, but I will give the following four remarks. First, I find it remarkable that Kloekhorst (2022) and Peyrot (2022) – in the same volume (Olander 2022d) – argue for the same re-interpretation of the reconstructed proto-Indo-European stop system based on the very different developments in Anatolian and Tokharian. Based on the fortis-lenis distinction in Anatolian, Kloekhorst argues for a more Anatolian-like distinction without voice in Proto-Indo-European. The merger of the "weaker" rows, *^{ʔ}t t – transitionally reconstructed as *d d^h, into the Proto-Anatolian lenis stops *t is indeed incompatible with the reconstruction.

In most circumstances, Tocharian merges all three rows into a single voiceless row, with crucial details that *b^h is lost after *m – indicating that the mediae aspiratae were different from the other two rows and some point in the prehistory of Tocharian, and that *d yields a different result (*ts*, palatalised to *ś*; also lost before *i̯ and *u̯) than *t and d^h (merging in *t*, palatalised to *c*) (Peyrot 2022: 91–2). On the basis of this, Peyrot concludes (references reformatted by me):

> Although Tocharian offers no direct evidence for the reconstruction of glottalic stops in Proto-Indo-European, the fact that *d has a different reflex from *t and *d^h is neatly compatible with it, since under Kortlandt's glottalic theory (e.g. 1985; 2018) *d [ˀd] on the one hand is set apart from *t and *d^h on the other.

I agree that the Tocharian material does point to an earlier three-way distinction, but I cannot see how this distinction is more compatible with Kortlandt's *t ^{ʔ}d d (or Kloekhorst's *$t\!:$ ^{ʔ}t t) than with the traditional *t d d^h - loss of *b^h after *m point to voicing, as Peyrot also explains, and the more frequent loss of *d does point to this row being "weaker" than the two others, which it is under all approaches.

Second, in the glottalic reconstructions, voicing and/or aspiration is seen as facultative: *$t(\!:\!)$, ^{ʔ}t~^{ʔ}d, $t(^h)$~d (see the summary of different approaches in Kümmel 2012: 293). It is of course very plausible that the phonetic realisation of the phonemic contrast would be optimised as much as possible; but the phonemisation of such a subphonemic traits are not trivial. Since Greek *t d t^h matches Indo-

Iranian $*t\ d\ d^h$ (based on Indic $*t\ t^h\ d\ d^h$ and Iranian $*t\ \theta\ d$) more than Anatolian $*t\!:$ t, phonemic voicing of $*d$ and aspiration of $*d^h$ would have to be a shared innovation at some point in their shared prehistory (Kümmel 2012: 309–10).

Third, Kortlandt's (1981a) claim that the preglottalized stops of Sindhi should directly continue the Proto-Indo-European state of affairs is very difficult to imagine in a phylogenetic perspective. I will even argue that it is equally impossible to imagine under a wave-model approach. Since neither the closely related New Indic languages, nor any Middle Indic prakrit, nor any variety of Old Indic preserved the Sindhi glottals, all other Indic, Indo-Iranian and Indo-European branches would have had to undergo identical but independent sound shifts. Even under the most generous hypothetical circumstances, if Sindhi were to be the outlier within Indic, and Indo-Iranian were to be an outlier within Indo-European, the innovation would at least have happened in non-Sindhi-Indic, Nuristani, Iranian and non-Indo-Iranian-Indo-European. Under the assumption that the innovation spread as an areal feature, this wave would have spread across the Indo-European speech community over incredible distances in time and space.

Fourth, and returning to the differences between "Indo-Greek" and Anatolian, also other reinterpretations of the peculiar asymmetric Proto-Indo-European stop system. While the glottalic reconstruction fit (some) of the synchronic data somewhat better than the traditionally reconstructed stops – i.e. breathy voice is lost everywhere but in Indic and possibly some Armenian dialects – it is very difficult to explain especially the rise of the aspiration contrast diachronically (Kümmel 2007: 47–53; 2012: 296–9). Weiss (2009) and Kümmel (2012: 303–6) have explored the possibility that the peculiarities of the Proto-Indo-European stop system could be due to an original contrast between voiceless, voiced and implosive consonants in Pre- or early Proto-Indo-European, **t ɗ d*. A shift of this system to $*t\ d\ d^h$ is perhaps possible. But while it gives a more typologically plausible starting point, it does not remove the problematic and typologically uncommon doubly marked "middle step". The system further has three downsides: First, it is just as far away from the observable stop systems in the ancient languages as the glottalic systems. Second, cross-linguistically, **ɓ* is the most common implosive consonant. In this reconstruction, it would be the one shifting to **b*, the rarest Proto-Indo-European stop. We would thus have to incorporate the morphophonemic processes speculated above also into this system (Kümmel 2012: 304–5). Third and conversely, guttural implosives are cross-linguistically rare, which makes PIE $*\hat{g}^h\ g^h$

g^{wh} even more suspicious if they were indeed [ɠʲ] (or [ʄ]?) [ɠ] and [ɠʷ] – or, in Kümmel's own reconstruction **ɠ ʛ ɠʷ* shifting to **g̈ʱ G̈ʱ g̈ʱ* (Kümmel 2012: 306; 2007: 324–7) – all of which are incredibly rare or even unattested sounds in the languages of the world. Thus, the typological issue might have been solved by positing another typologically implausible system.

At any rate, under this reconstruction, Greek and Indo-Iranian both belong to the group of languages that have undergone "the Central IE sound shift" – that is, they show the shift of implosives to voiced stops and of original voiced stops to breathy voice-stops (Kümmel 2022: 257). Under a glottalic reconstruction, only Indic underwent this shift.

4.5.3. Voiceless aspirates?

As elaborated above, the breathy voice stops of Indic correspond to voiced stop in Iranian and to voiceless aspirates in Greek. Traditionally, under a non-glottalic reconstruction, it is assumed that these were originally breathy voice consonants that were devoiced in Greek – because spontaneous and unconditioned voicing, aspiration and breathy voicing are typologically implausible.

There is, however, a small group of cognates in which the voiceless aspirates of Greek correspond to voiceless aspirates of Indic – and to voiceless fricatives in Iranian. This initially lead scholars to reconstruct a series of voiceless aspirates for the Proto-Language (Brugmann 1886: 406–8). It has later been backed by typological evidence – having D^h without T^h is excessively rare – but trustworthy examples of T^h are still hard to come by. It is certainly the case that some Iranian fricatives **f θ x* correspond to Vedic tenues aspiratae p^h t^h k^h, but the connection with Gr. *φ θ χ* is much less certain.

The clearest examples are with the dental: Ved. gen.sg. *patháḥ,* YAv. *paθō* < **patʰás* 'path'; Ved. *rathá-*, Av. *raθa-*, OP *u-raθa-* < **ratʰa-* 'chariot'; Ved. *pr̥thú-*, YAv. *pərəθu-* < **pr̥tʰu-* 'broad' and PERF.IND.2SG **dʰadʰātʰa* 'you have put' > Ved. *dadhā́ta,* Oav. *dadāθa*. After */s/, there is no fricative in Iranian, and after *ruki-*š*, /tʰ/ is realised as [ṭʰ]: **s(t(ʰ))i-stʰa-* 'stand' > Av. *hišta-*, Khot. *ṣṭa-*, Ved. *tíṣṭha* (Cantera 2017: 491; Kobayashi 2017: 332). For the labial, **ćapʰá* 'hoof' > Ved. *śaphá*, Av. *safa* and **kapʰa* 'slime, foam' > Skt. *kapha*, YAv. *safa* are relatively clear (Hoffmann & Forssman 1996: 94). Sibilants also block the fricativisation of labials in Iranian: PRES.INJ.3.SG **spʰr̥H-at* > Ved. *sphūrat*, YAv. *-sparat̰*. The velars are a little more difficult as both languages palatalise further. Relatively clear examples are PIIr. **sák*^{h}*āi̯*- ‘companion’ > Ved. *sákhāy*-, Av. *haxāii*- (dat.sg. *haš́e* < **sak*^{h}*ai̯*), **k*^{h}*umb*^{h}*a*- ‘jar’ > Ved. *khumbhá*-, YAv. *xumba, xuṇba* and perf.ind.2sg. **u̯au̯akt*^{h}*a* ‘you have spoken’ (of the root **vak/vač*-) > Ved. *uváktha*, YAv. *vauuaxδa*.

Some scholars reconstruct these to PIIr. **p*^{h} *t*^{h} *k*^{h} - and some even automatically add the exotic **ć*^{h} < **k̂*^{h} and the palatalised **č*^{h} < **sk*^{h}*’* (Hoffmann & Forssman 1996: 93). The last two should be dismissed: PIE **k̂*^{h} only exists after **s* as a result of Sieb’s Law (*LIV*^{2}: 547; Kobayashi 2017: 332; Ringe 2024: 124). It may be phonetic, but it is hardly in phonemic contrast. See further (Alexander Lubotsky 2001) on the merger of **sć* < **sk̂* and **sč* < **sk*/_[+front]. Similarly, Indic seems to merge the palatalised **sk*^{h}*’* with palatalised **sč* < **sk* and palatal **sć* < **sk̂*, so that the outcome is *skh*-, but all three merge in *ch*- before a front vowel.

However, we should not reconstruct **p*^{h} *t*^{h} *k*^{h} for Proto-Indo-Iranian either. Those sequence that are not due to devoicing by Sieb’s law are better explained as preservation of the laryngeal **pH tH kH*. This gave rise to the series *p*^{h} *t*^{h} *ṭ*^{h} *c*^{h} *k*^{h} in Indic when the laryngeal was lost – and local loanwords with aspirates were obtained (Kobayashi 2017: 332). In Iranian, the fricatives are easily explained as the regular preconsonantal reflex. Additionally, the weak stem of OAv. *dadā*- ‘put’, YAv. *daδā*- is OAv. *dad*-, but YAv. *daθ*- which seems to indicate that “Proto-Avestan” did not have */θ/ or */t^h/ yet, but actually */d^{h}H/ which could cause devoicing (Kümmel 2018a: 164–5; 2022: 263).

Some scholars reconstruct the voiceless aspirates beyond Proto-Indo-Irania. By the logic that the thematic vowel does not have a zero-grade, the derivational chain **ret-áh*$_2$- > **ret-h*$_2$*-ó*- is impossible, Rasmussen (1989b) reconstructs PIE **t*^{h} to account for the the /θ/ of Irish *reth* corresponding to Indic /t^h/. Most other scholars now agree that many of these arose through contact with laryngeals.

This gives rise to the question *if* the phonemisation could be shared between Indo-Iranian and Greek (despite the solution preferred above). There are a few cases of aspirates seemingly shared with other branches. The connection with Greek has been treated and disproven in recent years (De Decker 2011b; 2015; Norbruis 2023). As stated above, at least one laryngeal must have been present in Proto-Indo-Iranian, and the immediate surface connection does not warrant a reconstruction of a fourth series for the Proto-Language (*pace* Olsen 1999: xli, n. 6):

> Armenian and Slavic seem to show x < *kh_2, cf. *$t\hat{k}$-$\acute{a}h_2kh_2/(t)\hat{k}h_2k\acute{a}h_2$- 'branch' > Arm. *cʿax*, CSl. **soxà* (c) = Ved. *śā́khā-*, *Sogd. šāx* (beside MPers. *šāg*), but this does not necessarily presuppose an intermediate stage with aspiration. No other evidence is found in languages without phonological aspiration.
>
> (Kümmel 2022: 251 (notation of *χ changed to *h_2))

Opinions differ on how to understand *sx-/š-* in Armenian. Martirosyan (2013: 104) connects Arm. *sxalem* 'to err, be mistaken; to stumble; to fail, miss' (and (perhaps further palatalised) *šeł* 'slanting, crooked, oblique', *šil* 'squinteyed') with Skt. *skhalati*, Lat. *scelus*, gen. *sceleris* n. 'misdeed, crime', Gr. *σκελλός* 'crook-legged' and reconstructs **skHel-*. Others connect *sxalem* and Skt. *skhalate* (Br.+) with Gr. *σφάλλομαι* 'to fall, stumble' and derive it from *$sg^{wh}al$- (Kobayashi 2017: 332; Ringe 2024: 124). At any rate, the outcome /k^h/ is does not belong to the protolanguage. Word-initially, Greek only shares aspirates with Indo-Iranian when they can be explained by Sieb's law, e.g. *σχίζω*, *chinátti* 'splits, cuts' and *σφαραγεῦντο*, *spū́rjati* 'thunders' (Kobayashi 2017: 332; Ringe 2024: 124).

There is no inconclusive for the series of voiceless aspirates that should have arisen from "preaspiration" or laryngeal metathesis in branches without phonological aspiration either. Indeed, the outcome of the alleged *p^h t^h k^h always merge with existing phonemes anyway, e.g. Arm. *pʿkʿin* 'arrow' < *$(s)p^hih_1k^hih_1no$- but Lat. *spīca*) (Olsen 1999: 773–4). However, I believe that it must be true for at least Italic that the sequences *h_1T and *h_2T merged with **Đ*, the outcome of *D^h - the prime example being Latin *stabulum* < **staðlom* < "*$stat^hlom$" < *$stah_2tlo$- (Olsen 1988). There are great Greek examples as well (e.g. *γένεθλον* < *$\hat{g}enh_1tlom$), but also plenty of counterexamples (*νῆτρον* < *neh_1trom).

Word-internally, the reconstruction of a common series *p^h t^h k^h $(\hat{k}^h)$ k^{wh} is made difficult by the fact that Greek offers plenty of counterevidence. For the roots quoted above, Greek does not show aspiration: IIr. **pr̥tHu-* corresponds to Gr. *πλατύς*, **pant-/pat*^{h}- to *πόντος* and *$d^hugHtar$- to *θυγάτηρ*. There is further *μέγα* 'big' < *$me\hat{g}\mathring{h}_2$ > **majH̥* > Ved. *mahí*, OAv. *maz-* (Lubotsky 2018: 1882). The verb *$steh_2$- 'stand' (Ved. *tíṣṭhati*, Av. *hišta-* < **s(tH)i-stH-*) is aspirated in the present *ἵστημι* < *$s(t)i$-$steh_2$-ti – which if explained by laryngeal aspiration should be analogical from the weak stem (Weiss 2020: 50 n. 11) – it is not the synchronic weak stem which is *ἵστα-* (PRES.3PL. *ἵστᾱσιν*, IMPF.OPT.3.SG *ἱσταίη*, nor is it generalised to the aorist *ἔστη*

< *h_1e-steh$_2$*- nor to the verbal adjective *στατός* (neither **σθατός nor **σταθός) < *sth$_2$tos*, cf. Ved. *sthitá*- (Ringe 2024: 111 (but to *stah$_3$-*)). The only case of aspiration of this root would be placenames in *Ὀρεσθ*- which is hardly solid evidence (De Decker 2011b: 94–6; *pace* Kümmel 2022: 250)

In the grammatical morphemes where we would hope to find regular correspondences, we find the superlative and ordinal suffix *-*istHo*- as *-*ištHa*- in Indo-Iranian (Ved. -*iṣṭha*-, Av. -*išta*-). It is sometimes assumed that Greek did undergo this aspiration but analogically levelled it (Hoffmann & Forssman 1996: 80). While this is possible, the positive evidence for Greek ever having had *th_2 > *t^h is almost non-existent (Ringe 2024: 90; Norbruis 2023: 228).

De Decker (2011b) found only four non-circular examples of t^h corresponding to *th_2: *οἶσθα, κόγχος, μόθος* and possible *σθήν* "although in all cases other explanations are possible and some leveling is needed". De Decker (2015) in a very thorough evaluation, found mostly counterevidence (especially 5 secure cases of *th_2 > Gr. **t*, IIr. *t^h), but – to him – most importantly 3 instances where Gr. t^h corresponds to IIr. t^h but cannot go back to a laryngeal, namely **mat^h*- 'rob', **ment^h*- 'agitate' and the 2.sg.middle ending *-*t^{h}ēs*.

I believe the matter has now been settled by Stefan Norbruis (2023). He argues that the Greek voiceless aspirates that seemingly correspond to these Indo-Iranian ones do not have the same origin, but that they came about in the clusters PIE **psd, dst, dsk* that first became regressively assimilated to **bzd tst tsk* and then – just like *-*TsR*- > **T^{h}R* later – became an aspirates clusters **tst*, **tsk*, **ksk* > **t^{h}st^h*, **t^{h}sk^h*, **k^{h}sk^h*, but that a root-final velar was usually restored, as was productive suffixes. Accordingly, the *t^h of *οἶσθα* is to be explained as the regular reflex of **u̯oit^{h}st^{h}a* < **u̯oitsta*- < **u̯oid-th$_2$a*. Unlike the present active 2pl of the same verb *ἴστε* (not as expected by Norbruis **ἴσθε) < **u̯it^{h}st^{h}e* < **u̯itste* < **u̯id-th$_1$e*, there was no "normal-looking" ending to regularise, the 2pl. having been replaced with -*ας* on the model of the sigmatic aorist in all other verbs (Ringe 2024: 231).

4.6. Thorn or metathesis

It has also been suggested that Indo-Iranian and Greek belong to the "core" of the Indo-European languages that show a metathesis of dental-velar clusters **TK*>**KT* (Ringe 2010). In recent years, this has come under fire. The Brugmannian extra phonemes *þ* and *ð* have long been abandoned, and with the discovery that Hittie and Tocharian do not metathesise these clusters, they have come under scrutiny (Schindler 1977): PIE **dʰéĝʰ-ōm* > Hitt. *tēkan*, gen. *taknaš*; TB *keṃ*, TA *tkaṃ* < **tkænə*.

The analysis of Cuneiform Luwian *inzagan* as /īnt͡sgan/ < **en-dzgom* or **en-dʲgan* < *$h_1$en dʰĝʰṓm* has furthered the understanding and shown that the process is instead closer to the dental clusters than to a full metathesis (Melchert 2003; Jasanoff 2018). If there ever was this core-Indo-European metathesis should also be questioned (Kloekhorst 2014).

The classic evidence for the metathesis is Indo-Iranian, where all thorn-clusters merge in Vedic -*kṣ*-, but are kept distinct in Iranian: **tk̂* > *š*, **tk(ʷ)* > *xš* (palatalises to *š*) and similarly the voiced aspirated pair: **dʰĝʰ* > Ved. *kṣ*, Av. *z*; **dʰg(ʷ)ʰ* > *γž* (palatalises to *ž*).[21] It certainly seems metathesised from the PIE input; but its is worth remembering, with Lipp (2009a; 2009b), that many consonants merge in preconstantal position. Just like **s(u̯)ek̂s* became PIIr. **s(u̯)aćš* > **saćṣ* > **saṭṣ* > *ṣaṭ* 'six', so did the *-*tk̂*- > *-*tć*- > *-*tṣ*- > *-*ṭṣ*- > *-*kṣ*- (also Sihler 1995: 225). Kloekhorst's formulation, the outcome of **tk* and **tkʷ* are unknown, since he follows Lipp (2009b: 299–300) in rejecting the connection of Gr. *κτάομαι* 'acquire, win', Ved. *kṣáyati*, Av. *xšaiia*- 'to rule'. Instead, he argues that the *TK* clusters were preserved prevocalically: **tk̂ei̯*- > **tćai̯* > Ved. *kṣay-*, Av. *šaē*- 'to live', but was reduced to **K* preconsonantally. This does somewhat fits Ved. *kṣám*- < **dʰĝʰem*-, but *jmaḥ* < **(dʰ)ĝʰmós* (and Gr. *χαμαί*). Kloekhorst seems to not have made his mind up whether *zā̊* or ***žā̊* would be expected as the outcome of **dʰĝʰ*- (2014: 42, n. 16) or **dʰĝ* (2014: 61) – the latter would not fit Gr. *χθών*.

In all other branches, the cluster is reduced and the "metathesis" is not visible. In Balto-Slavic, only the velar survives which would be an odd cluster reduction. Celtic is the only branch in which there is a metathesis specific to "thorn": Cisalpine Gaulish *TeuoχTonion* < **dei̯u̯o-gdon-i̯o*- '(of) Gods and men?' from **dʰĝʰom*-. Otherwise, only the dental survives.

Obviously, the Greek outcome is also metathesised, but this is probably not of PIE date given that the so-called *τίκτω*-rule is synchronically active: **tek̂*- (aor. *τεκ*-) 'produce' gives *τέκνον* 'child', but the present stem **ti-tk-ō* yields *τίκτω* not ***τίτκω*.

---

[21] It is a common feature of PIIr. that the thorn-cluster **tk̂* merges with **k̂s* in PIIr. **ćš* > Ved. *kṣ*, Av. *š*, Nur. **ċ* (Kümmel 2022: 254; Lubotsky 2018: 1885).

Accordingly, *_tétk̂on-_ 'craftsman, carpenter' > _τέκτων_, Ved. _tákṣan-_, Av. _tašan_ and *_tV-tk̂-e/o-_ > _τίκτω_, Ved. _takṣati_ are to be explained by the same rules.

4.7. The fricative /s/ and its variants

The fricative *_s_ must be reconstructed for Proto-Indo-Iranian – and the latest common ancestor shared with Greek. Word-initially, it turns to _h-_ in Iranian and Greek, but it is kept in a number of consonant clusters in both branches. *_s_ is preserved in Indic and in Nuristani (where it may palatalise to _š_ before a front vowel (Nelson 1986: 94).

4.7.1. Ruki

The special and presumably not unitary developments of *_s_, that is especially the Iranian and Greek lenition to *_h_ and the Indo-Iranian (as well as at least Balto-Slavic) retraction to *_š_ (_š, ṣ, x_) following *_r r̥ (l l̥) u u̯ k̂ k kʷ ĝ g gʷ ĝʰ gʰ gʷʰ i i̯_ must obviously be later than the postulated proto-language.[22]

However, the details are not identical. In Balto-Slavic, _ruki_ is not triggered by *_l_ (because it does not become **_r_) and the palatals (Olander 2015: 53–4), but it is by the former labiovelars and devoiced *_g(ʷ(ʰ))_ (Daniels 2017: 1430–1). Examples after *_k_ are uncertain (Kim 2018b: 1649). It only becomes phonemic when *_k̂_ merged with _ruki-š_ in Baltic and non-Ruki-_s_ in Slavic (Andersen 1968).

Although disguised by subsequent innovations, _ruki_ did operate in Nuristani (Cathcart 2011; Hegedűs 2012).

As argued in Chapter 1, it is possible to view this as a potentially shared innovation between Balto-Slavic and Indo-Iranian, which then subsequently gained (more)

---

[22] In Indic, the reflex of *_k̂t_ > _*ćt_ merges with the _ruki_-outcome of *_st_ > _*št_ > Ved. _ṣṭ_, but Iranian has *_ćt_ > _xšt, xš_ but *_št_ > _št_ (Lubotsky 2018: 1884). Nuristani seems to agree with Indic in merging the two as PN *_št_, but this is obscured by *_st_ > _št_ in non-ruki environments in Kati, but depalatalization of *_št_ in Ashkun and Prasun (Cathcart 2011: 4–5): *_ĵʰasta-_ 'hand' > Ashkun _dost_, Prasun _lušt_. PIIr. *_aćta-_ 'eight' > PN *_aṣṭa/ašta_ has unexpected retroflex reflexes. Examples of *_k̂t_ are very rare (Cathcart 2011: 5, n. 2; Strand 2023: 800). The opposing view, that they did not merge in Nuristani is held by Ėdel'man (Ėdel'man 1999: 98)

productivity in Indo-Iranian. It is clear that Greek does not take part in this innovation.

4.8. Assimilation and dissimilation: Bartholomae's and Grassmann's Laws and the allophones of */s/

For Proto-Indo-European itself, **s* is regularly noted as **z* when regressively voiced – but most often accompanied by square brackets or an explanation of the subphonemic status, e.g. (Fortson 2004: 60; Fritz & Meier-Brügger 2021: 85, 112). But other automatic assimilations are rarely quoted as such: **sed-to-* 'sat', not ***setto-*, the *to*-participle of the root **sed-* 'sit'.

However, in clusters that include voiced aspirates at least Indic and Iranian (before they merged with the regular voiced stops), we can observe a progressive voicing assimilation. To some, it is even older than Proto-Indo-Iranian, and it may thus have been present in the latest common ancestor:

> The sibilant fricative ∗/s/, which was underlyingly voiceless, seems to have been voiced to ∗[z] before voiced stops (e.g. in ∗nisdós 'seat, lair, nest'); it probably also had a breathy-voiced allophone before breathy-voiced stops (e.g. in ∗misd^{h}ó- 'reward')
>
> (Ringe 2024: 17)

4.8.1. The age of Bartholomae's Law

It is difficult to assess the age and the relative chronology of Barthomolae's Law, its domain in Proto-Indo-Iranian, the devoicing of **z* in Indic and the relative age of Grasmann's Law to Bartholomae's.

To some researchers, e.g. Kümmel (2022: 247–8), it is easier to understand Bartholomae's Law as an archaism of PIE date. Examples like PGmc. **huzda-* 'treassure' if from **kud^{h}-to-* 'hidden' (with Bartholomaes Law and assibilation of the resulting voiced dental cluster: **kud^{z}do-*, to Gr. *κεύθω* 'to hide') (Kroonen 2013: 260) could also point to the PIE age of the rule. However, most examples outside Indo-Iranian are disputed, ambiguous or unconvincing (Sihler 1995: 200–1; Beekes 2011: 130; Ringe 2024: 112–4). Even researchers who adopt the rule for Proto-Indo-European do not note its application, i.e. **kud^{h}to-*, not ***kudd^{h}o*.

Ringe (2024: 123) sees _χθες_ 'yesterday' < _*ĝ^h-di̯es-_ with "deictic _ĝ^h_" and _ἐχθρός_ 'hatred, hostile' < _*eĝstr̥os_ as examples of Bartholomae in Greek, but I would rather explain these as examples of Norbruis' (Norbruis 2023) cluster affrication rule.

The fact that the Iranian merger of the mediae and the mediae aspiratae happened _after_ Barthomolae's law gives a single, but very vague, clue as to when it happened, before the split between Indic and Iranian. However, the seemingly different domain in Iranian than in Indic raises some questions on the complete unity of the rule. Similarly, we cannot tell whether or not Grassmann's law ever applied in pre-proto-Iranian, which is relevant for its dating and role in the so-called aspiration throwback in Indic.

4.8.2. The age of Grassmann's Law

It is more difficult to assess the relative chronology in Indic. On the surface, forms like _dipsa-_ point to the fact that Grassmann's Law applied in such cases, meaning that the aspiration on the second consonant cluster must have been preserved at the time this law operated. Because of the Iranian merger of the voiced aspirates with the voiced series of consonants, we cannot tell directly if this was already Proto-Indo-Iranian. It might have been, if we accept circumstantial evidence, such as the line of argumentation tentatively suggested by Kümmel (2022: 247–8), namely that the potentially joint Indic and Iranian replacement of the expected velar with its palatalised counterpart in the 2sg imptv. _*ȷ́ad^{h}í_ 'slay!' > Ved. _jahí_, Av. _ja^{i}δi_, if this was done to avoid homophony with _*gad^{h}í_ 'come!' > Ved. _gahí_, Av. _ga^{i}dī_, after Grassmann's Law made _*g^{h}ad^{h}í_ 'slay!' and _*gad^{h}í_ 'come!' merge in the phonetic form of the latter. This is, however, far from certain and could very well be later independent paradigmatic levelling of the initial consonant. Conversely, it has been suggested that the Gleichung of Ved. _kúmbha-_ 'jar, pitcher' and Av. _xumba_ from < _*k^{h}umb^{h}a-_ seem to indicate that Grassmann's Law did not apply in Proto-Indo-Iranian. However, the reconstruction of this ultimate loan word is far from certain (Kümmel 2022: 248).

According to Kobayashi (2004), Vedic forms of the Grassmannian type like _dipsa-_ are relics from a time when Barthomae's law applied, and the aspiration was preserved at the end of the consonant cluster: _*d^{h}ibz^{h}a-_. He goes on to claim that the subsequent Vedic phenomenon of "aspiration throwback", which synchronically means that the aspiration appears on the – from an Indic perspective

– unexpected root-initial consonant. *dʰipsa-* from the root *dabʰ-*, proves the reality of the underlying form **dibzʰa-*. This form would then surface as *dʰipsa-* because of the loss of the voiced allophone **z* in Vedic. As this either merged with [s] or was lost (with compensatory lengthening or triggering retroflexion), and as the phonotactics forbid ***bʰs* which neutralises to *ps*, the only place for the aspiration would be on the root-initial consonant.

I find this explanation problematic for several reasons. Most importantly is the existence of the "older" forms of the *dipsa*-type. Kobayashi does not explain them as analogical after other forms of the verb, in which Grassmann's Law applied and the root-final consonant cluster was devoiced; instead, he implies that they are original, and that the *dʰipsa*-type is the *later* and regular outcome, meaning that the speakers actively created formations with an underlying knowledge of the voiced and aspirated character of the *ps* as *//bzʰ//. This is hardly credible outside a generativist mind; at least it leaves to be explained why the inherited forms fell out of fashion and were clumsily replaced by the productive pattern.

In my opinion, it would be worth considering the obvious parallel, namely the aspect stems of the verb *ἔχω* 'to hold, have' in Greek. This is the usual textbook example of how the Greek Grassmann's Law also applies to *h-* from **s-*, and how it can be blocked or bled by an intervening **s*: *ἔχω* in the present stems continues **hekʰō* < **seĝʰ-oH* where the law applied regularly, and in the aorist, *s* is retained before the consonant and thus there is no ground for the application of the law: *ἔσχον*. In the future stem, however, the *s*-initial morpheme neutralised the aspirate **seĝʰ-s-* > **seks-*, which in turn also blocked the g of Grassmann's law in **seks-* > *heks-* > *ἕξω*.

A very similar relative chronology could apply to the interplay of Bartholomae's and Grassmann's laws in Indic, but only if Kobayashi is *not* right in the assumption that the *dipsa*-type are relic forms, whereas *dhipsa-* is the productive formations. We know from cases where Bartholomae's law was trigged by stops and not **s*, that it applied first; or at least that it did not have an influence on Grassmann's Law: **bʰudʰto-* > *buddha-*. However, the only evidence for *aspiration* being a part of the "extended" Bartholomae's law can come from these cases of aspiration throwback. Sure, In Iranian, the voicing assimilation is progressive unlike in regular consonant clusters, but that can neither prove the preservation of aspiration in the new Bartholomae-cluster nor that this potential aspiration was released on or after the

**z*. Should this chronology be correct, the forms of the *dipsa*-type would be analogical after the many forms in the paradigm where the root *dabh* synchronically appeared in a "Grassmannian" zero grade.

Conversely, Lipp (2009a: 252) argues that the *dhipsa*-type is analogical, essentially adopting the same chronology as Kobayashi, although this solution does raise the obvious question of what the analogical model is.

Ultimately, it is necessary to recall, as Lubotsky (2018: 1879) reminds us, that whatever its nature and age was, Bartholomae's law remained subphonemic since the realisation of the cluster was automatic and did not contrast with anything else.

Some scholars have tried to connect (Kiparsky 1973; Pozza 2019) Grassmann's law in Indo-Iranian and Greek, but the relative chronologies make it difficult. Even in Iranian and Greek, where initial **s*- turned to *h*-, the rule is not identical; *h* is deleted in Greek, but kept in Iranian: *hāxa*- < **sākH*-. A further parallel is found in Tokcharian (Kümmel 2022: 248; Peyrot 2022).

4.8.3. Bartholomae's /s/

The usual notation for the Proto-(Indo-)Iranian reflex of Proto-Indo-European **s* in a "Bartholomaeic" context is **z^h*, e.g. (Sihler 1995: 201; Beekes 2011: 130; Lipp 2009a: 252; Kobayashi 2017: 334; Cantera 2017: 490; Lubotsky 2018: 1879). When comparing forms like Av. *diβža*- to Ved. *dipsa*-, desideratives of the root **d^{h}ab^h*- < **d^{h}eb^h*- 'deceive' (Cheung 2007: 42–3), we run unto all of these problems. A transposition to PIE, or at least pre-Barthomolaeic Indo-Iranian, automatically projecting the root in the zero-grade and the desiderative morpheme **-sa*- back to just after the vowel merger, would look like **d^{h}ib^{h}sa*-. In Iranian, this /s/ obviously underwent some form of Bartholomae's Law and became first voiced and ultimately palatalised: **d^{h}ib^{h}za*- > **dibža*-. In the usual, automatic notation, this is quoted PIIr. **d^{h}ibz^h*-, partly due to the developments in Indic.

This is certainly an automatic notation, and it does not necessarily reflect the conscious reconstruction by researchers who employ it. However, as I will outline below, it is most likely impossible, at least rests on hitherto unattested phonemic combinations of any human language and is therefore highly unlikely (Jacques 2011: 1520). Further, I believe that the notation has led some researchers astray in that they based arguments on it, assuming that it is indeed linguistically real. Note that the following critique does not concern $*\acute{z}^h$ as the Proto-Indo-Iranian outcome of $*\hat{g}^h$ or $*\check{z}^h$ as the palatalised outcome of $*g^w$ and $*g^{wh}$. While **[ʑʱ] and **[ʒʱ] would be equally problematic, $*\acute{z}^h$ and $*\check{z}^h$ are clearly just a cover symbols for uncertain outcomes of the PIE stops. It is certain that they were proper affricates in Proto-Indo-Iranian, at the very least evident by the preserved occlusion in Nuristani (Lipp 2009a: 156).

Automatic notations are no issue per se, and we have plenty of examples in Indo-European linguistics. We could include the notation of the laryngeals with the cover symbols $*h_1\ h_2\ h_3$, or – for some researchers – the consonants $*t\ d\ d^h$ kept as such for convenience, although adherents of the glottalic theory would reconstruct their phonetics rather differently.

4.8.4. /s/ and its allophones

There is nothing wrong with Barthomolae's law affecting a cluster with *s* and even it retaining aspiration in (pre-)Proto-Indo-Iranian, but there is something very suspicious about the assumption that the aspiration was realised on the sibilant part of the cluster. The issue is not that the phoneme */s/ would have 6 allophones, rather that at least two of them do not exist in any attested language:

1. **s* – the unmarked variant, e.g. word initially and after unvoiced consonants
2. **z* – the regressively voiced variant, e.g. before voiced stops
3. **š* – the unvoiced (default) RUKI-variant, e.g. after **r l u u̯ i i̯ k* k^w $\hat{k}$ and before a vowel
4. **ž* – the regressively voiced RUKI-variant, e.g. following **i u* and preceding a voiced stop
5. $*z^h$ – the progressively voiced Barthomolaeic variant, e.g. following a voiced aspirate
6. $*\check{z}^h$ – the progressively voiced Barthomolaeic RUKI-variant, e.g. following a voiced velar aspirate

While most of these are quite observable, $*z^h$ and $*\check{z}^h$ are very dubious. Aspirated sibilants are extremely rare in the languages of the world, and they most often co-occur with aspirated affricates. As such, it would not be typologically strange to find a cluster like $**ts^h$. However, the reconstructed voiced variant is concerning: First, it is strange for $*dz^h$ to occur without $**ts^h$, and second is not at all restricted to dental affricates, but $*z^h$ is reconstructed for all Bartholomaeic clusters ending in **s*.

Voiced aspirated sibilants are only apparently only reported in a single language variety in the world (Jacques 2011), namely the dialect of Dikundu !Xũ.[23] To make matters more complicated, Jacques's survey quotes Köhler 1981 for the occurrence which is also less than ideal: First, it is only a preliminary ("provisoire") analysis (Köhler 1981: 563); second, it seems that none of the other varieties of !Xũ or the Khoisan Languages attests such voiced aspirated sibilants (Sands et al. 2013; esp. Miller 2013: 47–50; Heine & König 2015: 40–1); and third, none of his colleagues seem to have found speakers of this variety for further study (Heine & König 2015: 18, n. 4; Heine & Honken 2010: 8, n. 5). In total, it is rather unlikely that the sounds *zʰ* and *žʰ* ever existed as phonemes outside the notation of linguists.

4.9. In conclusion: Conservative PIE

Sadly, this tedious endeavour has revealed very little. Indo-iranian and Greek cannot be traced back to anything other than a conservative (one daresay Indo-Greek) reconstruction of Proto-Indo-European. All stops (including 9 velars), laryngeals, syllabic and consonantal liquids and nasals were preserved. It is impossible to point to exclusive common innovations, though **a* and **b* had probably become phonemes – if they were not already.

The nasal **m̥* was just as sonoric as it ever was (*pace* Zair 2018). There is no need to posit that postvocalic laryngeals had caused compensatory lengthening since postconsontal laryngeals need to have been preserved (*pace* Trager & Smith 1950). Accordingly, there is no evidence to reconstruct a series of voiceless aspirates, the Indic series stemming from **TH* (corresponding to Iranian fricatives and Greek voiceless stops), and the Greek ones arising in clusters of *-TsT-* (Norbruis). Grasmann's law was probably not active (*pace* Kiparsky 1973). Ruki and Bartholomae's law *could* have been subphonemically active, but if the latter was and influenced clusters with **s*, the aspiration was realised as breathy voice of the stops, not aspiration of **z*.

[23] The variety spoken around the village of Dikundu (western Caprivi Strip, Northeastern Namibia) of the Žu|'õasi (Ju|'oasi) or the southeastern dialect of !Xũ (!Kung or !xõõ) – a member of the Khoisan (Khoesaan) language family. Žu|'õasi is referred to as "E1" by specialists, and the Dikundu variant described by Köhler as "E2" (Heine & Honken 2010: 8; Heine & König 2015: 18; König 2008: 999).

5. Synopsis of morphological isoglosses of Greek and Indo-Iranian

5.1. Introduction

Unfortunately, this chapter will only be brief synopsis of the morphology of the latest shared ancestor of Indo-Iranian and Greek. As shown in Article 1, the nature of their relationship is very often only described in vague terms, and it is rarely substantiated by morphological innovations, let alone exclusive morphological isoglosses. The previous chapter reveals that even a tedious bottom-up reconstruction does not offer much more than an introductory grammar of Proto-Indo-European. Therefore, I will focus on the few proposed potentially shared innovations and spend less time on the isoglosses that are probably archaisms because they are shared with other branches.

5.1.1. Nominals

The ”Indo-Greek” noun must have been very conservative and copious. All cases reconstructed for Proto-Indo-European (except the Anatolian allative) are attested in Vedic and Avestan (vocative, nominative, accusative, genitive, dative, locative, ablative and instrumental) in the three numbers (singular, dual (turning to a numerative in Middle East Iranian (Sims-Williams 1989: 183)) and plural) – and in three genders (Kümmel 2018b: 1892). Although there are plenty of case synchretisms in Greek, the system is derivable from the Indo-Iranian one (on which the PIE is reconstructed) (van Beek 2022a: 179). Mycenean preserved the dative and locative distinct and perhaps the ablative and genitive distinct in the singular, but otherwise these merged by Proto-Greek (García Ramón 2017: 654).

Plenty of accent and ablaut patterns across multiple stem-classes can be reconstructed (Ringe 2024: 55–8; García Ramón 2017: 657–63; Kümmel 2018b: 1890–7), including root nouns, thematic nouns and stems in **ah₂/ā, *ih₂, *uh₂, *-s-, *-r-, *-n-, *-nt-* and **-r/n-* (Euler 1979).

5.1.2. Oblique plural: dat.-abl.pl. **-m/bʰos* instr. **-bʰi(s)*

Indo-Iranian and Greek both belong to the “core”-IE languages in the sense they they seem to have innovated a separate feminine gender, but it is debated if this was lost in Anatolian or acquired later (Ringe 2024: 21–3; Lundquist & Yates 2018: 2094–2100). They also share the widespread gen.sg. o-stem ending **-osi̯o* not attested (directly) in Anatolian or Tocharian (which as *-*nsə*). However, it cannot be ruled out that it was not replaced by the athematic ending in these languages. *-*osi̯o* might have been pronominal in origin, but it is (also or exclusively) nominal in Greek, Indo-Iranian, Albanian, Armenian, Italic and Celtic (Ringe 2024: 54; Lundquist & Yates 2018: 2087–8; Olander 2015: 136–8)

In the oblique plural cases, Indo-Iranian and Greek to the extend it is visible belong the languages that – unlike Anatolian[24] – show a labial consonant in the ending, and specifically *-*bʰ*-, e.g. Myk. -*pi*, Hom. -*φι* (although it is not quite a case), Ved. dat.pl. -*bhyas*, instr.pl. -*bhiṣ*. They share this trait with Italic and Celtic. On the other hand, Germanic, Baltic and Slavic show *-*m*- in the dative (and instrumental) plural. It is by no means exclusive, Greek does not preserve the dative which is replaces by the old locative plural. The **m/bʰ*-distinction is probably the result of generalisations of the consonants of opposite original endings (Jasanoff 2009; Melchert & Oettinger 2009). Indeed, a conflation of *-*bʰi* with *-*m-os* would be probable (Gotō 2013: 11), but unfortunately this reconstructed distribution is not attested anywhere:

> Because *-bʰ- is most clearly at home in the PIE instrumental plural ending, and *-*m*- cannot have arisen out of thin air, it is likely that the Germanic and Balto-Slavic dative plural endings are archaic
>
> (Pronk 2022: 280)

Kümmel (2022: 261) suggests that it is an areal trait of "southern" core-IE languages and cannot be a true genetic innovation since the differences vary greatly. From a purely linguistic perspective, this is at least compatible with the material, but cf. Chapter 6.

5.1.3. Oblique dual: gen.du. –(*oi̯*-)*Hos*, loc.du. *(-*oi̯*)-*Hau̯*

Iranian is the only branch of Indo-European to show distinct endings in the locative and genitive dual (Kümmel 2018b: 1893): YAv. *narå̊* 'of both men' going back to PIIr. gen.du. *-*Hās*, and OAv. *aŋhuuō* 'in both lives' going back to PIIr. loc.du. *-*Hau̯* which seems to be an Iranian archaism, since the corresponding Ved. -*oḥ* looks

[24] **bʰ* is unattested in Tocharian, unless it hides in *ṣ/ṣp/ṣäp* 'and' < **se-bʰi* (Ringe 2024: 52), although this is not generally accepted (Peyrot 2022: 88; Adams 2013: 731).

like a conflation of the two: PIIr. *-*Hau̯-š* (Mayrhofer 1986: 18). There is positive evidence for the ending-initial laryngeal (it makes matric position in both branches) (Kümmel 2018b: 1893 with lit.), and from the Verschärfung of the genitive of the numeral 'two' in Germanic: **du̯oi̯-Hou̯* > **twajjō* > Got. *twaddjē* (Weiss 2020: 227). However, the genitive -*ā̊* is on its way to oust the locative already in Old Avestan (Hoffmann & Forssman 1996: 115). In the *o*-stems, Indo-Iranian extends the *-*o*- to the *-*oi̯*- of the oblique *plural* stem, usually considered pronominal in origin (Sihler 1995: 265; Olander 2015: 212): PIIr. *-*ai̯-Has,* *-*ai̯-Hau̯* > YAv. *vīraiiā̊* 'of both men', OP *gaušāyā^h* 'of both ears'; OAv. *zastaiiō* 'in both hands'; Ved. *hastáyoḥ* 'in/of both hands' (Skjærvø 2007: 893; Euler 2010: 89). Greek only has one ending gen-dat.du -*οιν*, Hom. also -*οιιν* which "has the general look of something originally proper to the o-stems" (Sihler 1995: 265). The archaic Arcadian form *Διδυμοιυν* 'of both twins' which Euler (Euler 2010: 89) sees as evidence for *-*oi̯-u* comparable to Ved. -*ayoḥ* and reconstructed as proto-"Ostindogermanisch" *-*oi̯ou*. The vocalism of this claim is difficult, as is the survival of *-*i̯*- in Greek (Sihler 1995: 265). Weiss (2020: 227) suggests that *Διδυμοιυν* directly continues the loc.pl. *-*oi̯su* with a final *-*n* comparable to the nasalisation of the Old Irish dative dual. Instead, he suggests that Arc. *μεσουν* 'between' is an old dual, and that this form is more easily connected with the Indo-Iranian and Balto-Slavic forms which implies a preform **med^hi̯oHou̯-m*.

Slavic has merged the gen.-loc.du, and the ending (OCS -*u*, ORuss. -*u*) seems to reflect the *s*-less loc.du. *-*au̯* of Iranian < PIE *-*Hou̯*- (Olander 2015: 205–6). Thus, Arc. *μεσουν* and Slavic -*u* (< *-*Hau̯*) disagree with (Indo-)Iranian *-*ai̯Hau* on the introduction of the element *-*oi*- into the obl.du. In the numeral 'two' and in the pronoun, Slavic and Iranian do, however, resemble each other: YAv. *duuaiiā̊*, YAv. *uuaiiā̊* 'of both', OAv. *ubōiiō* 'in both', YAv. -*uuaiiō* 'id.); OCS *dъvoju, toju*.

Is the introduction of the plural pronominal marker *-*oi̯*- into the *o*-stem genitive dual a common Innovation of Indo-Iranian and Greek to the exclusion of Balto-Slavic? Hardly.

5.1.4. Comparison

In Indo-Iranian and Greek, *-*tero*- forms productive comparatives. However, as the corresponding superlative suffix is not identical it is not likely to a common innovation (Kümmel 2022: 261). On the basis of inherited *-*tm̥Ho*- seen in Lat.

intimus 'innermost', Ved. *ántama* 'nearest' (the regular "secondary" superlative), Greek has innovated **-tm̥-tos-* > *-τατος* replacing. However, as both languages simultaneously attest the "primary" comparative **-i̯es/-i̯os/-is-* (in Greek (as in Germanic) at some point extended to **-is-on-* > *-ιων* (van Beek 2022a: 180; Beekes 2011: 222), but the suffix is unattested in Mycenaean; in Indo-Iranian continued as **-i̯ās-,-i̯as-,-is-* and also the base of the conglomerate superlative **-ištha-* > Ved. -*iṣṭha-*, Av. *-išta-*), it is difficult to formulate what the innovation actually is (Porzig 1954: 158; Kümmel 2018b: 1898; Gotō 2013: 49; 2017: 352). Additionally, the suffix **-tero-* was present, but was used in a contrastive function. In most branches, only relics formed to adverbs survive, but in Greek and Indo-Iranian the suffix could freely be attached to nominal stems and later form comparatives. (Ringe 2024: 77, 252–4).

5.1.5. Three notes on derivation

This topic has been covered in great detail (Euler 1979), but not in a particularly strict methodological framework – what is lacking is a critical assessment of the relative age of the types and the attestation patterns outside Indo-Iranian and Greek. This should be done in the future. Here, I will only add a few minor details.

In a largely unknown lexicostatistical survey (Bird 1993), there are two derivational suffixes shared uniquely by Indo-Iranian and Greek: **-bʰo* and **-mei̯o-*. However, the data is based on Mann (1984) and the "uniqueness" does not hold. **-bʰo* is found in *śalabhá-* 'locust', *śarabhá-* 'deer' (*śal-* 'jump'), *ἔλαφος* 'deer', *ἔριφος* 'young buck'. The suffix **-bʰo* is not exclusively "Indo-Greek" in a narrow sense, although it is most widespread here. It is most often found in colour-terms and animals names (Olsen 1999: 805; Euler 1979: 181, n. 878; Debrunner 1954: 746). If Hitt. *šalpa-* 'dog excrement' truly belongs here, the suffix is PIE. There is further the semantically different Hitt. $^{\text{TÚG}}$*u̯ašpa-* 'clothing' if from **u̯es-bʰo-* (Kloekhorst 2007: 984). **mei̯os* is rare in Greek, but productive in Indix: *ἀνδομέος* 'human- (used of flesh)'; *go-máya-* 'for cattle', *aśmanmáya-* 'of stone', *ayasmáya* 'of iron or copper' – and, e.g.. *tarasa-maya-* 'made of meat' (*tarasa-* is probably a Dravidian loan) (Mayrhofer 1986: 628–9). It may be an archaism, as it was not widespread in Greek, or it could be an innovation of these two branches (and whomever else lost it) which fell out of favour in Greek (Euler 1979: 88, and n. 340; Debrunner 1954: 768–9).

A minor isogloss which might be a shared innovation is the formation of distributive adverbs in *-*k̂n̥s* or *-*k̂as* (Klingenschmitt 1975), but the isogloss requires quite some extra assumptions. It stems from the explanation of Gr. *ἕκαστος* 'each, everyone' as derived from the adverb *ἑκάς* 'afar, separate, away' which should be derived from **su̯e-* 'oneself' and the suffix in question (Sihler 1995: 401; Robert Stephen Paul Beekes 2010: 395; Gotō 2013: 146). Other than it this word, it occurs only in *ἀνδρακάς* which occurs once in Homer (Od. 13.14) in the meaning 'man for man' in a distributive meaning (each man is supposed to gift a cauldron and a tripod). In later authors, the meaning is closer to 'each'; but it is clear form the scholia that the form was no longer understood in antiquity (Heubeck & Hoekstra 1990: 163; Ameis & Hentze 1884: 1884; West 2017: 272). The connection to Indo-Iranian is the suffix of Skt. *dvi-śás* 'two by two' (Schwyzer 1939: 630). However, the distributive meaning with lower numerals is exclusively classical Sanskrit (Wackernagel & Debrunner 1975: 429–30). The Rig-veda only has *sahasraśás* (8.34.15) 'by the thousands, but also *r̥tuśás* 'at the proper time'. Artharvaveda has *śataśáḥ* 'in hundreds', and in the later language there it is added to the lower numerals and to nouns like in *gaṇaśás* 'by troops'. In Avestan, we find YAv. *navasō* in *navasōs-ča baēvąn* 'and nine times tenthousand' which is not exactly distributive (Bartholomae 1904: 1046). The connection is not exactly semantically forcing. According to Pedersen, it is not a proper suffix, but a compound with a noun from the root *śas-* < **k̂es-* 'cut' (Pedersen 1895: 39; Vaillant 1974: 676). Pedersen also sees the "suffix" in Slavic: ORuss. *mitusъ* or *mitusь*'alternatively'. If that is the case, it not exclusive – although the Slavic etymology is formally difficult.

5.1.6. Nominal forms of the verb

The Greek infinitives in *-εῖν* < **-ehen* < **-es-en* have been reconstructed as an endingless locative of a proterodynamic neural *s*-stem. Mycenaean *e-ke-e* /(h)ekʰehen/ < **hekʰehen* reveals that the suffix contains *-*s*- or *-*w*-. It has been connected to the Vedic infinitives in *-sáni* (e.g. *sakṣáni* < **seĝʰ-sén+i*), but the forms are not identical (Sihler 1995: 608, 610; Ringe 2024: 232). The addition of the extra locatival *-i* on top of an ending less locative is probably recent in Indic, but they go back to otherwise unattested locatives of amphikinetic **ségʰ-ōs, segʰ-sén+i* (Stüber 2000: 134). I believe this difference could be overcome. On the other hand, most other Vedic infinitives are formed directly to the root, whereas the ones in *-sáni* tend to be formed to the present stem: *gr̥ṇīṣáṇi* 'to sing' formed to *gr̥ṇā́ti*. Latin infinitives in *-ere* < **-esi* match Ved. *-áse* in, e.g. *bharáse* and also belong to these *s*-stem nouns derived from verbal roots or stems (Weiss 2020: 474; Stüber 2000: 164; Sihler 1995: 610).

In Bird's survey, there is "future passive formants" in **-teu̯i̯o-*, only found in Indo-Iranian and Greek. Mann (1984: 1393) connects Gr. *δοτέος* with Skt. *dātavyaḥ* (also Birwé 1956: 65–8). The reconstruction **-teu̯i̯o-* is impossible for the Greek. It would fit *φατειόν* which occurs verse-finally in Hesiod (*οὔ τι φατειόν* "not to be said"), and which might me a metrical lengthening of a verbal adjective in *-τεος* (Robert Stephen Paul Beekes 2010: 1567). Euler (1979: 78) also made the connection between this form, the Skt. gerunds in *-tavya-* and, but remarks on the problematic chronology: The verbal adjectives in *-τέος* do not occur before the 5th century BCE, and the Vedic gerunds are only found from the Atharvaveda. On the basis of the connection with Welsh *cara-dwy* 'worthy of love' and the Mycenean "Gerundiva" *ki-ri-te-wi-ja* and *a-mo-te-wi-ja*, Euler further reconstructs **-teu̯i̯o-/*-tou̯i̯o-* for "Gemeinindogermanisch". If the connection were to hold, the isogloss should not have been unique as it is in Bird's table. However, the Mycenaean forms are not gerundives, but adjectives derived from nouns: *ki-ri-te-wi-ja* /krī(s)tʰĕ̄u̯i̯ā/ '(cult-)woman of the barley' from *κρῑθᾱ* 'barley' and *a-mo-te-wi-ja* /(h)ar(h)motēu̯i̯ā/ and *a-mo-te-wi-ja* 'decorated with chariots or wheels' (of a ewer) or 'of the wheel-maker/charioteer' (García Ramón 2016: 243; Bartoněk 2003: 171, 213; Killen & Bennett 2024: 783). The Artharvavedic gerundives are derived from *-tu-* (infinitive) + the Rig-vedic gerundive suffix *-ya* < **-ii̯o/-iHo-* (Gotō 2013: 141).

5.2. Pronouns

In Hajnal's reconstruction, PIE had two relative pronouns: restrictive **kʷ(e)i̯-* and appositional **(H)i̯o-* (Hajnal 1997: 64). Accordingly, "Graeco-Aryan" innovated by generalising **(H)i̯o-* to both functions, as opposed to Anatolian and Italic. Notably, Balto-Slavic must belong to "Graeco-Aryan" here, since Indo-Iranian and Baltic has completely replaced the stem **kʷei̯* with **Hi̯o-*, whereas Greek would be an outlier in this "subgroup" as it preserves relics of the stem **kʷi-* in restrictive relative function (so Hajnal). Evident by the fact that that Celtic innovated in the "Graeco-Aryan" way, and Germanic in the Anatolian way (Jørgensen 2022: 143; Hansen & Kroonen 2022: 166), we can probably assume that both stems were preserved relatively late. While this is perhaps not the ideal reconstruction of PIE, it is worth noticing that Anatolian (Hitt. *kuiš*) and Tocharian (TB *k$_u$se*, TA *kus* < **kʷo*) preserve another stem, the interrogative **m-*. Peyrot (2022: 96) argues that while loss is a weak (non-identifiable) phylogenetic argument, the replacement of a centrally positioned stem is salient, and it could therefore be an innovation of the non-Tocharian core to have lost the interrogative **m-*. However, In Anatolian and Tocharian, *k^w- functions as the interrogative and relative stem. In fact, the forms in **m-* seem to be relics: Hitt. *⸗ma, mān, maši maḫ̮ḫ̮an*, Pal. *mān* 'when'; Lyc. *-mẽ* 'so, likewise'< PA **mon*; TB *mant*, TA *mänt* 'so' (Kloekhorst 2007: 552). The argument does not hold if OIr. *ma, má* 'when' is to be connected – unless, of course, this was lexicalised early on and constitutes no evidence for the interrogative stem (Kloekhorst 2007: 552; Adams 2013: 472–3). Since there is no evidence for the relative stem *(*H)i̯o*) seen in Greek, Indo-Iranian, Balto-Slavic and Celtic makes it impossible to judge if **(H)i̯o* was an original relative lost in Anatolian and Tocharian (Adams 2013: 200), or if was originally a demonstrative pronoun (Sihler 1995: 400). The intimate relationship between interrogative, indefinite, relative and demonstrative pronouns make it very difficult to reconstruct the ancestral state of affairs (see also Ringe 2024: 70).

In the (oblique) stem of the accented non-singular pronouns of the first and second person, Greek and Indo-Iranian exclusively share the suffixes *-*u̯e-* of the dual and *-*me-* of the plural (Kümmel 2022: 259–60).

However, it is impossible to see if this is an archaism as the accented forms are lost in most branches. It could be considered an innovation of Greek, Indo-Iranian (and any other branch not preserving the accented forms), if they extended an original distinction in person or exclusiveness to number (Kümmel 2022: 259–60). The matter is complicated further by Celtic 2pl **swīs* < **us-u̯e-s* and Hitt. *anza-* < **n̥s-u̯é-* with "dual" *-*u̯é* in the plural.

	1du	2du	1pl	2pl
PIE	**nh_3-u̯é-*	**uH-u̯é-*	**n̥s-mé-*	**us-mé-*
Gr. (acc.)	**nōu̯-* > *νώ*, *νῶϊ, νῶε*[25]	–	**hãhmé*[26] > Lesb. *ἄμμε*, Dor. *ἁ̄μέ*, Hom. *ἡμέ-ας*[27]	**huhmé* > Lesb. *ὔμμε*, Dor. *ὑ̄μέ*, Hom. *ὑμέ-ας*

[25] *-*i* is unclear, -*e* is the productive dual ending (Ringe 2024: 259; Sihler 1995: 381)

[26] With "hauchumsprung" (Ringe 2024: 259; Sayeed 2019)

[27] recharacterised with the acc.

PIIr.	*āu̯a-Hám	*i̯uu̯a-Hám[28]	*asma-Hám[29]	*ušma-Ham
Ved. (acc.)	*āvā́m*	*yuvā́m*	*asmā́n*	*yuṣmā́n*[30]
Iran.[31]	–	YAv. gen. *yᵃuuākəm*	OAv. acc. *əh-mā*, YAv. *ah-ma*; OP gen. *amāxam*	YAv. gen. *yū̆šmākəm*,

5.3. The verb

5.3.1. "Irrelevant archaisms"

The "Indo-Greek" verb has been the centre of a lot of attention. Much of this is treated in Article 1 (though in terms of terminology, not reconstruction). It is, however, clear that the verb of Indo-Iranian and Greek, copious as it is, seems to continue more archaisms than innovations. At least, it shares most of its features with other branches, although Greek and Indo-Iranian continue the most. There is also a growing consensus that the Tocharian verb is also more compatible with the "Indo-Greek" than Anatolian state of affairs (Peyrot 2022; Friis 2024; 2021; Malzahn 2010). In this section, I will only address a few shared traits mentioned as potential innovations or isoglosses in the literature. More work is definitely needed on the stem formation – both formally and functionally.

Many verbal isoglosses indeed connect "Indo-Greek" more to other core-languages, but these are irrelevant for the relationship between Indo-Iranian and Greek. There is no doubt that their latest common ancestor had simple, reduplicated sigmatic and thematic aorists (Birwé 1956: 26–32). The age must instead be assessed from other branches (Toch.: Friis 2024: Ch. 3: them. aor., chaps. 4-5: s-aor.; Anat. s-"aor".: Søborg 2020; Bendahman 1993). The same goes for the present classes where we find root-, reduplicated, simple thematic, *i̯e/o-*, nasal-infix and nasal-suffix (**-neu̯* in **deḱ-neu̯-* > Hom. *δεικνύμενος*, Ved. *dāśnóti,*[32] **-nah₂* in **dʰgʷʰi-náh₂-* Gr. *φθίνω*,

[28] with **uHu̯á* >> **yuu̯á-* after 2pl. **i̯uH-*.

[29] Remarked with the particle **-(H)am* (Kümmel 2018b: 1904) or the acc. ending (Gotō 2013: 66–7)

[30] with *y-* from the nom. **i̯uH-* > **i̯ū-* > *yū-yam*

[31] (Skjærvø 2007: 524–5)

[32] Uniquely Indo-Greek. Next to (Narten) root-present **dēḱ-*: *δέχεται*, Ved. *dāṣṭi*, Ir. **dać-* (*LIV*²: 110–1). Recall the Graeco-Armenian *ἕννῡμι, z-genown* < **u̯es-nu-mi-*.

Ved. *kṣiṇāti*, OAv. *jināiti*, Khot. *jändä*[33]) and even *d^he-presents (*$pleh_1$-d^he- > Gr. *πλήθη*, OAv. *-frādaṯ*)[34] etc. The thematic optative in *-oih_1- is also usually considered an innovation of the non-Tokharian branches. But this is far from certain: Anatolian shows no evidence, Italo-Celtic *-*ā*- is difficult if not uninformative, and Tocharian which only has an athematic non-ablauting *-ʻ*ə*- < *-ih_1- has regularised the paradigm substantially, and it would be conceivable to arrive at the Tocharian forms from paradigm internal analogies (Friis 2024: chap. 6; Friis fortch.). The sigmatic future (from the desiderative *-h_1s-*e/o*- or s-aor. subj.) is also not confined to the branches, in fact Indo-Iranian shares *-(h_1)*si̯o*- with Balto-Slavic (Birwé 1956: 21–25; Lundquist & Yates 2018: 1241).

Under the header "irrelevant features = shared archaisms", Kümmel (2022: 262) lists the preservation of the perfect (as opposed to loss or a merger with the aorist in most other branches), preservation of the simple imperfect (formed just with secondary endings, not with novel temporal suffixes as in Tocharian, Armenian, Italic, Slavic), the preservation of the subjunctive and optative (as opposed to loss of the former in Germanic, Baltic-Slavic and latter in Celtic, Armenian) and finally "vocabulary and poetic language", which he does not consider salient evidence.

On the other hand, Drinka's (*LIV*^{2}: 150–1)(2013b: 403) lists almost the same features as "shared innovation due to late contact in the eastern area": "a remarkably similar temporal-aspectual system", "lengthened theme-vowel subjunctive", "an obligatory use of reduplication for marking the perfect, found much less frequently in the west". But these arguments cannot stand. Since other branches (Germanic, Italic and others) show frequent relics of perfect reduplication even though they do not continue the category, and forms like the Lat. fut. *erō* continues *h_1es-*o*-*oH* (Hom. *ἔω*, Ved. *asā-ni*) (Weiss 2020: 453), these forms can hardly be anything else than archaisms.

While most of the Indo-Iranian and Greek similarities fall in the unfortunate category where they share positive evidence against nothing in the others, the features listed by Drinka are indeed *identifiable* and cannot be taken as evidence of prolonged contact. They are definitely compatible with such a scenario, but we should also consider the asymmetry of the reconstructed verbal system. Contact

[33] Not unique, also in Germ. and Ital.

[34] Uniquely Greek and Iranian.

could have played a role in keeping the tense-aspect system alive, sure, but there was also a clear motivation to either merge semantically and formally similar categories (such as the perfect, (reduplicated) aorist), to recharacterise unmarked forms (like preterite) or to fill out the gaps in the system using the inherited blocks (pluperfect, perfect middle).

5.3.2. The simple imperfective and the augment

The tense/aspect system was clearly asymmetrical to begin with. Only Indo-Iranian and Greek share (traits) of the system where past tense was marked primarily by secondary (non-present) endings on the eventive ("imperfective" and "aorist") aspect stems, but not the resultative-stative ("perfect") stem, here illustrated with the 3.sg.act.:

	Imperfective	Aorist	Perfect
non-past	*-*ti*	–	*-***e***
past	*-*t*	*-*t*	

This was clearly felt as an odd asymmetry for the ancestors of the speakers of all Indo-European branches. In the branches of "the West", formally different solutions were found to the problem, but they all reworked the tense-aspect system into a purely temporal one, by merging the aorist and perfect stem formation and endings and innovating new strategies, e.g. suffixes (Ital. *-*bā*-), periphrasis > suffix (Germ. weak preterit) or by grammaticalising ablaut (Germ. strong preterit) or by utilizing the merged non-indicative aspect-stem.

The age of the augment is another hotly debated issue. It is only synchronically functioning in Greek, Phrygian, Armenian (in past 3sg forms that would otherwise have become monosyllables) and Indo-Iranian, but alleged traces of it has been found in every other branch (Olander 2019a; Goldstein 2022). Clearly, its existence is not an exclusively shared innovation – and it might as well be a shared archaism of the "Eastern" languages. According to one view, the fact that the augment might take part in the accent-ablaut interplay in pre-Vedic (Frederik Henri Herman Kortlandt 1983) speak for an Indo-European age. I am, however, less certain that everything that shines like ablaut must be of Proto-Indo-European age. It is also clear from the existence of the injunctive in from the earliest attested stages of Indo-Iranian and Greek where it had not yet have become an obligatory marker of past that even the complete grammaticalization of it should be regarded a parallel rather than shared innovation (Kümmel 2022: 260). The motivation for doing so was also very clear, since there was very little if any functional opposition between forms like *b^héret* and *h_1e-b^heret*.

The specific combination of facultative augment and unextended preterits is thus a unique Indo-Iranian—Greek isogloss, but most likely only remnants of an earlier unstable system, which all branches solved in their quest for a marked past tense.

5.3.3. To **t* or not to **t* – the present middle or stative endings

The well-accepted connection between the mediopassive and the perfect, as well as the implications of the Anatolian *ḫi*-conjugation for the reconstruction of the Proto-Indo-European verb cannot be treated here. Unfortunately, the PIE reconstruction has immense consequences for the development in the daughter languages. The reconstruction is not just made difficult by the different theoretical stances on the Proto-Indo-European state of affairs, but also by the fact the forms do not agree very well in most branches. With no formal and little semantic agreement reconstructing the prehistory of the middle endings is truly groping in the dark.

What I can say is that it is beyond any reasonable doubt that the perfect endings are ancestral to the middle endings in some way (Kloekhorst 2007: 150). Some are identical or clearly derived from each other. In the 3.sg., all branches – Anatolian as well as in the core – that the active version (Hitt. *hi*-present, core-IE perfect (act.)) is **-e*, whereas the mediopassive (Hitt. *a(ri)*-middle and core-IE eventive mediopassive (or stative, see below)) is **-o*. Already by the proto-language, the middle and perfect were no longer formally of functionally the same, but it is rather opaque exactly what happened.

In the Anatolian languages, there is no system of tense-aspect stems, but what corresponds to perfect is another present class – in the synchronic middle, there is a similar split, but this is not tied lexically to the same verbs as in the present, i.e. the split between *mi*- and *hi*-verbs is completely disconnected from the split between *a(ri)* and *tta(ri)*-verbs (Kloekhorst 2007: 150). The system in Hittite is as follows:

	active				mediopassive	
conjugation	*mi*-conj.		*ḫ̮i*-conj.		*-tta(ri)*	*-a(ri)*
tense	pres.	pret.	pres.	pret.	pres.	pret.
PIE	*-*ti*	*-*t*	*-***e***	*-*s* (?)	*-*t***o***	*-***o***
Hitt.	*-zzi*	*-i*	*-t*	*-s*	*-tta(ri)*	*-a(ri)*

In contrast, what corresponds to the Hittite *hi*-endings, mediopassive present and mediopassive preterit, matches the non-Anatolian perfect, mediopassive present and mediopassive preterit. To make matters worse, it is a matter of dispute what the semantics of this category was in Proto-Indo-European, and if there indeed existed a stative (Kümmel 1996) or intransitive (Kortlandt 1981b) next to what I call the perfect – and indeed how this should be reconstructed (Jasanoff 2003). This stems from the fact that both Anatolian as well as the core attest a descendent of the naked "mediopassive" *-*o* next to *-*to*. In Hittite, there is no functional difference. Some verbs take *-a(ri)* in the 3sg, others take *-tta(ri)*. Luwian verbs do, however, not always agree with Hittite in the choice of ending: *ḱei̯- 'to lie' is *kitta(ri)* in Hittite, but *ziyar(i)* in Luwian. In Indo-Iranian, the same double forms are attested synchronically: *ḱéi̯oi̯ > *śáye* next to *ḱéi̯toi̯ 'lies'. There is also a type with stress on the ending, e.g. *duhé* 'milks'. In a few verbs, there is a functional distinction between these two: Ved. *bruvé*, OAv. *mruiiē* < *mruH-ái̯ 'is called' corresponds to Ved. *brūté*, YAv. *mrūite* < *mluHtái̯ 'invokes, calls (to oneself)'. The verbs that can take the naked ending in the 3.sg. are also accompanied by a special ending of the 3.pl., namely Ved. *-re* < *-r-ai̯, whereas the regular middle ending is *-nte* < *-ntai̯.

It is enigmatic how this system came into being. If one reconstructs a separate "stative" for the proto-language, this would have to have merged with the middle in every single branch but Luwic and Indo-Iranian. On the other hand, reconstructing "two allomorphs of the 3sg.mid. ending [for] PIE, older unproductive *-*o(r)*, and younger productive *-*tor*" and asserting that "[a]rchaic *-*o(r)* was gradually replaced by productive *-*to(r)* within the IE languages, but was exceptionally retained under certain conditions" (Lundquist & Yates 2018: 2144) requires accepting allomorphy of the same endings from the earliest split of the proto-language and until the attested Old Indo-Iranian states. Sometimes Old Irish passive/impersonal verbs, like *berair, -berar* 'is borne' are connected here (Jasanoff 2003: 49), but the "stative" is only formed to athematic verbs in Indo-Iranian and to thematic ones in Old Irish. Celtic – and Italic (umbr. *ferar*) – *t*-less middle forms should rather have another explanation (Villanueva Svensson 1999).

Similarly, according to Yoshida (2013) the attested development of the distribution of the endings in Hittite reveal that *-tta* is significantly younger and should probably not be reconstructed for the proto-language. This reconstruction, in turn would make force us to reconstruct the introduction of the *-*t*- onto the "middle" ending *-*o* at least thrice: Once in Hittite, once in Palaic (which Yoshida (2013: 163) assumes is a parallel innovation to account for Pal. *kītar*) and at least once in the non-Anatolian languages, whereby it did not oust "naked" *-*o* before after the break-off of Indo-Iranian. To him, this is more attractive than assuming that the process of replacing *-o- with -to- was active for more than 3000 years (Yoshida 2011). In light of the similarity between 3sg *-*o* > *-*t-o* and 3pl *-*r̥(-)* > *-*nt-o*, it is reasonable to assume that the initial consonant was imported from the corresponding secondary ending. The same also happened in the 2.sg., where Greek, Latin and Iranian replaced original "perfect"-looking mediopassive pres. *-*th₂a*-, pret. *-*th₂e* (Hitt. pres. *-tta(ri),* pret. *-ttat(ti)*) with *-*so*- in analogy with the active (Lat. pres. *-ris* << *-re,* -RUS << *-*so*-; Gr. pres. -*σαι* (anal.), Arc-Cypr. -*σοι*, pret. -*σο*; Av. pres. *-he, -še*, pret. *-sa* < *-*sa(i̯)*). In Indic, this analogy is only partial: the secondary ending is -*thāḥ*, but the primary is *-se, -ṣe* (Weiss 2020: 411; Sihler 1995: 475; Kümmel 2018b: 1914) It is, however, a remarkable Analogy to *insert* an ending, and not add it at the end of the word, as is usually the case, e.g. Ved. *asā́m* 'I was' < **h₁e-h₁es-m̥+m* ~ Hom. *ἦα*. Nowhere in Indo-European do we find the "usual" pattern: **-*o-t*.

5.3.4. The primary and secondary mediopassives

Not just is the introduction of *-*t*- into the mediopassive endings (or the parallel loss of the stative) phylogenetically problematic, to is the other end of the ending, so to speak. For structural reasons, the secondary middle ending is often reconstructed as *-*o* (e.g. Weiss 2020: 411). This is arguably closer to the cradle, the connection with the "perfect" *-*e*. The naked *-*o* is, however, not attested as a secondary ending. Greek which has no traces of the *t*-less forms does have pret. *κεῖτο* '(s)he lay'. Otherwise, the naked *-*o* is only attested in present function; in Anatolian as the *a(ri)*-class (e.g. Hitt. *eša(ri)*; CLuv. *zīi̯ari*, Lyc. *sijẽni* < **k̂ei̯-o*-). Enigmatic as it is, the Tocharian verb *ste* '(s)he is' might also continue this ending directly if it does in fact go back to **sth₂-ó* 'is stood, stands' directly – this is, however, not universally accepted (Malzahn 2010: 691, 398 n. 48). Otherwise, the secondary ending is only found with the *-*t*-/*-*nt*-, e.g. *áśayat* = *ἔκειτο*).

In the present, the naked *-*o* is also widely unpopular. In Anatolian, it is accompanied by -*r* or -*ri*. This has been variously explained. Notably, Yoshida (1990; 2011) has argued that they go back to accented and unaccented variant of *-*or*: *-*́or* would lead to loss of the *-*r*, whereby the naked ending *-*o* would (re)appear, *-*ór* would be preserved and later recharacterized with the -*i* of the active preterit -*tti*.

A classic Indo-European cladistic isogloss is indeed the distribution of the mediopassive endings *-*tor,* *-*sor* and *-*ntor*. Before the discovery of Anatolian and Tocharian, it was held to be a dialectal feature; the languages in the West would have -*r* as a middle counterpart of the active hic-et-nunc particle -*i*. It has now become clear that it is rather a question of chronology than of geography. Anatolian is an *r*-branch, but this -*r* is repeatedly recharacterized by the -*i* of the active over the course of its attested history. Tocharian, Italic and Celtic are somewhat consistently *r*-languages to the extend that they preserve the endings in question, and it is widely held that Greek, Indo-Iranian, Germanic and Albanian innovated by replacing the *-*r* with the particle *-*i*. There are several problems with this analysis (Villanueva Svensson 2014).

Comparison between Latin and Sabellic reveal that Proto-Italic had a different primary-secondary distinction than *-*tor/-to* (Weiss 2020: 413), namely primary *-*tro*, secondary *-*tor* which gave Umbrian pres. -*ter, -nter*; pret. -*ntur*. There may be phylogenetic arguments for Italo-Celtic here instead (Jasanoff 1997a). In Tocharian, preterit active endings do not usually reflect the PIE secondary endings, and primary *-*tor* would yield **-*ter*, not -*tär* (Peyrot 2022: 95). It is also strange why the understandable present-middle marker *-*r* would be replaced by a distinct present-active marker – and why this would show up in the Tocharian preterit middle (Friis 2024: 32).

In Greek, Indo-Iranian, Albanian and Germanic all show no trace of *-*r* but consistently mark the present with *-*i̯* as in the active (Alb. -*et* < -*etë* < *-*oi̯-toi̯* (Matzinger 2006: 126), Goth. -*ada* < *-*ai-dai*[35]). Armenian, on the other hand, does not give any evidence. The connection of *berēr* < **bereyr* < **ber-e-tr-o-* does not hold (Schmitt 1981: 141). Phrygian is usually said to be the crux of this comparison since

[35] Passive *haitada* 'is called'. Proto-Norse attests **haite** 'am called' < *-*ai̯* < *-h_2*ai* where the passive is remarked by the hic-et-nunc, but the strong preterite is not: -Ø or -*a* (Poulsen 2020).

it attests both *-*toi̯* and *-*tor* as the only Indo-European language. This is, however, another mirage (Kortlandt 2016; Obrador-Cursach 2019). Old Phrygian only has -*toy*, and while there are neo-phrygian examples of -*τορ*, but they only occur in the same formulas as the same verbs ending only in -*t*. They probably do not belong here (Ringe 2017b: 40, n. 29).

It is, on the other hand, possible that the middle endings had no hic-et-nunc marker to begin with (Kortlandt 1981b), and that the source of the -*r* lies elsewhere – the usual suspect is the perfect 3.pl. *-*r̥š* in Indo-Iranian (Jasanoff 1997b).

5.3.5. Unique endings

5.3.5.1. The verbal dual

Like the Iranian nominal dual endings mentioned above, there are a few archaic or isolated verbal endings in Indo-Iranian. The mediopassive *-*r*- are not attested anywhere also, also not in Anatolian, where the 3.pl. *-*r*-endings are only Active. As such, it is perhaps not impossible to imagine other archaisms surviving in Indo-Iranian, long after Anatolian had innovated. The Indic 2.pl.perf. -*a* which is difficult to explain as an innovation belongs this category (Kümmel 2018b: 1912; 2022: 263).

In the verbal dual, the specific set of endings is unique to Indo-Iranian and Greek (Kümmel 2022: 259). But as most branches do not continue the (verbal) dual, the isoglosses are likely unidentifiable or uninformative.

5.3.5.2. The secondary 2.du. *-*tom*

Greek -*τον*, Vedic -*tám* and Av. -*təm* and OP -*tam* all seem to reflect 2.du. *-*tom* which is not found anywhere else. Baltic and Slavic continue -*tah*$_2$ in this function (Olander 2015: 342–345). Since Baltic merges primary and secondary endings, it is not certain that Lith. -*tà* reflects the secondary ending, but it matches the others formally. As the does the **-tu** of the Umbrian 2.pl.imp **eta-tu** ‘you must go!’ *-*tå̄*. Formally, this is a clear Indo-Greek isogloss, but it is likely an archaism. Functionally, it is a little more difficult since the Iranian forms (-*təm, -tam* < *-*tam*) are used as the 3.du., although they match the Vedic and Greek 2.du.

5.3.5.3. The secondary 3.du. *-*tom*

The 3.du. secondary ending of Greek and Vedic is *-*tah₂m*: Gr. Dor. -*τᾱν*, Hom. -*την*; Ved. -*tā́m*. As stated, Avestan and Old Persian use the ending *-*tam* which formally matches the 2.du. In Slavic, the ending *-*te* is compatible with the reconstructions required for the Indo-Iranian primary endings 3.du. *-*tes* (Ved. -*tas*, Av. -*tō*) and 2.du. *-*th₁es* (Ved. -*thas*, Av. 3.du. -*θō*) (Olander 2015: 346–348).[36]

The Avestan secondary 3.du.act. ending -*tąm* quoted by Kümmel (2022: 259) should probably not be used as evidence. Although it would indeed look like the expected secondary 3.du. ending, I do not believe the ending belongs to a finite verb. In fact, -*tąm* is not listed as a 3.du.-ending in the usual grammars (Hoffmann & Forssman 1996: 179; Gotō 2013: 87; Martínez & de Vaan 2014: 84; Skjærvø 2007: 875; 2017b: 537). The ending is attested in the hapax legomenon *carətąm* (F. 19/602), which is treated as 3.du.aor./inj. 'they (two) who make' by Tremblay (2008: 28) and by Klingenschmitt (1968: 181).[37] The form is also unmentioned as an attestation of the verb *kar-/car-* 'do, make' (Cheung 2007: 236; *LIV*²: 291–2; *LIV*³), but it is occasionally mentioned as a possibility (e.g. de Vaan 2003: 579–80, with no further discussion). It would be fantastic if this hapax truly preserved the original 3.du. ending in 3.du. function in the light that the 3.du.: The primary endings are unattested in Old Persian, but in Avestan -*θō* (corresponding to Ved. 2.du. -*thaḥ*) and -*tō* (corresponding to Ved. 3.du. -*taḥ*) are used interchangeably for the 3.du.

Since Klingenschmitt's unpublished dissertation is hard to come by I will give his argumentation in full (italics added and references standardised by me):

> (602) *carətąm krt'l'n´* ‚welche machen': Der von Bartholomae (1904: 582) für das sonst unbelegte *carətąm*[1] vorgenommene

---

[36] The Tokharian 3.du.prt.act. endings TB -*ais*, TA -*enas* are enigmatic (Malzahn 2010: 48). According to Hackstein (1993: 55–6), they share the morpheme *-*ai*-, apparently from the 3.du.pres.med. -*ai-tär* which he connects to Ved. them.3.du.med. -*ete*, Av. -*aēte* < *-*oi̯-toi̯*. Klingenschmitt (1994: 411; 2005: 435) explains the endings as TB -*ais* < *-*āi̯sə*, TA -*enas* < *-*ā-i̯-na-sə* where the -*na*-infix of TA should be the nom.acc.du., and *-*i̯sa* would have a complicated analogical origin, also connected to the *-*i̯*- of the Indo-Iranian middle dual. At any rate, PT *-*ai̯*- or *-*ā-i̯*- is unconnected to the other languages.

[37] I know of this piece for *Forschungsreschichte* thanks to Kümmel who explained the ending in a pers. comm. to Thomas Olander (19 July 2022).

Stammansatz *carət- ‚machend‘ widerspricht den Bildungsgesetzen, die in diesem Falle ein *kərət- erwarten ließen. Möglicherweise ist *carətąm* 3.Du.Aor. (*ákartām* AV.) der Wz. *kar* ‚machen‘. Zur Wiedergabe einer finiten av. Verbalform durch ein mp. Nomen agentis wäre z.B. *tašaṯ*, PÜ [Pehlevi-Übersetzung] *t'šyt'l* Y. 29.7 zu vergleichen.

> 1) Reichelt (1901: 138) erwägt wegen *frašō.carəθrąm*, PÜ *plškrt-krt'l'n´* Y. 24.5 eine Korrektur von *carətąm* zu *carəθrąm*.

(Klingenschmitt 1968: 181)

The *Farhang-ī ōīm* is an Avestan-Pahlevi wordlist, and it does not contain any context of which forms it translates. *carətąm* simply occurs in the section on words beginning with *č*. Thus, the context is not ideal for an attestation of a verbal dual. Additionally, the translation with a verbal noun is puzzling. While Klingenschmitt does provide and example,[38] it is worth mentioning that the wordlists gives the same translation for the alleged 3.du. *carətąm* as it does for the noun -*carəθrąm* 'creator', namely *krt'l'n´* /kartārān/ (Bartholomae 1904: 1008). Accordingly, Reichelt's (1901: 138) suggestion – which Klingenschmitt does mention – that *carətąm* is to be read as *carəθrąm* should perhaps be reconsidered, even if the letter forms do not resemble each other strikingly.[39]

I would rather think that *carətąm* should not be taken as evidence for the 3.du. being preserved in Iranian. But even if it were, I would presume that it would have been used for the 2.du. (-*θo*) and 3.du. (-*to*) interchangeably (though they are both only attested in 3.du. function), just like the primary endings and its cousin, the etymological 2.du.-ending -*təm* which is only attested in as a 3.du.

---

[38] The example is: *tašaṯ* 'he formed, created' which is translated as a finite form *tāšīt* (F 18/599 (Reichelt 1900: 204; 1901: 141; Damaspji & Haug 1867: 29, 68; Klingenschmitt 1968: 180)).

[39] t <𐬙> and θr <𐬚𐬭>

	Pers.	Greek	**PIE**	PIIr.	Vedic	PIr.	Avestan	OP
SE	2.	*-τον*	***-tóm**	*-tám	*-tam*	*-tam	– ↓	– ↓
	3.	*-τᾱν,* *-την*	***-táh₂m**	*-tā́m	*-tā́m*	*-tām	*-təm,* [*-tąm*]	*-tam*
PE	2.	*-τον*	***-tHés**	*-tʰás	*-thas*	*-θah	– ↓	–
	3.	*-τον*	***-tés**	*-tás	*-tas*	*-tah	*-θō, -tō*	–

5.3.5.4. The primary 2. and 3. du.

Indic is the only branch to distinguish all four of the primary and secondary second and third person dual. Iranian, as stated, clearly merges the primary endings of the 3.du., and likely did the same for the 2.du., but the original 3.du. ending – and the entire 2.du. function is unattested. Greek, on the other hand, keeps only the 3.du. secondary ending distinct and continues only the ending corresponding to the Vedic secondary 2.du. as the secondary 2.du. and primary 2.-3.du.

In the primary 3.du., Tocharian surprisingly agrees with Greek in continuing **-tom*: TB *-teṃ* (unattested in TA) (Malzahn 2010: 48) probably goes back to **-tom*, but it cannot do so directly as word final (non liquid) consonants were dropped. Kümmel (2022: 259) suggests that the Graeco-Tocharian correspondence "might support the use of **-tom* for the 3[du.]". If that is the case, a particle could have been added in pre-Proto-Tocharian times, but this is unclear (Hackstein 1993: 50-1). If **-tom* was originally only the ending of the 2.du., it could have been expanded and later recharacterised to disambiguate it. Hackstein (1993: 50-5) elaborates this scenario as follows: The 2du. **-tom* > PT **-tæ* could have replaced the (pre-)PT 3du. **-tå* < **-tām* on the pattern of the primary 2-3du. where PIE 2du **-th₁es* and 3du **-tes* would have merged in PT 2-3du. **-cə*. This **-cə* would incidentally be identical to the 2pl act PE, so, according to Hackstein, the new ending **-tæ* was spread to the PE as well in order to disambiguate the dual from the plural, resulting in a tense-neutral ending of the PE/SE 2-3du. (maybe even identical to the MP SE **-tæ* < **-to*). This situation led to the addition of *-ṃ*, the marker of the third person in the 3pl-ending, in TB (Hackstein 1993: 55; Klingenschmitt 1994/2005: 435).

In conclusion, the Greek and Tocharian mergers are probably distinct processes. Vedic most likely continues the original distribution, which makes the Indo-Iranian-Greek correspondence an archaism, and the Balto-Slavic-Umbrian connection innovations.

5.3.6. Extending the tense-aspect system

As I have sought to elaborate above, the innovations in the “Indo-Greek” verbal system are difficult to derive from the Anatolian system – and vice-versa. There are, however, another line of innovation which Greek and Indo-Iranian underwent – separately. Drinka, Rix and Birwé would assign these to areal contact in prehistory, I would rather explain them as pressure from within the system, given that no shared morphology can be reconstructed. Both branches have filled the gaps in the asymmetric tense-aspect system by incorporating the perfect stem fully into it. Two new categories were invented to do so: The perfect middle and the perfect past (pluperfect). This is similar to the creation of a complete preterit to the *hi*-conjugation and the mediopassive preterite, whereby the original “proto-perfect-middle” ended up filling several slots in the tense-aspect grid.

Kümmel (2020: 18) gives the following grid as the starting point for PIE or P-CIE. I believe it is very instructive of the asymmetry of the system. On the one hand, the “perfect” is oddly isolated in not having an active/middle distinction, but on the other hand it is simultaneously the connected to the present and preterit of the *hi*-verbs but also to the mediopassive – directly to “Type 1” (the Anatolian *a(ri)*-verbs) and by influence from the (eventive) *mi*-verbs also “Type 2”:

	*m active		*χa-active		Perfect	Middle	
		Present		Present		Type 1	Type 2
1SG	*-m	*-m.i	*-χa	*-χa.j	*-χa(.j)	*-χa (*-χ?)	*-m.o?
2SG	*-s	*-s.i	*-tχa	*-tχa.j	*-tχa(.j)	*-tχa	*-s.o
3SG	*-t	*-t.i	*-s	*-e.j	*-e(.j)	*-o	*-t.o
3PL	*-(é)nt	*-(é)nt.i	*-ēr/-r.s	?	*-r/ ēr.i	*-r.o	*-nt.(r)o

The reconstruction of the perfect and the middle is complicated enough as it is, but the pluperfect and perfect middle reveal that these categories cannot be inherited.

According to Jasanoff (2003: 34–43), the (core-)Indo-European starting point did have a contrast in voice in the perfect – but the reconstruction of a pluperfect for this stage is *not* the majority view (*pace* Lundquist & Yates 2018: 2140; see Ringe 2024: 221). Rix (1976: 257) suggested that PIE simply built a pluperfect by adding the augment to the perfect stem. This could be the case as soon as the augment came into being – but it is completely unattested. Instead, the Greek and Indo-Iranian paradigms differ which suggests that the branches took the full consequence of incorporating the perfect into the tense/aspect system when they were no longer in the vicinity of each other.

In Indo-Iranian, the pluperfect is straightforward. It is built by adding the augment to the perfect stem and inflecting the verb with secondary endings. It is possible that Avestan *-ərəš* in *cikōitərəš* shows that there was a distinction between a primary perfect **r̥* > *-ur, -arə* and a secondary perfect **-r̥š*, but the details are difficult (Kümmel 2018b: 1914; 2000: 42ff.; Gotō 2013: 120; Skjærvø 2007: 874). Otherwise the "regular" perfect uses "primary" perfect endings, and the pluperfect takes the secondary endings of the imperfect-aorist system: *a-ved-am* 'I knew' < **h_1e-u̯oi̯d-m̥* (transposition), 2.-3.sg *á-jagan* 'went' < **h_1e-gʷe-gʷom-s/t*. In Greek, on the other hand, even the dialects vary. There are very few pluperfects in Homer, and epigraphical material rare offers additions. Therefore, the situation is clearest in classical Attic where it is formed by adding the augment to the perfect stem, extending this with an unclear element *-*e*-. This gives forms like Hom. 1sg *πεποίθεα* < *-*e-m̥*. In Archaic Attic, this *-*e-a* contracts to -*η*, but the ending is replaced by -*ει-ν* (e.g. *ᾔδειν* < **e-u̯id-e-ã-m*). In the 3.sg., the ending could have been that of the perfect, originally: Hom. 3.sg. *ᾔδεε* < **e-u̯id-e-e* seems to suggest that, as does Cypr. *o-mo-mo-ko-ne ὀμώμοκον* 'they had sworn' (Ringe 2024: 221; Rix 1976: 258).

Similarly (or perhaps rather, dis-similarly), both branches built a perfect middle (and a pluperfect middle!), but the formations are not identical. Again, they use the building blocks they have at hand: The perfect stem and the primary middle endings. Mediopassive perfects are surprisingly well-attested already from Mycenean – but this is obviously also an artefact of the type of texts. Mediopassive perfect participles are indeed ideal to identify objects in an inventory, i.e. *ke-ka-u-me-no* 'burnt. These forms also reveal that the mediopassive perfect was productive – and regularised itself. There are no relic endings, *-*s*- is always restored (*μεμνήσαι*) and most forms are completely detached from the inherited accent-ablaut system, i.e. they have multiple full-grades: λέλειπται 'is left, remains' < PGr. **le-lei̯kʷ-tai̯* < "PIE" **le-l(e)ikʷ-(t)e(i̯)*, and they frequently occur formed to secondary verbs (Ringe 2024: 222). In Indo-Iranian, on the other hand, the perfect middle is formed with a semi-unique set of endings. The 1sg, 2sg, 2du, 3du, 1pl, 2pl are the same as the primary endings (and thus to some degree the same as the Greek ones before the restructuring). The 3sg and 3pl are, however, always those of the -*t*-less perfect formations: 3sg *-*ai̯* and 3pl -(*ā*)-*rai̯* > Ved. *dadé,* YAv. *daiδe* < **de-dh$_3$-oi̯*; Ved. *riricré k*< **lilikʷ-roi̯* (Gotō 2013: 121). On the basis of a very few lexical correspondences where the Greek ablaut has not been remodelled _as_ much, e.g. *_k̂lei̯-_ 'lean' to which Greek and Indic form perf. middles, Ved. _śiśriye_ 'has been placed against' < *_k̂e-k̂li-roi̯_, Gr. Hom. _ἐκέκλιτο_ 'had been placed against' < *h_1_e-k̂e-k̂li-to_, Ringe (2024: 222) nevertheless states:

> Since Greek and Indo-Iranian clearly started out in the same area of the Central IE dialect continuum, it is possible that this example is a common inheritance

At this point, I would rather agree with van Beek, who thinks that a parallel innovation is possible because of the discrepancy between the Greek use of the proper middle endings against the Indo-Iranian stative (van Beek 2022a: 197). However, given that the "stative" is unattested in Greek, and the whole paradigm is continuously regularised, I think it is _more_ than likely that the formations postdate the latest common ancestor of the two branches.

Both branches also form mediopassive perfect participles – but as Euler concludes:

> Die Perfektpartizipien des Mediopassivs im Indoiranischen und Griechischen können im Gegensatz zu den Aktivpartizipien wie auch die finiten Perfektkategorien kaum indogermanisches Alter besitzen
>
> (Euler 2017: 53)

It is indeed the general tendency of Indo-Greek morphology: Late innovations built on archaic inheritance and a drive to fill in the blanks of an asymmetric verbal system.

6. Dating and locating linguistic prehistory

6.1. Dating of linguistic prehistory

6.1.1. Introduction

There are no apparent and undisputed means for dating language. While linguistic sources may to some extend be dated scientifically, language itself cannot. This only gets truer for reconstructed language which in essence is materially disconnected from historical records. However, there are some methods for estimating the especially the *terminus post quem* of linguistic speciation events.

Among linguists, the most trusted method is archaeolinguistic dating. It is beyond the scope here to give a thorough introduction, for which I will refer to important handbooks and ground-breaking works (Carling 2024; Olander 2019c; Campbell 2020: chap. 15; Anthony & Ringe 2015; Anthony 1991; Mallory & Adams 2006; 1997; Mallory 2013; 1989; 1996; 1975).

What is important to have in mind at present is that archaeolinguistic dating revolves around connecting reconstructed vocabulary to artefacts or real-world phenomena. If these are culturally specific, not universal to the human experience and archeologically dateable, we can use this as corroborating evidence for dating. In a nutshell, if two related languages share the same etymon (preferably not just the lexical root, but a specific formation) in the meaning of a phenomenon which did not occur until a specific archaeologically dateable time in prehistory, we can infer that the ancestral speakers of these languages knew these phenomena and coined the term when they had not diverged yet, and thus these two languages had not yet split by the time of the invention.

It should be acknowledged that this method is by no means flawless – and there are linguistic as well as archaeological problems connected with it, such as interference of inherited productive morphology, semantic shifts, early loanwords, lack of attestation, lexical loss, poorly preserved archaeological materials, non-dateable items and many more. Accordingly, the method should not be employed to post-rationalise conflicting evidence into fitting one model, and we should keep in mind the grave difference between positive evidence, negative evidence – and lack of evidence.

In fact, critique of the method in general (e.g. Sims-Williams 1998), findings in Indo-European (Clackson 2013; Heggarty et al. 2023b: 19–22) – or from other language families (e.g. Kaufman & Justeson 2009: 223–7) is welcomed by the community of historical linguists working in this field.

6.1.2. *Wörter und Sachen* from the Steppe or Lexical roots from Anatolia

Regarding the spread of the Indo-European language family and the dating of its protolanguage, it is well-known that there are two competing hypotheses. Interestingly, the divide between scholars does not just follow fields – which is perhaps to be expected in an interdisciplinary debate – but also the preferred data within linguistics. In a nutshell, historical linguists relying on archaeolinguistic data (cognates for culture words) side with archaeologists, anthropologists and geneticists who link the dissolution of Proto-Indo-European to the expansion of the Serednij Stih and Jamnaja cultures in the fourth and third millennia BCE with "elite-dominance" as the social model (Gimbutas 1952; 1970; Anthony 1990; 2017; Olander 2019c: 24–7), whereas computational linguists who prefer quantitative statistical methods rely on lexical data (cognates in culture-free semantic slots) tend to side with archaeologists who connect the spread of Indo-European to the spread of agriculture from Anatolia into Europe dated already in the seventh millennium BCE (crucially, Renfrew 1987; see Olander 2019c: 12–6 for further refs.).

It should be emphasized that the findings of *aDNA* (Lazaridis et al. 2025; Lazaridis et al. 2022; Haak et al. 2015; Allentoft et al. 2015) are much more compatible with the Steppe Hypothesis which is supported linguistically by the correlation of shared vocabulary of semantic fields with archaeological evidence, than with the Anatolian Hypothesis which is based on archaeological theories and supported by the dating of linguistic speciation events estimated from the shared number of words in basic meanings.

6.1.3. Bayesian inference and linguistic dating

The results of the latter type of studies (Dyen, Kruskal & Black 1992; Gray & Atkinson 2003; Rexová, Frynta & Zrzavý 2003; Bouckaert et al. 2012; 2013; Heggarty et al. 2023a) have been met with scepticism by most historical linguists (Ringe 2022: 67) – or downright ignored (Adams 2023: 217). The reason for this is not just the difference in methods; in an ideal case, different methods and different data would yield similar results, if these were indeed correct. There are also grave differences in the "linguistic evidence" examined in such studies as opposed to archaeolinguistic research. It is these we wish to clarify.

The recent study by Heggarty et al. (2023a: 8), for instance, argues for a hybrid model of the "steppe" and "Anatolian" hypotheses – despite their efforts to sweep the findings of Narasimhan et al. (2019) under the rug, namely the presence of Steppe Ancestry including DNA associated with the Corded Ware culture in modern Indo-Iranian speaking populations.

Importantly, the migration routes plotted on maps are *only* based on the estimates of dates of linguistic splits. These estimates are calculated by a complex mathematical model applied to lists of *cognates* – words descending from the same common ancestor – with the same *basic* meaning.

Contrary to the archaeolinguistic evidence, such data have no inherent social model and tell nothing about specific cultures; they are purposefully sought out to be universal to the human experience. By contrast, we find it more fruitful to examine the reconstruction of meaning in semantic fields that can be correlated with the archaeological records. If all Indo-European languages share identically formed *cognates* all referring to a phenomenon whose invention can be dated archaeologically, we can reconstruct a common ancestor and assign an absolute date to it based on the archaeological phenomenon. Words can indeed change their meaning, but they are not likely to do this multiple times independently. Additionally, words can be lost or replaced; but again, it would be quite the coincidence if all branches of Indo-European independently lost all traces of inherited words for an entire semantic field, such as agriculture, but preserved the cultural practise.

6.1.4. Basic vocabulary as a source for linguistic dating

The reason for examining the "basic" vocabulary is that this is the type of lexicon which is the least prone to borrowing and supposedly develops more consistently over time than other types of linguistic material. However, most linguists agree that vocabulary, basic or not, is not very useful for establishing language relationships (Clackson 2022). It remains unproven if the lexicon is a good proxy for other linguistic innovations. Absolute dating of linguistic divergence remains controversial (Piwowarczyk 2022b: 37–8; *pace* Greenhill & Gray 2012; Greenhill, Heggarty & Gray 2021). Previous results of Bayesian analyses have been proven time and time again to be highly unreliable (Ringe 2017a: 67; Chang et al. 2015; Pereltsvaig & Lewis 2015; Ehret 2011: 22–4; McMahon & McMahon 2006).

More recent Bayesian studies still suffer from critical issues, other than relying on a questionable basic assumption. The necessary amount of characters and the amount of "runs" of the software required to obtain correct results remain unknown (Nichols & Warnow 2008: 784–5), but this may be irrelevant if the current results come "close enough" after all. For finding the right phylogeny, the data offers inherent issues. Undetected prehistoric borrowings are inevitable (Scarborough 2020), and it is impossible to create a model of prehistoric and unknown contact situations (McMahon & McMahon 2006: 156). It is also well known that small changes to the model can yield very differing results (Piwowarczyk 2022b: 37; Ringe 2017a: 68). Dating of phylogenetic speciation events are, of course, never more reliable than the actual phylogeny – if the branching is wrong, the date of the branching does not even exist. Even for the best Bayesian analyses on the best available data, the higher-order subgroupings are very unlikely (with probability values as low as 0.21–0.25 (Heggarty et al. 2023b: 57)). It is also clear that there is no equivalent of a molecular clock in linguistics; it goes without saying that more time allows for more innovations, but no model has been able to estimate the rate reliably. Even if such a model existed, the lack of directionality in cognacy data – that is which branch innovated and which branch, if any, preserved the ancestral state – would not allow us to "use it to extrapolate into prehistory with any confidence" (Ringe 2022: 60). The alternative to the "fixed clock", the parameters that make up the "relaxed clock", are neither transparent nor reliable:

> Until we have more and better data, more extensive and structured cross-linguistic data bases available for quantitative study, more sophisticated simulations, and hence a clearer idea of the signals we can expect under particular social and linguistic circumstances, we cannot hope to provide accurate and generalizable dates
>
> (McMahon & McMahon 2006: 159)

Even when the models allow for each branch or each meaning (cognate class) to have a unique "mutation rate", these are only calculated on the basis of the calibration of the dates of attestations (Heggarty et al. 2023a: 10; 2023b: 53; Pellard, Ryder & Jacques forthc.: 22). No matter how great the model is it is crucial to keep in mind that there is no further chronological input that the historical dating of the languages sampled.

Interestingly, the completely reworked database (IE-CoR) which underlies Heggarty et al. 2023 does not include Romani and Domari, spoken by the Roma and Sinti peoples, among the 17 modern Indic languages. These languages are generally regarded as examples of quite extreme contact situations (Matras & Bakker 2003; Matras & Tenser 2020), and it would therefore have been very interesting to see how the model would date their split from the other Indic languages (Pereltsvaig & Lewis 2015: 95–106). The fact that exactly these languages have been omitted does raise the suspicion that even a "relaxed clock" is not universally applicable.

6.1.5. Issues of the Bayesian inference of Heggarty et al. 2023

Additionally, Heggarty and colleagues' (2023b: 57) grouping some of the well-attested and thoroughly researched branches is remarkable. They find both modern standard registers of Norwegian to be more closely related to Danish than to the other West Nordic languages, which is untenable based on phonology and morphology. There is also something peculiar about the West Germanic languages, where most medieval varieties are returned as outliers, and the modern varieties of Frisian, Dutch, Flemish, Luxemburgish, German and Swiss German descend from the same unattested dialect which is only the cousin of Old High German, Old Saxon, Old Frisian and Middle Dutch. These varieties, in turn, all died out without descendants. Although some details of Medieval Western European linguistic history remain opaque, the Bayesian study by Heggarty et al. neither confirms what centuries of research has already established, nor sheds new and comprehensible light on recent history in one of the most well-documented eras and areas of the Indo-European speaking world.

These issues are discussed by the team. Of course, they highlight that the MCC tree is merely a visualisation tool of a probability distribution. They do engage directly with a couple of the problematic findings of their phylogeny (Heggarty et al. 2023b: 89–90) and acknowledge that these surprising differences probably arose because the West Germanic languages formed a dialect continuum, whereby the lexicon of these languages was not just subject to genetic change over time and detectable loans but was also affected by "semantic (re)alignment" whereby inherited lexemes changed their meaning because of pressure from a neighbouring speech community. This is indeed likely, and such realignments or semantic calques between diverging dialects is exactly a reason for not trusting cognations in the basic vocabulary as evidence for phylogenetics. If the model struggles with known facts, it is doubtful why its inferences on prehistory should be trusted.

In general, the Bayesian date estimates are irreconcilable with archaeological and genetic data, and exist only in an isolated ecosystem. For the dates to be applied to other linguistic material than the "basic" vocabulary, all other evidence must be explained away as independent innovations or pure chance. Adherents of Bayesian inference insist that the result of the analysis is only likelihoods, not a finite tree; but for all practical purposes, this difference is miniscule – by this I mean "practical" quite literally: If the inference has any truth to it, it should be treated accordingly and thus the model should have radical effects on the reconstruction of the protolanguage. If it does not, and the protolanguage is still to be reconstructed without regard for the phylogeny, the inference has added little more than uncertain numerical values of the uncertainties of the past.

6.2. Excursus: Ancestry, descendance, lists and languages

6.2.1. Latin and Romance

The study of Heggarty et al. does not force any ancestry constraints on the languages. That means that the authors let the model explore whether the ancient and medieval languages in the database, such as Old English and Latin, can be the direct ancestor of their modern descendants, such as English and French. In fact, a direct lineage is only returned in a few cases of the sampled languages, namely between Classical and Modern Armenian and between some varieties of Greek (Mycenaean, Classical Attic and New Testament Koiné) (Heggarty et al. 2023a: 3).

Forcing ancestry constraints is a hot topic within Bayesian inference. It is usually not done so, but Chang et al. (2015) "demolished" (so Ringe 2022: 59) these earlier works by showing how the dates shifted dramatically when ancestry constraints were applied. Heggarty et al. do not apply them "given the known objections to doing so" – without further clarification (Heggarty et al. 2023b: 72). They raise objections to forcing ancient *written* languages to be the ancestors of modern *spoken* ones (Heggarty et al. 2023b: 91). This strict methodological approach is applaudable, but it also enforces some limitations on what can be generalised from the analyses. Reducing *language* to 170 semantic slots is quite a stretch, and accordingly, such analysis based on 170 semantic slots can only answer whether or not a wordlist can go back to the wordlist of the presumed ancestor, not the language as a whole (Ringe 2022: 60; Chang et al. 2015: 207). The chronology og Chang et al. is not unproblematic. Their model dates the split of Proto-Romance to 1000 ce which is later than the earliest attestations of medieval Romance (Heggarty 2021: 382).

Although Heggarty and colleagues would like to add to the debate on whether Latin is the direct ancestor of Proto-Romance, their material can, in fact, not answer this question; only whether the Romance *wordlist* can go back to the wordlist coded for Classical Latin. Even the lexical relationship between them is debated. As is, of course, the larger question of continuity between Old Latin, Classical Latin, Vulgar Latin and Proto-Romance (Vincent 2016: 6–7). Chang et al. (2015: 207) claim that "[l]iterary and colloquial Latin share the same basic vocabulary", whence it – to them – is not misleading to view Romance as the direct descendant of Latin and assigning the differences in cognations no semantic shifts "already nascent" in Latin. Heggarty et al. (2023b: 72), however, insist that "[e]ven one difference, in a single meaning of the 170 in the IE-CoR reference set logically entails separate sublineages, and that ancestry is not fully direct" (2023a: 3). Which, again, can only hold true for the word list, not for the entire language for which such a logic would be absurd. Thus, Heggarty and colleagues (2023b: 72) fall into their own trap when criticising Chang et al. (2015: 266) for recognising early structural differences in *the vowel systems of* the Latin regional texts. If cognations were a reliable proxy, the phonological system should be ignored when discussing ancestry as it has no say in the preservation of shared lexemes.

6.2.2. The survival of Classical Latin

In the discussion of the differences in *register* between the coded varieties of "written, literary, formal" and "fossilizing" Latin on the one hand and the spoken Romance languages on the other, a key point of interest is revealed:

> In the meaning MOUTH, for instance, it is not the issue whether any tokens of *bucca* are found in Classical texts, and/or any tokens of *os* in supposedly 'Vulgar' texts […] On this strictly cross-comparable question, evidence from across Romance languages points clearly to the default assumption that Proto-Romance

> already did not have *os* as its default word for MOUTH, even if the lexeme was still known
>
> (Heggarty et al. 2023b: 92)

This is quite provoking to many linguists. It may be that it plays no role for the Bayesian dating estimate whether the token existed in the proposed ancestor or not, but it does indeed for etymology and the reconstruction of linguistic evolution. This is also evident in the in the IE-CoR protocol for coding romance characters whose Latin ancestor does not have any cognates (Heggarty et al. 2023b: 34). To other linguists, it is indeed an "issue" if the inferred or reconstructed ancestor of Proto-Romance is found in Classical Latin, and that it only requires a small semantic change to arrive at the ancestor of the Romance languages. It does not follow logically that Proto-Romance cannot be derived directly from the Latin language, albeit not the wordlist. It makes quite the difference whether the shared cognation is to be explained as a semantic shift in a near-synonym, a newly coined term following the rules of derivational morphology or if it is completely new lexical root.

Heggarty et al. (2023a: 3; 2023b: 60, 72) rightfully quote Clackson (2016: 15) for the statements that the Romance spoken languages do not go back to a formal written register, and that many Classical words do not survive into Romance (Clackson & Horrocks 2007: 283). However, the relevance of these facts for phylogenetic purposes is overstated for the following reasons: First, loss is not an unusual lexical development, and even though it is perhaps striking in the core vocabulary, there are phonological and structural explanations for the Romance cases – for instance, the prime example *ōs, ōris* 'mouth' "was an insubstantial word anyway, and once the distinctiveness of vowel length disappeared, it was inconveniently similar in many forms to *os*, genitive *ossis*, 'bone', originally with short [o]" (Herman 2000: 99). Second, loss is, in fact, not very prevalent in the Romance core vocabulary (Dworkin 2016; Söhrman 2016; Chang et al. 2015: 207). Third, although the Romance cognates may not continue the Classical Latin wordlist, the material continues Latin lexicon, but through semantic changes, derivations and compounds (Herman 2000: 97–105; Clackson & Horrocks 2007: 283–4; Clackson 2016: 10–2; Wright 2011: 77–9) – in these cases, the Romance innovations are very similar to the Germanic cognates in the meaning WOMAN (*wijf/wîp* and *vrouw/Frau*) discussed by Heggarty et al. (2023b: 89): Just like some languages generalised an already existing near-synonym over the other in the semantic slot WOMAN, some romance languages did the same (see below). Fourth, and most importantly, when romance shares cognations, it is always with Latin, not with other branches to the exclusion of classical Latin. In the line of though of traditional linguistic phylogenetics, Romance sharing cognates with any other language or branch than Latin would be the only trait that could disprove a direct lineage between them, since either Latin or Romance would have innovated to the exclusion of the other.

6.2.3. The rise of Romance

Heggarty et al. (2023b: 60–2) discuss the salience of different types of discrepancies between the Latin and Romance vocabulary. They distinguish three types where the primary term in Latin is not cognate with the primary term in Romance from one converse type where the primary term in Romance is cognate with the primary term in other branches, but not in Latin. The latter would indeed be very significant for determining the relationship between Latin and Romance. The former three are, however, not very interesting for phylogenetics: 1) a few Romance languages preserve the Latin cognate, whereas most others replace it (archaisms), 2) Latin agrees with Sabellic against Romance (archaism in Latin, innovation in Romance) and 3) Latin has a singleton (unique) cognate against Romance, but Sabellic is uninformative because it does not preserve any form in the proper meaning. It is noteworthy – and adds to the credibility of the model – that it correctly infers the ancestral and innovative states, e.g. for type 3: “the higher probability is that they are retentions in Latin, while alternative cognate sets in Romance were innovated on the ‘spoken’ branch to Proto-Romance”. However, it becomes increasingly clear that this correctness is little more than a lucky punch since the model does not engage with the innovations and their salience themselves when we compare these three subtypes to the contrasting type where Romance shares cognates with other branches than Latin.

These are the highly interesting cases because they posit the potential counterarguments to the hypothesis that Romance directly continues Classical Latin. If Latin or Romance shared innovations with other branches, or if there is absolutely no trait of the Romance ancestral form in Latin. Heggarty et al. (2023b: 61–2), however, only identify two such examples, and they do assign any significance to them: Some Celtic languages and Neapolitan share the continuation of PIE **kok̂s-ah*$_2$- ‘limb, joint’ in the meaning LEG; and Romanian and Brythonic share a descendant of **deu̯k-* ‘to lead’ as CARRY. In both examples there is a low probability of them being innovations in Romance to the exclusion of Latin; but once again, this is based on the overall pattern of the data, not the quality of the arguments. Regarding the first example, Heggarty et al. explain how the Latin form *crux* represents the ancestral cognate (PIE **k̂reu̯H-*, also found in Arm. *srownkʿ* 'leg, shank'), and that the semantic shift of **kok̂s-ah*$_2$- 'limb, joint' to 'leg' in Celtic and Neapolitan (and to 'thigh' in most Romance languages) occurred independently. They also consider the semantic shift from 'carry' to 'bring' of the root **deu̯k-* independent innovations in Brythonic and Romanian. While these claims are reasonable in themselves, they prove just how fragile these arguments are. *Coxa* is indeed attested in Latin, only in the meaning 'hip', and the primary meaning of Old Irish *cos* is 'foot' (*eDIL*: s.v. *cos*) (although it does *also* function as the basic word for 'leg'). Similarly, Middle Welsh *dwc* primarily means 'leads, brings', but it is also used as 'carries' (*GPC*: s.v. *dygaf*).

For these concrete examples, the bulk of evidence has caused them to be analysed according to the general scholarly consensus, but others are not as fortunate. Although it is weakly supported, the study by Heggarty et al. returns an Anatolian-Tocharian clade. In their dataset, however, these branches only share two unique cognations[40]: *h_2*erĝ-* 'white' and *h_1*eg*^{wh}- 'drink'. The former is widely attested in other IE languages, only not as the most basic term for 'white' in these languages, and the latter is considered a lexical archaism.[41]

[40] iecor.clld.org/cognatesets?cladefilter=Anatolian%2CTocharian

[41] It is not possible to reconstruct just a single term for 'white' in Proto-Indo-European, but rather multiple near-synonyms (Kölligan 2018: 2256). This is also evident from the fact that there are 9 "PIE" terms for WHITE in *IE-CoR*, attested in 1-3 branches: **k̂u̯ei̯t-* in Germanic and Indo-Iranian, **b*^{h}*elH-* in Balto-Slavic and 6 more in only one clade each, iecor.clld.org/parameters/white#3/41.56/35.28). But even this is a truth with modifications: Greek *ἀργός* < *h_2*ar̥grós* primarily means 'shining, bright' in Homer, but is attested in the meaning 'white' from Aristotle (*LSJ*: s.v.; Kloekhorst 2007: 307). The latter is considered a lexical archaism. As a verbal root, *h_1*eg*^{wh}- 'drink' is only attested in Anatolian and Tocharian, but it survives in derived adjectives elsewhere: Gr. *νήφων* 'sober' < **n̥-*h_1*g*^{wh}*-on-*, Lat. *ēbrius* 'drunk' (Ringe, Warnow & Taylor 2002: 99; *LIV*^{2}: 231; *LIV*^{3}: 34; Friis 2024: 26). This is usually connected to the semantic innovation of the root **peh*$_3$- which means 'swallow, gulp down' in Anatolian (Kloekhorst 2007: 649), but came to be the standard word for 'drink' in the core-IE languages, pushing *h_1*eg*^{wh}- into the specialised meaning 'drink (alcohol)' before it eventually disappeared (Friis 2024: 26).

It would be highly remarkable, to say the very least, if an inherited cognate had survived in unrecorded registers of Latin only to flourish as the standard term in Romance later. It is, however, not impossible. Maybe Proto-Romance **rŏkka-* 'rock' (Fr. *roche*, Sp. *roca,* It. *rocca*) is such an example. According to the analysis of Garnier (2012: 15), followed by Pellard, Jacques and Ryder (forthc.: 23–24), Proto-Romance **rŏkka-* 'rock' and Latin *rūpēs* 'rock' descend from the same Proto-Italic paradigm, but with independent unique innovations on both sides. Accordingly, the Proto-Romance lexeme (and by extension, language) cannot be derived directly from Classical Latin. The idea is that both Latin *rūpēs* and Romance **rŏkka-* show independent innovation from the ancestral Proto-Italic form **rou̯pí-*. In Latin, the word followed the regular sound laws, but a new paradigm as a plurale tantum was built on the inherited genitive plural *rūpium* < **rou̯pi-i̯om*. Proto-Romance **rŏkka-* has a more complex explanation. The nominative singular of the diminutive **rou̯pika* would become **rṓpika* with a distinct "plebian" monophthongisation of **ou̯* > **ō* (not ***ū*). The genitive plural **rou̯pikā́rum* should have become **rṑpikā́rum* and then **rṑ̀pkárum* by "a sound law whereby unstressed short vowels between two strong syllables are deleted" which assimilated to **rŏkka-* with pretonic or rhythmic shortening of **ṑ* > **ŏ*. This theory is arguably speculative and quite complicated, and the form is, perhaps, not the clearest example of the alleged phenomenon. Nevertheless, Pellard, Jacques & Ryder go on to conclude:

> "While most proto-Romance etyma can be derived from Latin by regular sound correspondences, forms originating from etyma non-attested in Latin like **rŏkka-* 'rock' are not rare, and Classical Latin is thus not the exact ancestor of Romance languages, even though it is close to it."
>
> (Pellard, Ryder & Jacques forthc.: 24)

This must meet some objections. First, it is crucial that such "non-attested" forms are excessively rare in the core vocabulary. Second, it is very rare situation that Romance continues a form which is not derivable from the attested Latin but must go back to before the split of Latin, Romance and Sabellic. In this instance, it would require that this word survived in unrecorded registers or dialects not just of the Roman Empire, Republic, Kingdoms but between the dissolution of Proto-Italic (which Heggarty et al. (2023a: 9) date ca. 1431 BCE) and the first attestation of OFr. *roche* in 747 CE (von Wartburg 1962: 440).

6.3. Locating linguistic prehistory

6.3.1. "Unknown unknowns"

Turning now from the issues of dating linguistic prehistory and especially the problems of Bayesian inference, I will address the issues of locating languages and speakers gerographically in prehistory. It should go without saying that discussions of matters in which most variables are unknown are difficult to keep at a decent scientific level, but that does not mean that they are not important or that they should be avoided, or that people – researchers or not – will abstain from diving headfirst into them. This should in itself make the need for careful consideration and frequent reconsideration of the available evidence and argumentation. Otherwise, what seems to be the most decent proposal at some time ends up being repeated so often that it gets a life of its own as its own truth.

Accordingly, it goes without saying that – no matter the approach – it is difficult to know anything with certainty about the geographic location of languages in prehistory. Determining the linguistic area of an attested ancient language is difficult enough as it is: Does the record of written records match the distribution of the speakers? What dialects, registers and genres did *not* survive? These factors are unknown at present, and most of them will probably remain so. There are plenty more unknown variables to add. For instance, Garett's (1999; 2006) "pruning" theory posits that the Indo-European language family dispersed gradually as one large dialect continuum but only pinnacles of this continuum survived into attestation, leaving their interactions with and influence on the attested languages as "unknown unknowns" (Clackson 2022: 29; Kölligan 2023: 322). These "unknown unknowns" of prehistoric dialectology and ancient geography form a "cocktail effect" of unknowns when the topic is the geographic position of prehistoric languages. Yet another unknown factor is added when the donor language is unknown or reconstructed (e.g. Alexander M. Lubotsky 2001; 2020; Thorsø 2020; Bjørn 2022; Koivulehto 2001; Beekes 2014; cf. also Poulsen 2025a).

There is a high risk of circularity in the argumentation revolving around where prehistoric languages were spoken, and especially whom the speakers interacted with. The most fruitful attempts are tied closely to the lexicon of the languages – borrowings, especially larger groups of loan words falling in semantic spheres – reveal contacts in prehistory. Borrowing events between known languages can be dated relatively according to the relative chronology of sound changes, and sometimes they can be dated more-or-less absolutely according to the archaeolinguistic method (on critique of this, see above).

6.3.2. Real-world consequences

The cocktail effect of unknown unknowns only gets stronger when it is applied to other linguistic data than the lexicon which has some connection to the outside world. Clackson (2022: 28) gives an important critique of models that connect claims of reconstructed prehistoric dialects with their geographical dispersal: They are anachronistic when they plot languages and linguistic traits that might be millennia apart. Of course, we can try to integrate isoglosses on a map or a model, but the "areal" features are inherently circular: A shared isogloss is interpreted as evidence for prehistoric geographic proximity. It could be argued that this is not too different from orthodox cladistics where a shared isogloss is seen as evidence for a prehistoric subgroup – and thus a prehistoric protolanguage. One major difference is that a subgroup can be falsified: Better arguments in favour of an alternative or differing relative chronologies disproving the validity of the isogloss as a shared innovation. Under dialect geography, these differences that are irreconcilable to phylogeneticists only lead to adjustments of the map – or to shifting maps across time to encompass everything.

6.3.3. Short-circuits in dialectology

Similar traits shared isoglosses have often led historical linguists to draw far-reaching conclusions on the basis of slender material. For instance, on the nature of the similarities between Indo-Iranian and Greek, Rix states:

> Diese Übereinstimmungen (Isoglossen) beruhen [...] teils auf Sprachberührungen in vorhistorischer Nachbarschaft, die dem späteren Sprachaustausch zwischen Griechisch und Latein vergleichbar sind
>
> (Rix 1976: 8)

Concerning the same languages, adherents of earlier sophisticated work by dialect geographers (Porzig 1954; Birwé 1956; Euler 1979) who "sifted through large amounts of data to discern which similarities signal retention of an archaic form and which represent, conversely, a shared innovation, which imply genetic relationship and which point, instead, to areal spread" (Drinka 2013a: 388), are not shy from pinning these innovations to maps of prehistory. This paves the way for some very speculative interpretations of the spread of the Indo-Europeans, e.g. Drinka (1993) who finds the homeland of the proto-Indo-Europeans north of the Black Sea on the basis of the development of – among other things – the middle endings from the "stative" or proto-perfect. This data obviously contains no geography, and it is thus pure speculation to assign real-word geography to such a topological model.

Kim (2018a: 247–253) dedicates a paragraph to the rhetorical question if Greek and Armenian are "more than just neighbors". While this section deals with the conflation of the two hypotheses on why Armenian and Greek share certain linguistics traits (did they form a single speech community or were they spoken in close proximity to each other), it does not address when and where this close proximity should have been. It is thus just as circumstantial to posit that as to claim a genetic unity in prehistory.

Along the same vein is the claim that Wakhi belongs to the Khotanese-Tumshuqese dialect area within East Iranian, or that the ancestors of Wakhi and Khotanese were geographically close in prehistory (Emmerick 1989: 216). This based on a single phonological isogloss occurring only in one word, namely the preserved palatal character of the sibilant in PIr. **ć* in **aću̯a-* 'horse' > Khot. *aśśa-* 'id.', Wakhi *yaš* 'id.' as opposed to the other Iranian languages where **ć* loses occlusion and all palatal traits and becomes a fully-fledged dental sibilant /s/ (see further the Preamble of Article 3). While the distance between Khotan and Wakhan is a few hundred kilometres, the areas are still divided by a mountain range – and the languages by at least a millennium. This single isogloss cannot bear the weight of this claim alone.

Similarly, Cheung (2015: 53–61) concludes that Pashto was part of three consecutive Sprachbünde among the East Iranian languages, all based on very spurious evidence – notably very trivial and cross-linguistically common innovations like *d* > *l* or reduction of the clusters *št (> *tt)* > *t*. The few lexical isoglosses – and most importantly the "cultural loans" constitute much more valid evidence for prehistoric geographical proximity.

6.3.4. Phylogenetics and Archaeolinguistic location

Admittedly, also adherents to “orthodox” phylogenetics may also commit the tempting sin of drawing geographical conclusions based on linguistic evidence, such as the connection of satəm and *RUKI* with the spread of corded ware DNA in the ancestors of the Indo-Iranians, Balts and Slavs (Narasimhan et al. 2019: 58; Olander & Poulsen 2022a) which has also been heavily criticised (Heggarty et al. 2023b: 69–71).

Heggarty and colleagues (2023a) do not shy away from making their own claims on the prehistoric migrations of the Indo-Europeans, only they base them on their own topology. At this point, it is important to reiterate that linguistic phylogenetics offer no information on absolute dating. Traditional phylogenetics, relying on evidence from common innovations, only make claims on the relative chronology of changes. Bayesian studies, as described above, infer the date of linguistic speciation events on the basis of the amount of shared etymologically related vocabulary in specific semantic meanings. They then go on to match these findings with results from archaeological and chiefly genetic research. Notably, these matches can only be made on similarities in the dating as the data studied for Bayesian inference is purposefully curated to be free from interference from culture. Accordingly, if the dates of the topology are wrong, placing it on a map becomes absurd.

Just like traditional phylogenetics does not make any claims on the absolute time, the discipline does not *per se* make any claims on the geographical location of the speakers. Adherents of the methods are, however, often keen to jump to these conclusions but instead rely on archaeolinguistic arguments. It is important to stress that the connection of both traditional and Bayesian topologies and the archae-genetic reality is based on the grossly simplified assumption that it is possible to connect the immaterial culture that is prehistoric languages to material culture and artefacts of the physical world. Language is not encoded in DNA or a prerequisite for certain style of pottery, but it is an effective vector for advancing both. While it is a very large assumption, it is not unjustified as a working hypothesis.

The great advantage of archaeolinguistic dating is that it combines language with time and space. If the connection between the reconstructed terminology and the objects or phenomena they signify is indeed correct, it is possible not just to date the invention of a craft, but also to locate it geographically. In archaeology, it is obviously the other way around: An artefact is located before it is dated. For archaeolinguistic, the archaeogenetic dates and locations are inseparable. At this point, it is important stress that the discipline has come a long way since the initial attempts (Olander 2019c). Specifically, increasing awareness of the limitations of how precise lexical semantics are is what has driven the discipline to search for more accurate data. It is unlikely that we will encounter a reiteration of the infamous "birch-problem". However, the archaegentic dates and locations are especially compelling when they are connected to vocabulary which is already sifted through a phylogenetic approach. As elaborated below, there is much to be learned from not just the geographical but also the topological distribution of lexemes or semantic shifts; notably the absence of the word for 'plough' in Anatolian and the semantic shift of the root from which it is derived *h_2erh_3- after the break-off of the clade is indeed suggestive.

6.3.5. Evidence for or compatible with

These examples show that one should be careful not to stretch the conclusions beyond what the evidence can carry. It is important to stress that such results can merely be compatible with each other, but we cannot say anything about correlation. Therefore, it is much preferable to apply a stricter and falsifiable phylogenetic methodology *before* searching for geographic locations and contact edges than to attribute all non-tree like innovations to prehistoric contact (*pace* Nakhleh, Ringe & Warnow 2005) or to return to the "sophisticated work" by dialect geographers. Once more, an isogloss can be disproven as a common innovation in a phylogenetic sense, but prehistorical areal spread cannot be disproven so long as we do not know the details of the migrations and trade routes in prehistory. Sharing a few trivial traits cannot be taken as serious proof of prehistorical migrations – although sharing such traits can indeed be compatible with such a scenario.

While I write this critique from the standpoint of defending the validity and relevance of linguistic phylogeny over prehistoric dialectology, one could also hope for stronger methodological delimitations among (prehistoric) dialect geographers. Afterall, we are all painfully aware that linguistic innovations on macro-level all began as inter-speaker variation in much smaller communities. Sociolinguistics and dialectology supply valuable and absolutely crucial analyses and evidence for research in not just linguistic but also human prehistory. It is neither doing these fields just when phylogeneticists water down terms like "areal spread" or "dialect variation" by applying them to incalcitrant data not meeting the criteria for verifying their own hypotheses; nor when dialect geographers fail to distinguish between degrees of likelihood in their claims on prehistory.

6.4. The whereabouts of "Indo-Greek"

Entertaining the hypothesis, or at least the frequent allegation (see Article 1), that Greek and Indo-Iranian constituted a single clade at a time in prehistory following the dissolution of the Indo-European proto-language, or that they were spoken in such a proximity to each other that they could undertake areal, non-genetic innovations together, it would be necessary explore when and where this common ancestor could have been spoken. This latter scenario is what is concluded by Birwé (1956: 70) and Rix (1976: 8). If they ever formed a (more or less) exclusive clade (e.g. "Ostindogermanisch", Euler 1979: 257), when would this have been? And if the similarities are to be assigned to prehistoric contact, when and where were they neighbours?

6.4.1. Bayesian inference

Answering this question with Bayesian inference leads nowhere since such studies never find a particularly intimate relationship between the two branches, and accordingly, the dating of their latest common ancestor is ancient. Nevertheless, these datings of their latest common ancestor vary greatly: ca. 5300 BCE (Gray & Atkinson 2003: 437), ca. 5200 BCE (Bouckaert et al. 2013), ca. 4000 BCE (Atkinson et al. 2005: 215), ca. 3400 BCE (Heggarty et al. 2023a; 2023b: 78), ca. 3200 BCE (Chang et al. 2015). But bear in mind that in almost all of the resulting topologies of these studies, all branches but Anatolian and Tocharian are included – and in the ones that try to connect the inferred dates with the influx of DNA (which suggests migration in prehistory) (Heggarty et al. 2023a), the proposed populations would not have left the Homeland yet (Kroonen et al. 2023a).

6.4.2. Archeolinguistic *termini ante quos*

Before turning to the age of the latest common ancestor, it is perhaps necessary to state when the languages of both branches are attested. This will give us the *terminus ante quem* of when we certainly know the clade was broken up. In many cases, the *terminus ante quem* is easily settled, but it goes without saying that a reconstructed clade cannot be dated by historic evidence. At any rate, dating these are notoriously difficult – especially for Indo-Iranian. The Greek language was recorded (in the shape of Mycenaean) from ca. 1390 BCE in Knossos on Crete (Bennett 2024: 4; Olivier 2024: 74). The recorded variety is very much Greek – younger than the Proto-Greek we arrive at when reconstructing from comparative evidence including the alphabetic dialects. Naturally, this does not consider who arrived with the language, and when and where the dialects began to diverge.

For Indo-Iranian exact dates are more difficult. Because of the strong oral tradition of the Vedic chants, the earliest Indic documents probably postdate their composition by a millennium. Judging from the fact that the language is significantly older than the earliest dateable documents are the inscriptions of king Aśoka from 3rd century BCE and the grammatical commentary of Paṇini can be dated "no later than the late 5th to early 4th century BCE", an estimate is that the R̥g-Veda dates around 1000 BCE – give or take half a millennium (Cardona 2017: 313).

The attestation of "Indic" names in texts from the Hurrian-speaking Mitanni Kingdom (from ca. 1350 BCE onwards) and "Indic" loanwords (mediated through Hurrian) in the Hittite horse-training manual of Kikkuli strongly suggest that Proto-Indic should be dated even earlier (Daniels 2017: 29). However, the Indic nature of Mitanni-Aryan rest on scanty evidence. It is certainly Indo-Iranian, but the limited linguistic material does not give much opportunity to distinguish between Indic and Iranian phonology. The loanwords predate the monophthongisation of PIIr. **ai* > **ē* otherwise attested in Indic (but reconstructed for the time of the composition of the Vedic hymns and preserved in Iranian). The unhelpful cuneiform script and the Hittite-Hurrian medium makes it difficult to assess what the dialectal stance of the first and second palatalisations represent: PIIr. **ǰ*^{h} is written <š> (which is neither Indic (Vedic) **ǰ*^{h} > *h* nor Iranian **ǰ*^{h} > **ǰ* > **ź* , Av. *z*), and PIIr. **č* (from palatalised PIE **k*^{w} *k*) is written <z>, representing [ts]? (which neither in alignment with Indic or Iranian, as **č* is preserved in both: Ved. *c*, Av. *c* (in Hoffner's transcription, but *č* in Bartholomae's) (later turning to **c* /t͡s/ in most Iranian languages). Further, the voiced aspirates are not distinguished from the voiced stops, at least graphically, which on the surface is more Iranian than Indic: Mitanni *b/papru-nnu-* 'epithet of horses' < **bab*^{h}*ru-nu-*, cf. Ved. *babhrú-* 'reddish-brown'.

However, Mitanni-Indic is lexically more Indic than Iranian: In the word for one, PIE **h*$_{1}$*oi-*, Mitanni *aika-* in *aikau̯artanna* 'one lap (vel sim.)' < **h*$_{1}$*oi̯-ko-* clearly sides with the Indic suffix (_éka-_ < _*ai̯-ka-_ against Iranian _*ai̯-u̯a-_ > Av. _aēuua_, OP _aiva-_ (and "European" _*$h_1$oi-no-_, Gr. _οἴνη_ 'one on a die', Lat. _ūnus_, Germ. _*aina-_) (Lubotsky & Kloekhorst 2021: 331 nn. 8; 334). The theonyms and the onomastic material are also Indic rather than Iranian (Mayrhofer 1982: 76–7; Kroonen, Barjamovic & Peyrot 2018: 11).

The situation is comparable for Iranian. It is estimated that the Avestan hymns were composed before 1000 BCE, but they were not recorded before the Sassanian tradition beginning ca. 600 CE. The Old Persian inscriptions can be dated from Achaemenid royal inscriptions from 522/21 BCE (Skjærvø 2017a: 471–2). All this to say that the latest common ancestor of Indo-Iranian and Greek must have been dissolved _long_ before 1500 BCE, probably much earlier.

6.4.3. Archaeolinguistic _termini post quos_

Turning to archaeolinguistic evidence rather than historical data – that is, the meaning of the words, not just their attestation, brings us closer to an estimate, but not by much. The usual arguments for a much later date of PIE dissolution than those offered by controversial Bayesian inferences are less problematic for "Indo-Greek" – it is, by now, generally accepted that Anatolian split from the rest of the Indo-European language families first, and thus the "Indo-Greek" date would be later than "Proto-Indo-Hittite" anyway. The best archaeolinguistic arguments for a later dating of the dispersal of Proto-Indo-European than what is found by Bayesian inference is evidence for shared vocabulary in the semantic fields of metallurgy, "secondary products", agriculture and chariotry.[42]

6.4.4. Metals

Although Indo-Iranian and Greek do not share specific lexical roots or derivations exclusively, they belong to the core Indo-European branches sharing specific terms for metals, notably silver and cobber or bronze (Thorsø et al. 2023; Olander 2023). In Iranian, YAv. _ərəzata-_ and OP _ardata-_ mean 'silver', and Vedic has the slightly different _rajatá_ 'bright, silver-coloured (R̥V), silver (AV)' < PIIr. _*(H(a))r̥ȷ́ata-_. These correspond to Lat. _argentum_ 'silver', but not exactly to Greek _ἄργυρος_, Myk.

[42] In the following, I have relied on many references from the notes compiled for a paper in preparation (Kroonen et al. in prep.) as an elaboration of (Kroonen et al. 2023a; 2023b).

a-ku-ro ‘silver’ < **h₂(a)r̥ĝ-u-ro-*. Obviously, the forms are derived from the root **h₂erĝ-* ‘shine, be bright’, but as the derivations are not identical, and the semantics of the root definitely allow for independent innovations, as proven by Ved. *árjuna-* ‘light, bright, silver-coloured’ and PIE **h₂r̥ĝ-ró-* > Ved. *r̥jrá-* ‘reddish, shining, bright-coloured; fast, quick’, Gr. *ἀργός* ‘white, bright-shining; fast’ (*NIL*: 316–8). Silver objects are known in Europe, primarily in the Balkans, from ca. 4500 BCE onwards (Radivojević & Roberts 2021: 29). Indo-Iranian, but not Greek, shares a word for a utensil metal, PIIr. **áy̯as-* > Ved. *áyas-*, Av. *aiiah-*, with other IE languages where cognates mean copper, e.g. Lat. *aēs*. Copper smelting occurred in Europe and Caucasus already ca. 5000 BCE (Radivojević & Roberts 2021), but was not prominent in many areas until millennia later.

Naturally, sharing nouns for goods – and naming brightly shiny metals “brightly shining”, does not prove how well the societies knew metallurgy or if they merely traded the goods, but the facts are at least compatible with dating the latest shared ancestor of Indo-Iranian and Greek after the introduction of metallurgy. Note that already these dates are too late for the inferred dates of Gray & Atkinson (2003) and Bouckaert et al. (2013) – not just for Proto-Indo-European, but also for the latest common ancestor of Indo-Iranian and Greek.

6.4.5. Wheels and wheeled vehicles

Another excessively prominent point which has been debated extensively is the invention of the wheel. Most non-Anatolian Indo-European branches share the term **kʷekʷlo-* ‘wheel’ – including Gr. *κύκλος* ‘a wheel, a circle, a place of assembly’ (Hom.), Ved. *cakrá-* ‘wheel, circle; sun-disc’, YAv. *caxra-* ‘wheel’ [43]. This is a linguistically peculiar term, derived as a reduplicated thematic noun from the root **kʷelh₁-* ‘to move, turn, wander; to pasture; to settle and cultivate’. Since wheels do not occur in iconography (as clay figurines and imprinted on clay vessels) before ca. 3400 BCE, and the earliest wooden wheels occur around ca. 3200 BCE (Anthony 2007: 65–72; Bakker et al. 1999; Holm 2019), it requires quite some mental gymnastics to claim that the IE languages independently innovated the term (with

---

[43] A minor complicating factor in Indo-Iranian is that the roots descending from PIE **kʷelh₁-* ‘to move, turn, wander; to pasture; to settle and cultivate’ and **kʷer-* ‘do, make’ coalesce formally in derivations where the laryngeal is dropped: PIIr. **čarH-*; **kar-/čar-*. Cf. OAv. *caxri-* adj. ‘doing, turning (into)’ derived from **kar-* (Uesugi 2024: 73; Cheung 2007: 34, 238).

unproductive morphology), or all branches independently changed the semantics of an already coined term (*pace* Heggarty et al. 2023b: 19–21). What makes it all there more unlikely is that it is not just an ambiguous term **kʷekʷlo-* 'circle, wheel' that can be reconstructed, but an entire semantic field relating to wheels and wagons: **$h_3$nobʰ-* 'nave' (not Greek and Iranian, but Indic (Ved. *nábhya-* 'hub of a wheel') and others), **$h_2$ak̂s-* 'axle' > Ved. *ákṣa-*, Gr. *ἄξων*.

As to the lack of evidence form Anatolian, one should be careful with drawing conclusions based on the lack of evidence (*pace* Ehret 2015: 88). While the lack of terms is certainly compatible with a scenario whereby "Proto-Indo-Hittite" was spoken before the invention of wheel, they do not necessarily "indicate" it – the original terms could have been replaced graphically or linguistically in Anatolian. At least, arguments of this sort should be considered across entire and preferably multiple semantic fields.

6.4.6. Wool

There is also a reconstructible term for 'wool' which is the usual crux for archaeolinguistic dating as the earliest finds of wool are younger than the cultures otherwise associated with the earliest Indo-Europeans. The IE languages agree on a specific term on the surface: Hitt. *ḫulana-* 'wool', Ved. *ūrṇá-* 'id.', Av. *varəna-* 'id.', Gr. *λῆνος* 'id.' etc. all going back to **h_2u̯ĺ̥h_1-nah$_2$-*. However, the earliest preserved wool fibres are dated ca. 3200-2800 BCE – but they are in Iran. Wool does only shows up in the Caucasus or Pontic-Caspian zone much later: ca. 3000-2600 BCE and 2400 BCE, respectively (Viñas-Caron et al. 2024: 34; Shishlina et al. 2020; 2023). This dating is even too young for the Steppe-hypothesis which dates PIE to the Pontic-Caspian steppes and associates the language with the culture complex of Serednij Stih (ca 4400-4000 BCE) and its successors Suvoro (ca 4200-4000 BCE) and Jamnaja (ca. 3300-2500 BCE), and accordingly far too young for the dates implied by Bayesian inference.

This discrepancy has led to some speculation (summed up by Olander (2024b; Kroonen et al. in prep.). Anthony (2007: 269, 285) emphasises the linguistic evidence and goes on to postulate that wool was in existence but has not survived into the archaeological records. While it is true that wool is easily degraded, I prefer to leave archaeological dating to experts and not speculate on the lack of evidence. Hopefully, more research in bioarchaeology will shed more light on the issue – it is currently unknown when the mutation that led to the wooly sheep occurred, there are no nuclear DNA data from wool samples yet and no genetic studies of sheep on the Pontic-Caspian steppes (Viñas-Caron et al. 2024: 42; Kristiansen & Sørensen 2019). These will no doubt shed light on what happened between the domestication of sheep ca. 8500 BCE and the earliest wool samples.

The classic counterargument is that **h*$_2$*u̯ĺ̥nah*$_2$- meant 'un-spinnable wool; fleece or fur from a non-woolly sheep' in PIE (Renfrew 2001: 26) and shifted the meaning to 'wool' independently in all branches following the dissolution of PIE and the introduction of the woolly sheep. However, this not only goes against the attested semantics and economy of reconstruction, but also the derivational pattern reconstructed by Olsen (2023; 2018b; 2017): The existence of the Latin *s*-stem *vellus* 'wool shorn from a sheep, fleece' and Arm. *gełmn* 'fleece' as well as the verb Lat. *vellō* 'pluck (feathers, hair)' < **h*$_2$*u̯elh*$_1$- 'to pluck'[44] implies that they can all be connected: **h*$_2$*u̯elh*$_1$- → **h*$_2$*u̯ĺ̥h*$_1$*-men-* → **h*$_2$*u̯ĺ̥h*$_1$*-mn-ah*$_2$-. One would have to ask why the speakers who coined the terms would have "plucked" these sheep and referred to the "pluckings" if said sheep were hairy, not woolly – and why the term is not applied to other animal skins.

It has also been pointed out that there are formal problems with deriving the Anatolian and Greek forms from **h*$_2$*u̯ĺ̥h*$_1$- (Pronk 2021: 158–61; Kloekhorst 2023). I find it attractive to consider a separation of the Hittite "Wortsippe" (*ḫulan-* 'wool', *ḫulii̯a-* 'wool' and *ḫulāli-* 'distaff') seemingly from **h*$_2$*ul-* without a final laryngeal (to account for the lack of gemination) from **u̯lh*$_1$- without the initial laryngeal for "core-Indo-European" to account the lack of a prothetic vowel in Gr. *λῆνος* (although exact parallels are rare). The suggestion that *ḫul-* is ultimately a Hurrian loan (through Akkadian) in Hittite (Kronasser 1967: 45) is attractive for Akk. *ḫullānu* which seems to be derived with a Hurrian suffix. But we should keep in mind that **ḫul(l)-* is not actually attested in Hurrian (Richter 2012: 161), and that Akk. *ḫullānu* means 'a piece of clothing made of wool' in texts from (the Hurrian-influenced cities) Nuzi and Alalah. While a *ḫullanu* is certainly made of wool, the word is consistently spelled with *-ll-* in Akkadian (von Soden 1965: 354) making it a less obvious donor candidate.

[44] Note that the *LIV*^{2} (679) connects *vellō* to **u̯elh*$_3$- (or *u̯elh*$_2$-) 'strike'.

Finally, as Olander (2024) explains, taking the consequence of the dating of wool would imply that PIE can hardly be archaeolinguistically dated before 3000 BCE. This is less of a problem for the latest common ancestor of Indo-Iranian and Greek, and we can _at least_ date "Indo-Greek" _*(h$_2$)u̯l̥h$_1$-nah$_2$-_ 'wool' to the period in which wool existed – while waiting for bioarchaeological data to settle the matter in Proto-Indo-European proper.

6.4.7. Agriculture

Another controversy for the dating of Indo-European is the lack of agricultural terminology. As is well-known, the Anatolian Hypothesis posits that the Indo-European languages spread with agriculture from the fertile crescent. While it has been backed by the dates conjured by Bayesian inference, it is difficult to reconcile with the lack of evidence elsewhere, as Ehret points out:

> Not one single root word of unambiguous reference to even one of the ancient crops of Middle Eastern agriculture can be reconstructed with certainty to the proto-Indo-European language.
>
> (Ehret 2015: 89)

While I essentially agree with the point, it is yet another interesting case of a topic where negative evidence is easily conflated with lack of evidence. I find it much more compelling the Anatolian hypothesis is incompatible with the positive evidence for the existence of terms denoting artefacts that had not yet come into being, as Ehret (2015: 89) adds below the quote above:

> The linguistic testimony puts it beyond reasonable doubt that the far-flung expansions of Indo-European languages began no earlier than the fourth millennium and that they emanated outwards from the lands along the north of the Black Sea.

What actively speaks against a dating or a clade is much more compelling than absence of evidence of it. From a purely linguistic perspective, it is possible to imagine the early speakers of Indo-European were familiar with agriculture but went through a subsequent period of pastoralist economy. It would only take a few generations to lose knowledge of the craft and the specialised terminology associated with it. However, such a scenario is purely a thought experiment and not necessarily compatible with any archaeological reality.

Most core-IE languages share a term for 'plough': An instrument noun *h_2arh_3-tro-* derived from the root *h_2erh_3-* 'to plough'. It is found in all branches – but Anatolian, Albanian and Indo-Iranian: TB *āre*, Gr. *ἄροτρον*, Arm. *arawr*, Lat. *arātrum*, Lith. *ar̃klas*, ON *arðr* etc. Contrary to the example of the word for 'wheel', the verbal root *h_2erh_3-* is attested in Anatolian, but the instrument noun is not. Importantly, the descendant Hitt. *ḫarra-*^{i} does not mean 'plough', but rather 'crush, grind' (the Hittite verb for 'tilling (soil' is *ḫārš-*^{i} which seems related but might be a West Semitic loan (Kroonen et al. 2022: 12)). This has been taken to suggest that the meaning of the root shifted after the split-off of Anatolian, and that this root is direct evidence of the evolution from 'Hackbau' to 'Ackerbau' in agriculture (Kloekhorst 2007: 7). While this sounds almost too good to be true, it is indeed a stronger argument than pure lack of evidence.

Until recently (Van Willigen et al. 2024), the earliest evidence for ploughing in Europe stemmed from in the form of "races of criss-cross plough-marks" from the late 4000-3000 BCE (Sherratt 1983: 169). Depictions of the plough are known from Akkadian cylinder seals dated 2300 BCE. A "proto-plough" made from deer-antler is known from the Cucuteni-Trypillja culture in modern-day Ukraine at the end of the 6th millenium BCE (Kroonen et al. 2022: 12). Oxen bones from the Trypillian culture also show pathologies associated with hard labour, presumably animal traction (Paškevič & Videjko 2006: 129). Surviving wooden ploughs are only attested much later. It requires quite the historic coincidence for such marks to be preserved – they are coincidentally preserved under Middle Neolithic (ca. 3500 BCE) burial mounds. However, the following "prophecy" may have come true:

> […] we cannot be sure that animal traction was not harnessed earlier than this, and the means by which earlier gardens were cultivated is likewise unknown
>
> (Whittle 2015: 580)

It seems now that the earliest plough marks go all the way back to between 5100 and 4700 BCE – in Switzerland (Van Willigen et al. 2024). Although it would be uneconomical, we cannot rule out that *h_2arh_3-tro-* originally referred to some sort of pole or hoe used for 'crushing, grinding' (*h_2erh_3-*) the ground of a bed of sorts (which is essentially 'tilling'), and only shifted to designate the 'plough' used for 'ploughing' (*h_2erh_3-) a field after the introduction of the plough – independently in most branches. While one could perhaps expect the arrival of new technology to be accompanied by new terminology (like much later in medieval Europe where terms derived from late Lat. *ploxenum* and *plōstrum* largely replaced the descendants of *h_2erh_3trom with the spread of the moulboard plough). However, there is no guarantee the speakers would not simply have repurposed their existing word and apply it to the new tool instead of losing the word along with the obsolete technology.

That said, agricultural terminology is a major divide between Indo-Iranian and Greek. As mentioned above, Indo-Iranian does not attest a cognate of the word for plough. Instead, Vedic has *matyà-* 'some kind of agricultural tool, stick' maybe related to Lat. *mateola* 'mallet, beetle' and OCS *motyka* 'hoe' (Mallory & Adams 2006: 242–3; Mayrhofer 1996: 297). Indo-Iranian does not even attest the verbal root, found in Gr. *ἀρόω* (*LIV*[2]: 272). It may of course have lost some vocabulary, probably even this root, but strikingly Indo-Iranian seems to take up a semantic middle ground between the archaic branches Anatolian and Tocharian and the European branches when it comes to innovating specifically agricultural terminology (Kroonen et al. 2022: 31–2). For example, *$g^w r̥h_2$-u̯on-* becomes TB *kärweñe** 'stone', Ved. *grā́van-* 'pressing stone' but OCS *žrъny* 'millstone'. Accordingly, it seems that

> [...] Indo-Iranian participated in the initial core Indo-European shift from a pastoralist to an agro-pastoralist economy, of which some elements later were lost. On the other hand, Indo-Iranian was peripheral to the more recent and more radical shift towards a farming economy, as reflected in the vocabularies of the European branches.
>
> (Kroonen et al. 2022: 32)

6.4.8. Horses and chariotry

Notably, Indic and Iranian share a remarkable terminology related to horses (Kroonen, Barjamovic & Peyrot 2018; Kroonen 2021). Of course, being able to talk about horses in intricate detail does not automatically imply that the speakers knew domesticated horses – and if they did, language alone cannot tell us if they kept horses for meat consumption, riding and/or as draught animals. However, the existence of the nuanced terminology and personal names or epithets such as Ved. *ṛjrā́śva-*, YAv. *ərəzrāspa-* 'whose horses are swift (or shining?)' and Ved. *yuktā́śva-*, YAv. *yuxtāspa-* 'whose horses are yoked' that can be reconstructed of their common ancestor, PIIr. **(H)ṛ́jra(H)aću̯a-* and **i̯ukta-(H)aću̯a-*, strongly suggests an intimate cultural relationship with domesticated (yoked) horses in the Indo-Iranian Proto-Language. Although, to be fair, there is no way to exclude that these formations are not parallel – but the amount of them makes this extremely unlikely. Remarkably, Indo-Iranian has shared vocabulary related to horses uniquely with Balto-Slavic (e.g. **u̯olo-* 'horse hair, tail (?)', Ved. *vā́ra-*, Lith. *vãlas*) – and not with Greek (Kroonen 2021).

We can reconstruct nuanced and specific terms for chariotry to Proto-Indo-Iranian (Lubotsky 2023: 259). For the chariot and the driver, we can reconstruct **(H)ratHa-* 'battle chariot'[45] (Ved. *rátha-*, YAv. *raϑa-*, Khot. *rraha-*), **(H)ratHiH-* 'charioteer' (Ved. *rathī́-*, OAv. *raiϑī-*), **(H)ratHai-štaH-* (Ved. *rathe-ṣṭhā́-*, YAv. *raϑaē-štā-*) 'id.' (litt. 'chariot-standing'). The action of driving a chariot ran be reconstructed with the verb **Hi̯aH-* (Ved. *yā-* 'to drive', seen in Av. *yāman-* 'course'). There is further **Haua-saHana-* 'unharnessing of horses, resting place' (Ved. *avasā́na-*, OAv. *auuaŋhāna-*) and a number of specialised terms like **(H)raćanaH-* 'bridle' (Ved. *raśanā́-* 'cord, bridle', MP (Pahl.) *lsn* /rasan/ 'rope') and **(H)i̯auktra-* 'cord' (Ved. *yóktra-* 'thong, yoking cord', YAv. *-yaoxəδra-* 'halter, bridle'). According to Lubotsky (2023: 259), the term **Habʰi-dʰaHana-* 'bridle' (Ved. *abhidhā́nī-* 'horse halter', YAv. *zaraniiō.aiβiδāna-* 'with a golden bridle')[46] and the related compound verb **Habʰi-dʰaH-* 'to halt horses' (< *'to put against') seems to refer exclusively to horses. This mirrors the semantic situation of the agricultural terminology treated above where Indo-Iranian usually preserves a polysemous situation whereby general terms had acquired specialised meaning as well but not yet lost the former. This latter term suggests a further step from the starting point of domestication.

[45] Related to another PIE word for 'wheel', **(H)róth₂-* e.g. Lat. *rota*, Lith *rãtas* with a similar semantic shift as in Tocharian where **kʷekʷlo-* 'wheel' became TB *kokale*, TA *kukäl* 'cart'.

[46] Continued wiedely in East Iranian: Sogd. (Buddh.) *βδ''nh*, *βyδ'n* 'bridle', Khot. *byāna-* 'id.', Chor. *'βz'n-* 'id.', Pashto *mlúna* 'id.', Yidgha *awlān* 'id.', Sarikoli *viδun* 'id.', Yazgulami *avδén* 'bridle and bit').

The ancestors of modern horses were domesticated and became geographically widespread from ca. 2200 BCE (Librado et al. 2021a; 2021b), and chariots do not occur before ca. 2020 BCE (Chechushkov & Epimakhov 2023: 14). Taking these facts together strongly suggests that the ancestors of the Indo-Iranians and Greeks no longer formed a clade at this time.

6.4.9. Evidence from genetics and archaeology

Another line of evidence often used to locate the ancestors of the speakers of reconstructed ancestral languages in prehistory comes from archaeogenetic evidence. A word of caution is important here. Just like sherds and pottery do not speak, neither do genes. That does not mean that the evidence cannot be very compelling and compatible with linguistic scenarios, but as has been reiterated time and time again, languages can spread without genetic contact, and genes can spread without linguistic influence.

That said, it is now generally accepted that the Indo-Iranians can be traced to a chain of migrations of cultures feeding into each other. Proto-Indo-Iranian is broadly associated with the Sintašta-Petrovka culture (BCE 2100–1900 BCE, Southern Trans-Urals) "which was very compact in time and space" and was succeeded by the Andronovo culture (Lubotsky 2023: 259; Kuz'mina & Mallory 2007). Sintašta is a continuation of the Abaševo culture which dissolved from Fatyanovo-Balanovo ca. 2200 BCE (Nordqvist & Heyd 2020). This link is highly important since the Fatyanovo-Balanovo culture is located north of the Black Sea, part of the broader Corded Ware horizon – and genetically linked to these, most famously the introduction of the Y chromosome haplogroup R1a into the population. This incompatible with the grossly outdated "Out of India"-hypothesis (Hock 1999) and the "southern route" scenario whereby the ancestors of the Indo-Iranians never set foot on the Pontic-Caspian steppes (Heggarty et al. 2023a: 2).

Conversely, it is much more difficult to trace the migrations of the Greeks. There is a continuous influx of cultures and genes from the North into the Balkans. However, the Mycenaeans are "genetically quite distinct" from the Minoans and Cycladic cultures. Steppe-related ancestry does not occur in Greece before the Mycenaean period (Lazaridis et al. 2017; Clemente et al. 2021). This genetic influx is dated 2600-2000 BCE. Notably, the Steppe admixture is very small compared to many other parts of Europe and there is no sex bias unlike in Northern Europe. The Greeks are among the people who inherited only R1b Y-chromosome haplogroup and have no corded-ware associated admixture (Yediay et al. 2024; Lazaridis et al. 2025; Solans 2024).

6.4.10. Summary

The question of the nature of the linguistic relationship between Indo-Iranian and Greek cannot be settled by data from archaeology and genetics. However, these disciplines supply the scaffolding for human migrations of the past. It is clear that we cannot point to a point in time and space where the Greeks and Indo-Iranians formed an exclusive clade.

Archaeolinguistic evidence related to charioteering point to a date of Proto-Indo-Iranian no earlier than 2000 BCE. The same date has been suggested for a while (Trager & Smith 1950: 62). This vocabulary is not shared with any other branches, but there are some horse-related terms are shared with Balto-Slavic to the exclusion of Greek. Genetically, the Indo-Iranians can be linked to the migrations Sintašta (ca 2000 BCE) and Andronovo cultures (ca. 1900 BCE). Given that the Mycenean culture came into being around 1750 BCE (Linear B being recorded from 1400 BCE), it is no surprise that their ancestors were nowhere close to the Indo-Iranians at this time.

On the one hand, it seems that Indo-Iranian did not take part in the complete shift from a primarily pastoral (herder-gatherer) economy to an agro-pastoralist (herder-farmer) economy, which Greek did. On the other hand, we can link the Greeks genetically to a continuous influx of steppe-related genes from 2600-2000 BCE who did not show signs of R1a (corded ware) admixture. Conversely, while the Indo-Iranians seems to have been peripheral at the time of the introduction of ploughing and agriculture, which Greek took part in, their ancestors were part of the Corded Ware horizon and did bring Steppe ancestry with R1a admixture into Central Asia (Fatyanovo broke from Corded Ware ca. 2800 BCE, and Fatyanovo dissolved into Abaševo ca. 2200 BCE which turned to Sintašta by 2100 BCE).

Accordingly, Greek and Indo-Iranian were hardly part of the same speech community much after 3000-2800 BCE – and the divide could easily be much older. Additionally, it is very difficult to point to a point in time and space were the two branches could undergo areally induced innovations together (*pace* Birwé 1956; Rix 1976) to the exclusion of other branches – it is, however, impossible to exclude, but the burden of proof should be on those advocating for this geographic proximity.

7. Concluding remarks

Sparked by the curiosity of the frequent allegations of a special, almost mythical, relationship between the two most well-attested and researched branches of Indo-European, the present thesis has entertained the methodology of higher-order linguistic subgrouping and the relationship between Indo-Iranian and Greek in various ways.

Chapter 2 deals with the methodology of linguistic subgrouping, and the limitations of traditional and quantitative methods. With traditional phylogenetic methods, one will easily be led stray by the many tempting connections based on positive evidence from Indo-Iranian and Greek but absence of evidence in all other branches. From a quantitative point of view, one can only base the analysis on the least salient type of data in order to include all comparable branches on equal terms.

Chapter 3, 4 and 5 deal with the suggested isoglosses and the reconstruction of the latest common ancestor of Indo-Iranian and Greek. I can conclude that there is very little *indentifiable* or *informative* evidence in favour of an exclusive Indo-Iranian–Greek clade. Since the two branches attest much more material – and are generally more conservative than many of their potential close relatives, they take up an enormous space in the reconstruction of Proto-Indo-European. There are virtually no phonological innovations which they could share. Laryngeal aspiration and the vocalisation of nasals must be younger than their latest common ancestor. In morphology, the two branches share an enormous amount of material – but rarely to the exclusion of others, and almost never to the exclusion of *positive evidence* from other branches. In the noun, they might share the introduction of *-*oi̯*- into the oblique dual, but this is doubtful. They might share the introduction or expansion of the particle *-*u̯e* in the dual and plural first and second person pronoun, but it could equally well be an archaism or parallel innovations from a poorly understood earlier state. In the verb, they continue – and expand – the tense-aspect system of post-Anatolian Indo-European. Their shared features are mostly not unique (the augment, thematic optative, the mediopassive endings) or likely archaisms (unmarked preterit).

The fixation of the augment, the incorporation of contrastive *-*tero*- as a comparative, the creation of the perfect middle and the pluperfect are parallel innovations, which are more likely due to system-internal pressure than contact, although they are compatible with both.

Chapter 5 examined the archaeolinguistic consequences of their latest common ancestor. This endeavour was also negative. Through its cultural vocabulary of farming, Greek is tied to the European branches, whereas Indo-Iranian seems isolated in this regard. But in terms of genetics, the ancestors of the migrations that brought Indo-European speakers to Central Asia, share Y Chromosome haplogroups with the Corded Ware peoples of North-Western Europe, whereas the ancestors of the Greeks are isolated in this regard.

Article 1 dealt in detail with the terminology "Indo-Greek" and the phylogenetic claims made in published tree topologies. It reveals that the "Indo-Greek hypothesis" never existed in a (narrow) phylogenetic sense.

Article 2 is a methodological survey of the role of loanwords in phylogenetics from traditional and computational points of view as a way to tackle the overwhelming overrepresentation of lexico-derivational material in these branches. It examined two case studies of alleged borrowings into a common ancestor. It disproves – again – that "Indo-Greek" **pelek̂us-* 'axe' can be borrowed from a semitic source because of word formation, and that **(H)a(i̯)ĝ-* 'goat' only looks like its Caucasian connections on paper. It concludes that every step in the analyses can have major implications on the results, no matter the approach.

Article 3 is an application of a computational phylogenetic Maximum Parsimony analysis on a group of languages usually considered a dialect continuum or Sprachbund, the Middle and Modern East Iranian languages. While the resulting topologies are not credible, the paper offers practical insights that can further the incorporation of typological and morphological directionality in datasets of phonological and grammatical isoglosses.

Of course, further research is needed to examine many of the individual problems raised in this thesis. An attractive next step would be a systematic reanalysis of the lexico-derivational material collected by Euler (1979), the *NIL* and the *LIV*. Having shed some light on the dangers of drawing conclusions on the basis of *unidentifiable* or *uninformative* data, a phylogenetically informed re-reconstruction of Indo-European derivational morphology is another milestone for the field.

8. References

Abaev, Vasilij Ivanovič. 1973. *Istoriko-ėtimologičeskij slovar' osetiskogo jazyka*. Vol. 2. L-P. Moskva & Leningrad: Nauka.

Adams, Douglas Q. & James P. Mallory. 1997. Wool. In James P. Mallory & Douglas Q. Adams (eds.), *Encyclopedia of Indo-European culture*, 648–649. London & Chicago: Fitzroy Dearborn.

Adams, Douglas Quentin. 2013. *A dictionary of Tocharian B*. 2nd ed. (Leiden Studies in Indo-European 10). Amsterdam & New York (NY): Rodopi.

Adams, Douglas Quentin. 2023. Review of Olander (ed.) 2022. *Journal of Indo-European Studies* 51(1–2). 216–236.

Allentoft, Morten E., Martin Sikora, Karl-Göran Sjögren, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522(7555). 167–172. doi:10.1038/nature14507.

Ameis, Friedrich Karl & Karl Hentze. 1884. *Homers Odyssee. Für den schulgebrauch erklärt*. 7th ed. Vol. 2.1. Gesang XIII-XVIII. Leipzig: B. G. Teubner.

Andersen, Henning. 1968. Indo-European **s* after *i*, *u*, *r*, *k* in Baltic and Slavic. *Acta Linguistica Hafniensia* 11(2). 171–190.

Anthony, David. 1990. Anthony 1990 Migration in archaeology: The baby and the bathwater. *American Anthropologist* 92(4). 895–914.

Anthony, David. 2017. Archaeology and Language: Why Archaeologists Care about the Indo-European Problem. In Pam J. Crabtree, Peter I. Bogucki & Bernard Wailes (eds.), *European archaeology as anthropology: Essays in memory of Bernard Wailes*. Philadelphia (PA): University of Pennsylvania Museum of Archaeology and Anthropology.

Anthony, David W. 1991. The archaeology of Indo-European origins. *Journal of Indo-European Studies* 19(3–4). 193–222.

Anthony, David W. 2007. *The horse, the wheel, and language: how bronze-age riders from the Eurasian steppes shaped the modern world*. Princeton, N.J: Princeton University Press.

Anthony, David W. & Donald A. Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1. 199–219. doi:10.1146/annurev-linguist-030514-124812.

Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam: John Benjamins. (19 November, 2021).

Atkinson, Quentin D., Geoff Nicholls, David Welch & Russell Gray. 2005. From words to dates: Water into wine, mathemagic or phylogenetic inference. *Transactions of the Philological Society* 103(2). 193–219. doi:10.1111/j.1467-968X.2005.00151.x.

Bacovcin, Hezekiah Akiva & David Wilson. 2018. A New Method for Computational Cladistics: An Afro-Asiatic Case Study. *Transactions of the Philological Society* 116(3). 410–434. doi:10.1111/1467-968X.12128. DOI: 10.1111/1467-968X.12128.

Bakker, Jan Albert, Janusz Kruk, Albert E. Lanting & Sarunas Milisauskas. 1999. The earliest evidence of wheeled vehicles in Europe and the Near East. *Antiquity* 73. 778–790.

Bartholomae, Christian. 1904. *Altiranisches Wörterbuch*. Strassburg: Trübner.

Bartoněk, Antonín. 2003. *Handbuch des mykenischen Griechisch*. (Indogermanische Bibliothek. Erste Reihe). Heidelberg: C. Winter.

Batisti, Roberto. 2020. On Greek *Αἰθίοψ* "Ethiopian" and *Αἴσωπος* "Aesop" from a PIE Perspective. *Proceedings of the 31st Annual UCLA Indo-European Conference. November 8th and 9th, 2019*, 37–53. Hamburg: Buske. (24 October, 2024).

van Beek, Lucien Christiaan. 2022a. Greek. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 173–201. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.011 (23 September, 2022).

van Beek, Lucien Christiaan. 2022b. The Reflexes of Syllabic Liquids in Ancient Greek: Linguistic Prehistory of the Greek Dialects and Homeric *Kunstsprache*. *The Reflexes of Syllabic Liquids in Ancient Greek*. Leiden: Brill. https://brill.com/display/title/60857 (8 October, 2024).

Beekes, Robert S. P. 2010. *Etymological dictionary of Greek*. (Leiden Indo-European Etymological Dictionary Series). Leiden & Boston: Brill.

Beekes, Robert Stephen Paul. 2010. *Etymological dictionary of Greek*. (Leiden Indo-European Etymological Dictionary Series 10). Leiden & Boston (MA): Brill.

Beekes, Robert Stephen Paul. 2011. *Comparative Indo-European linguistics: An introduction*. 2nd ed. Amsterdam & Philadelphia (PA): John Benjamins.

Beekes, Robert Stephen Paul. 2014. *Pre-Greek: Phonology, Morphology, Lexicon*. (Ed.) Stefan Norbruis. Leiden: Brill.

Bendahman, Jadwiga. 1993. *Der reduplizierte Aorist in den indogermanischen Sprachen*. (Deutsche Hochschulschriften 642). Egelsbach: Hänsel-Hohenhausen.

Bennett, John. 2024. The archaeological and historical context. In John Killen (ed.), *The New Documents in Mycenaean Greek*, vol. 1. Introductory Essays, 3–24. Cambridge & New York (NY): Cambridge University Press. doi:10.1017/9781139029049.

Bernabé, Alberto & Eugenio R. Luján. 2020. *Introducción al griego micénico: Gramática, selección de textos y glosarios*. 2ª edición, corregida y aumentada. (Monografías de Filología Griega 30). Zaragoza: Prensas Universitarias de Zaragoza.

Bezzenberger, Adalbert. 1890. Die indogermanischen Gutturalreihen. *Beiträge zur Kunde der indogermanischen Sprachen* 16. 234–260.

Bird, Norman. 1982. *The distribution of Indo-European root morphemes: A checklist for philologists*. Wiesbaden: Harrassowitz.

Bird, Norman. 1993. *The roots and non-roots of Indo-European: A lexicostatistical survey*. Wiesbaden: Harrassowitz.

Birwé, Robert. 1956. *Griechisch-Arische Sprachbeziehungen im Verbalsystem*. (Beiträge zur Sprach- und Kulturgeschichte des Orients 8). Walldorf Hessen: Verlag für Orientkunde Dr. H. Vorndran.

Bjørn, Rasmus G. 2022. The lexicon of an Old European Afro-Asiatic language: Evidence from early loanwords in Proto-Indo-European. *Historische Sprachforschung* 135. 3–42.

Bonfante, Giulano. 1976. *I dialetti indoeuropei*. (Studi Grammaticali e Linguistici 12). Brescia: Paidaia.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337. 957–960. doi:10.1126/science.1219669.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, et al. 2013. Corrections and Clarifications. *Science* 342. 1446. doi:10.1126/science.1219669.

Brugmann, Karl. 1886. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. 2. Bearbeitung, 1897. Photomechanischer Nachdruck, Berlin, de Gruyter. Vol. 1. Einleitung und Lautlehre. Strassburg: Trübner.

Budruss, Georg. 1977. Nochmals zur Stellung der Nūristān-Sparchen des Afghanischen Hindukush. *Nachtrag zu MSS 37* 38. 19–38.

Burrow, Thomas. 1972. A Reconsideration of Fortunatov's Law. *Bulletin of the School of Oriental and African Studies, University of London* 35(3). 531–545.

Byrd, Andrew Miles. 2015. The Indo-European syllable. (Brill's Studies in Indo-European Languages & Linguistics volume 15). Leiden: Brill.

Campbell, Lyle. 2020. *Historical linguistics: An introduction*. 4th ed. Edinburgh: Edinburgh University Press.

Cantera, Alberto. 2017. The phonology of Iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 481–503. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter.

Cardona, George. 2017. The documentation of Indic. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 309–325. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter.

Carling, Gerd. 2024. Archaeolinguistics. Words, artefacts, and ancient DNA. *Linguistic Archaeology: An Introduction and methodological guide*, 124–137. New York: Routledge. DOI: 10.4324197810032916647.

Cathcart, Chundra. 2011. RUKI in the Nuristani languages: an assessment. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 22nd Annual UCLA Indo-European Conference, Los Angeles, November 5th and 6th, 2010*, 1–11. Bremen: Hempen.

Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1). 194–244. doi:10.1353/lan.2015.0005.

Chantraine, Pierre. 2009. *Dictionnaire étymologique de la langue grecque: Histoire des mots*. (Ed.) Jean Taillardat, Alain Blanc, Olivier Masson & Jean Louis Perpillou. 2ème ed. (Librairie Klincksieck. Série linguistique 20). Paris: Klincksieck.

Chechushkov, Igor V. & Andrey Epimakhov. 2023. Relative and Absolute Chronologies of the Chariot Complex in Northern Eurasia and Early Indo-European Migrations. In Eske Willerslev, Guus Kroonen & Kristian

Kristiansen (eds.), *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics*, 247–256. Cambridge: Cambridge University Press. doi:10.1017/9781009261753.020. https://www.cambridge.org/core/books/indoeuropean-puzzle-revisited/relative-and-absolute-chronologies-of-the-chariot-complex-in-northern-eurasia-and-early-indoeuropean-migrations/D89F5636E78BB048786528EA0FDA4683 (22 March, 2025).

Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb*. (Leiden Indo-European Etymological Dictionary Series 2). Leiden: Brill.

Cheung, Johnny. 2015. The linguistic and geographic position of Pashto within the east iranian language group: 1. A phonological and lexical analysis. *Iranian Languages and Literatures of Central Asia. From the Eighteenth Century to the Present*, 29–65. (Cahiers de Studia Iranica 57). Leuven: Peeters.

Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. (Publications of the Philological Society 30). Oxford & Cambridge (MA): Blackwell.

Clackson, James. 2007a. The genesis of Greek. In Anastasios-Phoibos Christidis (ed.), *A history of Ancient Greek: From the beginnings to Late Antiquity*. Cambridge: Cambridge University Press.

Clackson, James. 2007b. *Indo-European linguistics: An introduction*. Cambridge: Cambridge University Press.

Clackson, James. 2013. The origins of the Indic languages: The Indo-European model. In Angela Marcantonio & Girish Nath Jha (eds.), *Perspectives on the origin of Indian civilization*, 260–288. Center for India Studies, University of Massachusetts, Dartmouth.

Clackson, James. 2016. Latin as a source for the Romance languages. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 3–13. (Oxford Guides to the World's Languages). Oxford: Oxford University Press.

Clackson, James. 2022. Methodology in Linguistic Subgrouping. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 18–32. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.002 (23 September, 2022).

Clackson, James & Geoffrey C. Horrocks. 2007. *The Blackwell history of the Latin language*. Malden (MA) & Oxford: Blackwell.

Clayton, John. 2022. Labiovelar loss and the rounding of syllabic liquids in Indo-Iranian. *Indo-European Linguistics* 10(1). 33–87. doi:10.1163/22125892-bja10021.

Clemente, Florian, Martina Unterländer, Olga Dolgova, et al. 2021. The genomic history of the Aegean palatial civilizations. *Cell* 184(10). 2565-2586.e21. doi:10.1016/j.cell.2021.03.039.

Cowgill, Warren. 1960. Greek *ou* and Armenian *očʿ*. *Language* 38. 347–50.

Damaspji, Destur Hoshengji & Martin Haug. 1867. *An old Zand-Pahlavi glossary: Edited in the original characters with transliteration in Roman letters, an English translation and an alphabetical index*. Bombay, London & Stuttgart: Trübner.

Daniels, Peter T. 2017. The writing systems of Indo-European. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 26–61. De Gruyter.

De Decker, Filip. 2011a. Stang's Law and the Indo-European word for 'cow.' *Indogermanische Forschungen* 116. 42–59.

De Decker, Filip. 2011b. Evidence for laryngeal aspiration in Greek? Part I: The "recent" evidence. *Indogermanische Forschungen* 116. 87–109.

De Decker, Filip. 2015. What is the Greek counterpart of (Proto-)Indo-Iranian (*)th? *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 12. 89–164.

Debrunner, Albert. 1954. *Altindische Grammatik*. Vol. 2/2. Die Nominalsuffixe. Göttingen: Vandenhoeck & Ruprecht.

Degener, Almuth. 2002. The Nuristani languages. In Nicholas Sims-Williams (ed.), *Indo-Iranian languages and peoples*, 103–118. (Proceedings of the British Academy 116). Oxford: Oxford University Press.

Dehò, Delia. 1957. Isoglosse lessicali greco-sanscrite di origine origine indeuropea. *Rendiconti dell'Istituto Lombardo di Scienze e Lettere, Classe di Lettere* 91. 343–380.

Del Freo, Maurizio. 2016. La scrittura lineare B. *Manuale di epigrafia micenea: introduzione allo studio dei testi in lineare B*, 123–166. Prima edizione. Padova: Libreriauniversitaria.it edizioni.

Drinka, Bridget. 1993. The dispersion of Indo-European dialects: Clues from morphology. *WORD* 44(3). 413–429. doi:10.1080/00437956.1993.11435910.

Drinka, Bridget. 2009. Stratified reconstruction and a new view of the family tree model. In Angela Marcantonio (ed.), *The Indo-European language family: Questions about its status*, 7:1-39. (Journal of Indo-European Studies, Monograph Series). Washington (DC): Institute for the Study of Man.

Drinka, Bridget. 2013a. Phylogenetic and areal models of Indo-European relatedness: The role of contact in reconstruction. *Journal of Language Relationship • Вопросы языкового родства* 6(2). 379–410. doi:10.1163/19552629-00602009.

Drinka, Bridget. 2013b. Evolutionary models and language contact in Indo-European.

Duvick, Brian Marshall. 2007. *On Plato Cratylus*. (Ed.) Harold Tarrant. (Ancient Commentators on Aristotle). London & Ithaca (NY): Bloomsbury Academic. doi:10.5040/9781472552112.

Dworkin, Stephen N. 2016. Lexical stability and shared lexicon. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 577–587. (Oxford Guides to the World's Languages). Oxford: Oxford university press.

Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indoeuropean classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society* 82. iii–iv, 1–132. doi:10.2307/1006517.

Ėdel'man, Džoj Iosisovna. 1999. *Dardskie i nuristanskie jazyki*. (Jazyki Mira 8). Moskva: Indrik.

eDiAna = Rieken, Elisabeth, Jared Miller, Olav Hackstein, et al. 2022. Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages. https://www.ediana.gwi.uni-muenchen.de/index.php (25 October, 2024).

eDIL = Toner, Gregory, Máire Ní Mhaonaigh, Sharon Arbuthnot, et al. 2019. An electronic dictionary of the Irish language, based on the *Contributions to a Dictionary of the Irish Language* (Dublin: Royal Irish Academy, 1913-1976). https://dil.ie/.

Ehret, Christopher. 2011. *History and the testimony of language*. (The California World History Library 16). Berkeley (CA): University of California Press.

Ehret, Christopher. 2015. Agricultural origins: What linguistic evidence reveals. In Graeme Barker & C. Goucher (eds.), *The Cambridge world history*, 55–92. (2. A world with agriculture). Cambridge: Cambridge University Press.

Eichner, Heiner. 1973. Die Etymologie von heth. *mehur*. *Münchener Studien zur Sprachwissenschaft* 31(1972). 53–107.

Emmerick, Ronald E. 1989. Khotanese and Tumshuqese. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 204–229. Wiesbaden: Reichert.

Euler, Wolfram. 1979. *Indoiranisch-griechische Gemeinsamkeiten der Nominalbildung und deren indogermanische Grundlagen*. (Innsbrucker Beiträge Zur Sprachwissenschaft B30). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Euler, Wolfram. 2010. Der Schwund des Duals in der Flexion indogermanischer Einzelsprachen. *Studia Etymologica Cracoviensia* 2010(Volume 15, Issue 1). 77–111.

Euler, Wolfram. 2017. Perfektpartizipien des Aktivs und Mediums im Griechischen und Indoiranischen: Zur Frage von Redundanzen. *Wékʷos: Revue d'études indo-européennes* 3. 35–56.

Fellner, Hannes A. 2021. Polar life in the Tarim Basin. In Hannes A. Fellner, Melanie Malzahn, Michaël Peyrot & Georges-Jean Pinault (eds.), lyuke wmer ra*: Indo-European studies in honor of Georges-Jean Pinault*, 125–137. Ann Arbor & New York (NY): Beech Stave.

Fortson, Benjamin W. 2004. *Indo-European language and culture: An introduction*. (Blackwell Textbooks in Linguistics). Malden, MA & Oxford: Blackwell.

Fortson, Benjamin Wynn, IV. 2020. Never say never: A new proposal for Armenian *očʿ*. *Die Sprache* 53(1). 65–84.

Francis, Eric D. 1970. Greek disyllabic roots: The aorist formations. Yale University.

Friis, Louise Skydsbjerg. 2021. Tocharian B agent nouns in *-ntsa* and their origin. *Indo-European Linguistics* 9(1). Brill. 1–25. doi:10.1163/22125892-bja10012.

Friis, Louise Skydsbjerg. 2024. Studies in Tocharian verbal morphology relevant to the cladistic position of tocharian in Indo-European. Leiden: Universiteit Leiden Ph.D. dissertation. https://hdl.handle.net/1887/4093377 (3 October, 2024).

Friis, Louise Skydsbjerg. fortch. The thematic optative in Tocharian. In Michiel Arnoud Cor de Vaan & Sergio Neri (eds.), *the proceedings of the XVII. Fachtagung der Indogermanischen Gesellschaft (eds. Sergio Nero & Michiel de Vaan)*. in review.

Fritz, Matthias & Michael Meier-Brügger. 2021. *Indogermanische Sprachwissenschaft*. De Gruyter. doi:10.1515/9783110661767. https://www.degruyter.com/document/doi/10.1515/9783110661767/html (28 April, 2023).

García Ramón, José Luiz. 2016. Il greco miceneo. *Manuale di epigrafia micenea: introduzione allo studio dei testi in lineare B*, vol. 1, 211–246. Prima edizione. Padova: Libreriauniversitaria.it edizioni.

García Ramón, José Luiz. 2017. The morphology of Greek. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 2, 654–682. (41). De Gruyter.

Garett, Andrew. 1999. A new model of Indo-European subgrouping and dispersal. In Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the twentyfifth annual meeting of the Berkeley linguistics Society, February 12–15, 146–156.*, 146–156. Berkeley (CA): Berkeley Linguistics Society.

Garett, 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. (McDonald Institute Monographs). Cambridge, Oxford & Oakville (CT): McDonald Institute for Archaeological Research, University of Cambridge.

Garnier, Romain. 2012. Allomorphisme et lois de limitation rythmique en latin. *Bulletin de la Société de Linguistique de Paris* 107(1). 235–259.

Garnier, Romain. 2014. Italique commun **né⸗χe⸗ai̯ud* “jamais de la vie.” *Wékʷos: Revue d’études indo-européennes*(1). 93–109.

Garrett, Andrew. 1999. A new model of Indo-European subgrouping and dispersal. In Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society, February 12–15*, 146–156. Berkeley: Berkeley Linguistics Society.

Gąsiorowski, Piotr. 1999. The tree of language: A cladistic look at the genetic classification of languages. *Dialectologia et Geolinguistica* 7. 1–20. doi:10.1515/dig.1999.1999.7.39.

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

http://lingulist.de/documents/geisler-list-2013-do-languages-grow-on-trees.pdf (26 March, 2025).

Gercenberg, L. G. 1981. Xotansakskij jazyk. *Osnovy iranskogo jazykoznanija: Sredneiranskie jazyki*, 233–313. Moskva: Nauka.

Gimbutas, Marija. 1952. On the Origin of North Indo-Europeans. *American Anthropologist* 54(4). 602–611. doi:10.1525/aa.1952.54.4.02a00430.

Gimbutas, Marija. 1970. Proto-Indo-European culture: The Kurgan culture during the fifth, fourth, and third millennia B.C. In George Cardona, Henry M. Hoenigswald & Alfred Senn (eds.), *Indo-European and Indo-Europeans*, 155–198. Philadelphia: University of Pennsylvania Press.

Gippert, Jost. 1986. Zur Metrik der Gathas. *Die Sprache* 32. 257–275.

Gippert, Jost. 1997. Laryngeals and Vedic metre. In Alexander M. Lubotsky (ed.), *Sound law and analogy: Papers in honor of Robert S. P. Beekes on the occasion of his 60th birthday*, 63–79. Rodopi.

Goldstein, David M. 2020. Indo-European phylogenetics with R. *Indo-European Linguistics* 8(1). https://brill-com.ep.fjernadgang.kb.dk/view/journals/ieul/8/1/article-p110_3.xml (23 September, 2022).

Goldstein, David M. 2022. There's no escaping phylogenetics. In Laura Grestenberger, Charles Reiss, Hannes A. Fellner & Gabriel Z. Pantillon (eds.), *Ha! Linguistic studies in honor of Mark R. Hale*, 71–91. Wiesbaden: Reichert.

Gotō, Toshifumi. 2013. *Old Indo-aryan Morphology and its Indo-iranian Background: In cooperation with Jared S. Klein und Velizar Sadovski*. Wien: Verlag der Österreichischen Akademie der Wissenschaften. doi:10.2307/j.ctt1vw0r7r (16 February, 2024).

Gotō, Toshifumi. 2017. The morphology of Indic (old Indo-Aryan). In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 344–377. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 41). De Gruyter. (30 March, 2025).

GPC = Thomas, R. J., Gareth A. Bevan & Patrcik J. Donovan. 1987. *Geiriadur Prifysgol Cymru. A dictionary of the Welsh language.* (Ed.) Bwrdd Gwybodau Celtaidd Prifysgol Cymru. Caerdydd: Gwasg Prifysgol Cymru.

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–439. doi:10.1038/nature02029.

Greenhill, Simon J. & Russell D. Gray. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica* 29(4). 523–537. doi:10.1075/dia.29.4.05gre.

Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2021. Bayesian Phylolinguistics. In Richard D. Janda, Joseph, Brian D. & Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, 226–253. John Wiley & Sons. doi:10.1002/9781118732168.ch11 (28 September, 2021).

Grestenberger, Laura. 2024. Sound Change and Analogy, Again: Brugmann's Law and the Hunt For O-Grades in Indo-Iranian. *Transactions of the Philological Society* 122(1). 79–118. doi:10.1111/1467-968X.12272.

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522(7555). 207–211. doi:10.1038/nature14317.

Hackstein, Olav. 1993. On the prehistory of dual inflection in the Tocharian verb. *Die Sprache* 35(1). 47–70.

Hajnal, Ivo. 1995. Review of Volkart 1994. *Historische Sprachforschung* 108. 308–316.

Hajnal, Ivo. 1997. Definite nominale Determination im Indogermanischen. *Indogermanische Forschungen (1997)* 102(1997). 38–73. doi:10.1515/9783110243437.38.

Hale, Mark. 2007. *Historical linguistics: Theory and method*. Malden, MA, & Oxford: Blackwell.

Hamp, Eric P. 2013. The expansion of the Indo-European languages: An Indo-Europeanist's evolving view (with annotation and comments by Douglas Q. Adams). *Sino-Platonic Papers* 239. (1–14).

Hansen, Bjarne Simmelkjær Sandgaard & Guus Kroonen. 2022. Germanic. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 152–172. Cambridge, UK, & New York: Cambridge University Press.

Harrison, S. P. 1986. On the nature of subgrouping arguments. In Paul Geraghty, Lois Carrington & S. A. Wurm (eds.), *FOCAL II: Papers from the fourth international conference on austroneasian linguistics*, 13–21. (Pacific

Linguistics. Series C 94). Canberra: Department of Linguistics, Research School of the Pacific Languages, The Australian National University.

Hegedűs, Irén. 2012. The RUKI-rule in Nuristani. In Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen & Jens Elmegård Rasmussen (eds.), *The sound of Indo-European: Phonetics, phonemics and morphophonemics*, 143–167. (Copenhagen Studies in Indo-European 4). Copenhagen: Museum Tusculanum.

Heggarty, Paul. 2021. Cognacy databases and phylogenetic research on Indo-European. *Annual Review of Linguistics* 7. 371–394. doi:10.1146/annurev-linguistics-011619-030507.

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, et al. 2023a. Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* 381(6656). eabg0818. doi:10.1126/science.abg0818.

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, et al. 2023b. Supplementary materials for "Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages." *Science* 381(6656). eabg0818. doi:10.1126/science.abg0818.

Heine, Bernd & Henry Honken. 2010. The Kx'a family: A new Khoisan genealogy. *Journal of Asian and African Studies* 79. 5–36.

Heine, Bernd & Christa König. 2015. *The !Xun language: A dialect grammar of Northern Khoisan*. (Quellen Zur Khoisan-Forschung 33). Köln: Rüdiger Köppe.

Herman, József. 2000. *Vulgar Latin*. (Trans.) Roger Wright. University Park (PA): Pennsylvania State University Press.

Heubeck, Alfred & Arie Hoekstra. 1990. *A commentary on Homer's Odyssey*. Vol. 2. Books IX-XVI. (Clarendon Paperbacks). Oxford: Clarendon Press.

Hock, Hans Henrich. 1986. *Principles of historical linguistics*. 2. rev. and updated ed. Berlin & New York: Mouton de Gruyter.

Hock, Hans Henrich. 1988. *Principles of Historical Linguistics*. *Principles of Historical Linguistics*. 2nd ed. Berlin, New York (NY) & Amsterdam: Mouton de Gruyter. doi:10.1515/9783110871975.

Hock, Hans Henrich. 1991. Dialects, diglossia and diachronic phonology in Early Indo-Aryan. In William G. Boltz & Michael C. Shapiro (eds.), *Studies in the Historical Phonology of Asian Languages*. Amsterdam: Benjamins.

http://ebookcentral.proquest.com/lib/kbdk/detail.action?docID=793572 (20 February, 2024).

Hock, Hans Henrich. 1999. Out of India? The linguistic evidence. In Johannes Bronkhorst & Madhav Deshpande (eds.), *Aryan and non-Aryan in South Asia: Evidence, interpretation, and ideology: proceedings of the International Seminar on Aryan and Non-Aryan in South Asia, University of Michigan, Ann Arbor, 25-27 October 1996*, 1–19. (Harvard Oriental Series. Opera Minora 3). Cambridge (MA) & Columbia (MO): Harvard University, Department of Sanskrit and Indian Studies & South Asia Books.

Hoenigswald, Henry M. 1966. Criteria for the subgrouping of languages. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects: Proceedings of the Conference on Indo-European Linguistics, held at the University of California, Los Angeles, April 25-27, 1963*, 1–12. Berkeley (CA) & Los Angeles (CA): University of California Press.

Hoffmann, Karl & Bernhard Forssman. 1996. *Avestische Laut- und Flexionslehre*. (Innsbrucker Beiträge Zur Sprachwissenschaft 84). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Holm, Hans J. 2008. The distribution of data in word lists and its impact on the subgrouping of languages. In Christine Preisach, Hans Burkhardt, Lars Schmidt-Thieme & Reinhold Decker (eds.), *Data analysis, machine learning and applications: Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7–9, 2007*, 629–636. Berlin & Heidelberg: Springer.

Holm, Hans J. J. G. 2019. *The earliest wheel finds: Their archaeology and Indo-European terminology in time and space, and early migrations around the Caucasus*. (Archaeolingua). Budapest: Archaeolingua Alapítvány.

Hyllested, Adam. 2004. Greek *λωτός* "lotus" and the Indo-European word for "blue." In James Clackson & Birgit Anette Olsen (eds.), *Indo-European word formation: Proceedings of the conference held at the University of Copenhagen, October 20th– 22nd 2000*, 59–64. (Copenhagen Studies in Indo-European 2). Copenhagen: Museum Tusculanum.

Hyllested, Adam & Brian D. Joseph. 2022. Albanian. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 223–245. Cambridge: Cambridge University Press. (7 November, 2022).

IE-CoR = Heggarty, Paul, Cormac Anderson & Matthew Scarborough (eds.). 2023a. *IE-CoR*. Zenodo. doi:10.5281/zenodo.8089434. https://doi.org/10.5281/zenodo.8089434 (18 March, 2025).

IE-CoR = Heggarty, Paul, Cormac Anderson & Matthew Joseph Charles Scarborough. 2022b. Indo-European Cognate Relationships database project (IE-CoR). Zenodo. doi:10.5281/zenodo.8089434. https://zenodo.org/record/8089434 (28 July, 2023). Leipzig: Max Planck Institute for Evolutionary Anthropology. Supplementary materials of Heggarty et al. (2023).

Insler, Stanley. 1987. The Vedic causative type jāpáyati. In Calvert Watkins (ed.), *Studies in Memory of Warren Cowgill (1929-1985): Papers from the Fourth East Coast Indo-European Conference Cornell University, June 6-9, 1985*, 54–65. (Untersuchungen Zur Indogermanischen Sprach- Und Kulturwissenschaft, Neue Folge 3). De Gruyter. doi:10.1515/9783110853223.54. https://www.degruyter.com/document/doi/10.1515/9783110853223.54/html (9 October, 2024).

Jacques, Guillaume. 2011. A panchronic study of aspirated fricatives, with new evidence from Pumi. *Lingua* 121(9). 1518–1538. doi:10.1016/j.lingua.2011.04.003.

Jacques, Guillaume. 2019. Sanskrit *hŕ̥d* "cœr." *Wékʷos: Revue d'études indo-européennes* 4. 85–90.

Jacques, Guillaume & Johann-Mattis List. 2019. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1). 128–167. doi:10.1075/jhl.17008.mat.

Jasanoff, Jay H. 1997a. An Italo-Celtic isogloss: the 3 pl. mediopassive in **-ntro*. In Douglas Quentin Adams (ed.), *Festschrift for Eric P. Hamp. 1.*, 146–161. (Journal of Indo-European Studies Monograph Series 23). Washington (DC): Institute for the Study of Man.

Jasanoff, Jay H. 1997b. Gathic Avestan *cikōitərəš*. In Alexander Lubotsky (ed.), *Sound law and analogy: Papers in honor of Robert S.P. Beekes on the occasion of his 60th birthday*. (Leiden Studies in Indo-European 9). Amsterdam & Atlanta (GA): Rodopi.

Jasanoff, Jay H. 2003. *Hittite and the Indo-European verb*. Oxford & New York: Oxford University Press.

Jasanoff, Jay H. 2009. *-*bhi*, *-*bhis*, *-*ōis*: following the trail of the PIE instrumental plural. In Jens Elmegård Rasmussen & Thomas Olander (eds.), *Internal reconstruction in Indo-European: Methods, results, and problems: Section papers from the XVI International Conference on Historical Linguistics, University of Copenhagen, 11th–15th August, 2003*, 137–149. (Copenhagen Studies in Indo-European). Copenhagen: Museum Tusculanum.

Jasanoff, Jay H. 2018. Palatable Thorns. In Dieter Gunkel, Stephanie W. Jamison, Angelo Mercado & Kazuhiko Yoshida (eds.), *Vina diem celebrent: studies in linguistics and philology in honor of Brent Vine*, 133–140. Ann Arbor (MI) & New York (NY): Beech Stave Press.

Jørgensen, Anders Richardt. 2022. Celtic. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 135–151. Cambridge: Cambridge University Press. doi:10.1017/9781108758666. https://www.cambridge.org/core/books/indoeuropean-language-family/4B44B5ACF0D3BBA89B9408050F112A52 (23 September, 2022).

Joseph, Brian D. 2005. Some Ancient Shared Metaphors in the Balkans. *Studia Albanica* 42. 45–48.

Kalyan, Siva & Alexandre François. 2019. When the waves meet the trees: A response to Jacques and List. *Journal of Historical Linguistics* 9(1). 168–177. doi:10.1075/jhl.18019.kal.

Kassian, Alexei S. 2011. Annotated Swadesh wordlists for the Greek group (Indo-European family). https://starlingdb.org/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ier\grk&limit=-1.

Kassian, Alexei S., George Starostin, Anna Vladimirovna Dybo & Vasiliy Chernov. 2010. The Swadesh wordlist: An attempt at semantic specification. *Journal of Language Relationship • Вопросы языкового родства* 4. 46–89.

Kassian, Alexei S., Mikhail Zhivlov, George Starostin, et al. 2021a. Rapid radiation of the inner Indo-European languages: an advanced approach to Indo-European lexicostatistics. *Linguistics* 59(4). 949–979. doi:10.1515/ling-2020-0060.

Kassian, Alexei S., Mikhail Zhivlov, George Starostin, et al. 2021b. Supplementary materials of “Rapid radiation of the Inner Indo-European languages: an advanced approach to Indo-European lexicostatistics.” Zenodo.

doi:10.5281/zenodo.4046607. https://zenodo.org/record/4046607 (30 August, 2022).

Kaufman, Terrence & John Justeson. 2009. Historical Linguistics and Pre-Columbian Mesoamerica. *Ancient Mesoamerica* 20(2). Cambridge University Press. 221–231.

Keydana, Götz. 2018. The syntax of Proto-Indo-European. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, vol. 3, 2195–2228. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). Berlin & Boston (MA): De Gruyter Mouton.

Killen, John T. & John Bennett. 2024. Finished products 1: Vessels and furniture. In John Killen (ed.), *The New Documents in Mycenaean Greek*, vol. 2, 758–796. Cambridge & New York (NY): Cambridge University Press. doi:10.1017/9781139029049.

Kim, Ronald I. 2018a. Greco-Armenian: The persistence of a myth. *Indogermanische Forschungen* 123. 247–272. doi:10.1515/if-2018-0009.

Kim, Ronald I. 2018b. The phonology of Balto-Slavic. In Jared S. Klein, Jared Stephen & Matthias Fritz (eds.), trans. Joseph, Brian D., *Handbook of comparative and historical Indo-European linguistics*, vol. 3, 1974–1985. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). Berlin & Boston (MA): De Gruyter Mouton.

Kimball, Sara E. 2017. The phonology of Anatolian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, vol. 1, 249–256. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science Band 41). Berlin & Boston (MA): De Gruyter Mouton.

Kiparsky, Paul. 1973. On comparative linguistics: The case of Grassmann's law. In Thomas Sebeok (ed.), *Current trends in linguistics*, 115–134. The Hague: Mouton.

Klein, Jared S. 1988. Proto-Indo-European **$g^w iH_3$-* 'live' and related problems of laryngeals in Greek. In Alfred Bammesberger (ed.), *Die Laryngaltheorie und die Rekonstruktion des Indogermanischen Laut- und Formensystems*, 257–279. Heidelberg: Winter.

Klingenschmitt, Gert. 1968. Farhang-i Ōīm. Edition und Kommentar. Erlangen: Friedrich-Alexander-Universität Erlangen-Nürnberg Inaugural-Dissertation.

Klingenschmitt, Gert. 1975. Altindisch *śáśvat-*. *Münchener Studien zur Sprachwissenschaft* 33. 67–78.

Klingenschmitt, Gert. 1994. Die Verwandtschaftsverhältnisse der indogermanischen Sprachen. In Jens Elmegård Rasmussen & Benedicte Nielsen (eds.), *In honorem Holger Pedersen: Kolloquium der Indogermanischen Gesellschaft vom 25.–28. März 1993 in Kopenhagen*, 235–251. Wiesbaden: Reichert.

Klingenschmitt, Gert. 2005. Das Tokharische in indogermanischer Sicht. In Michael Janda, Rosemarie Lühr, Joachim Matzinger & Stefan Schaffner (eds.), *Aufsätze zur Indogermanistik*, 353–435. (Philologia. Sprachwissenschaftliche Forschungsergebnisse 74). Hamburg: Kovač.

Kloekhorst, Alwin. 2004. The preservation of $*h_1$ in Hieroglyphic Luwian: Two separate *a*-Signs. *Historische Sprachforschung* 117(1). 26–49.

Kloekhorst, Alwin. 2006. Initial laryngeals in Anatolian. *Historische Sprach-forschung* 119. 77–108.

Kloekhorst, Alwin. 2007. The Hittite inherited lexicon. Leiden: University of Leiden.

Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. (Leiden Indo-European Etymological Dictionary Series 5). Leiden & Boston (MA): Brill.

Kloekhorst, Alwin. 2011. The accentuation of the PIE word for 'daughter.' In Tijmen Pronk & Rick Derksen (eds.), *Accent Matters. Papers on Balto-Slavic Accentology*, 235–243. (Studies in Slavic and General Linguistics 37). Leiden: Brill.

Kloekhorst, Alwin. 2014. Proto-Indo-European "thorn"-clusters. *Historical Linguistics* 127(1). 43–67. doi:10.13109/hisp.2014.127.1.43.

Kloekhorst, Alwin. 2016. The Anatolian stop system and the Indo-Hittite hypothesis. *Indogermanische Forschungen* 121(1). 213–248. doi:10.1515/if-2016-0013.

Kloekhorst, Alwin. 2018. Anatolian evidence suggests that the Indo-European laryngeals $*h_2$ and $*h_3$ were uvular stops. *Indo-European Linguistics* 6. 69–94. doi:10.1163/22125892-00601003.

Kloekhorst, Alwin. 2019. The etymology of Hieroglyphic Luwian *izi(ya)-di* 'to do, to make': An athematic *i*-present in Anatolian. In Ignasi-Xavier Adiego, José Virgilio García Trabazo, Mariona Vernet, et al. (eds.), *Luwic dialects and Anatolian: Inheritence and diffusion*. (Barcina Monographica Orientalia 12. Series Anatolica et Indogermanica 1). Barcelona: Edicions de la Universitat de Barcelona.

Kloekhorst, Alwin. 2022. Anatolian. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 63–82. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.005. https://www.cambridge.org/core/books/indoeuropean-language-family/anatolian/82F6A46F99BB62BB588D3EFC130292CE (14 August, 2023).

Kloekhorst, Alwin. 2023. Wool and the Indo-Anatolian Hypothesis. Paper presented at the Tools, Trades, and Technologies – Exploring Prehistory Through Archaeolinguistics 5-6 October 2023, Uppsala. LAMP – Languages and Myths of Prehistory and the Centre for Studies in Indo-European Language and Culture at Stockholm University, Swedish Collegium for Advanced Study, Uppsala. https://www.academia.edu/123968955/Wool_and_the_Indo_Anatolian_Hypothesis (21 March, 2025).

Kobayashi, Masato. 2004. *Historical phonology of Old Indo-Aryan consonants*. (Study of Languages Anc Cultures of Asia and Africa Monograph Series 42). Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Kobayashi, Masato. 2017. The phonology of Indic. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 325–344. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter.

Köhler, Oswin. 1981. La langue !xũ. In Albert Valdman & Gabriel Manessy (eds.), *Les Langues dans le monde ancien et moderne*, vol. 1-2. Les Langues de l'Afrique Subsaharienne & Pidgins et créoles, 557–615. Paris: Editions du Centre national de la Recherche Scientifique.

Koivulehto, Jorma. 2001. The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. In Christian Carpelan, Asko Parpola &

Petteri Koskikallio (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archeological considerations: Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki, 8–10 January, 1999*, 235–263. (Suomalais-Ugrilaisen Seuran Toimituksia / Mémoires de La Société Finno-Ougrienne). Helsinki: Suomalais-Ugrilainen Seura.

Kölligan, Daniel. 2018. The lexicon of Proto-Indo-European. *Handbook of Comparative and Historical Indo-European Linguistics*, 2229–2279. De Gruyter Mouton. doi:10.1515/9783110542431-045. https://www.degruyter.com/document/doi/10.1515/9783110542431-045/html (30 September, 2023).

Kölligan, Daniel. 2023. Review of Olander (ed.) (2022). *Folia Linguistica* 57(s44–s1). 317–338. doi:10.1515/flin-2023-2032.

König, Christa. 2008. Khoisan Languages. *Language and Linguistics Compass* 2(5). 996–1012. doi:10.1111/j.1749-818X.2008.00084.x.

Korn, Agnes. 2016. A partial tree of Central Iranian: A new look at Iranian subphyla. *Indogermanische Forschungen* 121(1). 401–434. doi:10.1515/if-2016-0021.

Kortlandt, Frederik. 1981a. Glottalic Consonants in Sindhi and Proto-Indo-European. *Indo-Iranian Journal* 23(1). 15–19.

Kortlandt, Frederik. 1981b. 1st sg. middle *-H_2. *Indogermanische Forschungen* 86(1981). De Gruyter. 123–136. doi:10.1515/9783110243284.123.

Kortlandt, Frederik. 1983. Notes on Armenian historical phonology, III: *h-*. *Studia Caucasica* 5. 9–16.

Kortlandt, Frederik. 1985. Proto-Indo-European glottalic stops: The comparative evidence. *Folia Linguistica Historica* 6(2). 183–201.

Kortlandt, Frederik. 2016. Phrygian between Greek and Armenian. *Linguistique Balkanique / Балканско езикознание* 55(2). 249–255.

Kortlandt, Frederik. 2018. Proto-Indo-European glottalic consonants. *Münchener Studien zur Sprachwissenschaft* 71. 147–160.

Kortlandt, Frederik Henri Herman. 1983. Archaic ablaut patterns in the Vedic verb. In George Cardona & N. H. Zide (eds.), *Festschrift for Henry Hoenigswald*, 219–222. Tübingen: Narr.

Kortlandt, Frederik Herman Henri. 1978. I.-E. palatovelars before resonants in Balto-Slavic. In Jacek Fisiak (ed.), *Recent developments in historical* *phonology*. (Trends in Linguistics. Studies and Monographs 4). The Hague, Paris and New York (NY): Mouton de Gruyter.

Kortlandt, Frederik Herman Henri. 1985. Notes on Armenian historical phonology 4. *Studia Caucasica* 6. 9–11.

Kristiansen, Kristian & Marie Louise Stig Sørensen. 2019. Wool in the Bronze Age: Concluding Reflections. In Serena Sabatini & Bergerbrant (eds.), *The Textile Revolution in Bronze Age Europe*, 317. Cambridge: Cambridge University Press. https://www.academia.edu/41587128/WOOL_IN_THE_BRONZE_AGE_CONCLUDING_REFLECTIONS (22 March, 2025).

Kristoffersen, Tore Rovs. 2019. Monophthongization of *u*-diphthong before labial consonant in Greek. Copenhagen: Københavns Universitet (University of Copenhagen) MA thesis. University of Copenhagen MA-thesis: https://www.academia.edu/43045272/Monophthongization_of_u_diphthong_before_labial_consonant_in_Greek.

Kristoffersen, Tore Rovs. forthc. a. Monophthongization of *-Vu̯-* before Labial in Greek. Ammendation of Kristoffersen (2019).

Kristoffersen, Tore Rovs. forthc. b. The fat pasture: Greek *Βοιωτία* and Vedic *gávyūti*.

Kronasser, Heinz. 1967. Hith. In Wolfgang Meid (ed.), *Beiträge zur Indogermanistik und Keltologie: Julius Pokorny zum 80. Geburtstag gewidmet*, 45–49. (Innsbrucker Beiträge Zur Kulturwissenschaft 13). Innsbrucker: Institut für Sprachwissenschaft der Universität Innsbruck.

Kroonen, Guus. 2011. False exceptions to Winter's Law: On the effects of Kluge's law on the Proto-Germanic consonantism. In Tijmen Pronk & Rick Derksen (eds.), *Accent matters: Papers on Balto-Slavic accentology*, 251–261. (Studies in Slavic and General Linguistics). Amsterdam & New York: Rodopi.

Kroonen, Guus. 2021. Supplementary Discussion 1. Linguistic index of Indo-European equine and Indo-Iranian chariotry terminology. In Librado et al. (2021b). 38–39.

Kroonen, Guus, Gojko Barjamovic & Michaël Peyrot. 2018. Linguistic supplement to Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian. doi:10.5281/zenodo.1240524. https://zenodo.org/record/1240524 (14 August, 2023).

Kroonen, Guus, Anthony Jakob, Axel I. Palmér, et al. 2022. Indo-European cereal terminology suggests a Northwest Pontic homeland for the core Indo-European languages. *PLOS ONE* 17(10). e0275744. doi:10.1371/journal.pone.0275744.

Kroonen, Guus Jan. 2013. *Etymological dictionary of Proto-Germanic*. (Leiden Indo-European Etymological Dictionary Series 11). Leiden & Boston (MA): Brill.

Kroonen, Guus, Thomas Olander, Mikkel Nørtoft, et al. 2023a. Archaeolinguistic anachronisms in the Indo-European phylogeny of Heggarty et al. 2023 [eLetter]. Science. https://www.science.org/doi/10.1126/science.abg0818#elettersSection.

Kroonen, Guus, Thomas Olander, Mikkel Nørtoft, et al. 2023b. Archaeolinguistic anachronisms in Heggarty et al. 2023. Zenodo. doi:10.5281/zenodo.10113208. https://doi.org/10.5281/zenodo.10113208.

Kroonen, Guus, Thomas Olander, Mikkel Nørtoft, et al. in prep. [preliminary title:] Archaeolinguistic anachronisms in the Indo-European phylogeny of Heggarty et al. 2023. Antiquity [debate].

Kulikov, Leonid Igorevič. 2012. *The Vedic -ya-presents: passives and intransitivity in Old Indo-Aryan*. (Leiden Studies in Indo-European 19). Amsterdam & New York (NY): Rodopi.

Kümmel, Martin Joachim. 1996. *Stativ und Passivaorist im Indoiranischen*. (Historische Sprachforschung. Ergänzungsheft 39). Göttingen: Vandenhoeck & Ruprecht.

Kümmel, Martin Joachim. 2000. *Das Perfekt im Indoiranischen: eine Untersuchung der Form und Funktion einer ererbten Kategorie des Verbums und ihrer Weiterentwicklung in den altindoiranischen Sprachen*. Wiesbaden: Reichert.

Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.

Kümmel, Martin Joachim. 2012. Typology and reconstruction: The consonants and vowels of Proto-Indo-European. In Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen & Jens Elmegård Rasmussen (eds.), *The sound of Indo-European: Phonetics, phonemics and morphophonemics*, 143–167. (Copenhagen Studies in Indo-European 4). Copenhagen: Museum Tusculanum.

Kümmel, Martin Joachim. 2018a. The survival of laryngeals in Iranian. In Lucien Christiaan van Beek, Guus Jan Kroonen, Tijmen Pronk, et al. (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*, 162–171. Ann Arbor (MI) & New York (NY): Beech Stave Press.

Kümmel, Martin Joachim. 2018b. The morphology of Indo-Iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 3, 1888–1924. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter Mouton. doi:10.1515/9783110542431-032.

Kümmel, Martin Joachim. 2020. The development of the perfect within IE verbal systems: An overview. In Robert Crellin & Thomas Jügel (eds.), *Perfects in Indo-European and beyond*, 15–48. (Current Issues in Linguistic Theory 352). Amsterdam: John Benjamins Publishing Company. doi:10.1075/cilt.352.02kum.

Kümmel, Martin Joachim. 2022. Indo-Iranian. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 246–268. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.014. https://www.cambridge.org/core/books/indoeuropean-language-family/indoiranian/C01D66B2DA24680E0808EFEC10E389C9 (23 September, 2022).

Kuz'mina, Elena E. & James Patrick Mallory. 2007. *The origin of the Indo-Iranians*. (Leiden Indo-European Etymological Dictionary Series 3). Leiden & Boston (MA): Brill.

Lazaridis, Iosif, Songül Alpaslan-Roodenberg, Ayşe Acar, et al. 2022. The genetic history of the southern arc: A bridge between west asia and europe. *Science* 377(6609). eabm4247. doi:10.1126/science.abm4247.

Lazaridis, Iosif, Alissa Mittnik, Nick Patterson, et al. 2017. Genetic origins of the Minoans and Mycenaeans. *Nature* 548(7666). 214–218. doi:10.1038/nature23310.

Lazaridis, Iosif, Nick Patterson, David Anthony, et al. 2025. The genetic origin of the Indo-Europeans. *Nature* 639(8053). Nature Publishing Group. 132–142. doi:10.1038/s41586-024-08531-5.

Leskien, August. 1876. *Die Declination im Slavisch-Lithauischen und Germanischen*. Leipzig: S. Hirzel.

Librado, Pablo, Naveed Khan, Antoine Fages, et al. 2021a. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 598(7882). 634–640. doi:10.1038/s41586-021-04018-9.

Librado, Pablo, Naveed Khan, Antoine Fages, et al. 2021b. Supplementary information to Librado et al. (2021). *Nature* 598(7882). I–48. doi:10.1038/s41586-021-04018-9.

Lipp, Reiner. 2009a. *Die indogermanischen und einzelsprachlichen Palatale im Indoiranischen*. Vol. 1. Neurekonstruktion, Nuristan-Sprachen, Genese der indoarischen Retroflexe, Indoarisch von Mitanni. (Indogermanische Bibliothek, 3. Reihe: Untersuchungen). Heidelberg: Carl Winter.

Lipp, Reiner. 2009b. *Die indogermanischen und einzelsprachlichen Palatale im Indoiranischen*. Vol. 2. Thorn-Problem, indoiranische Laryngalvokalisation. (Indogermanische Bibliothek, 3. Reihe: Untersuchungen). Heidelberg: Carl Winter.

LIV[1] = Rix, Helmut & Martin Joachim Kümmel (eds.). 1998. *Lexikon der indogermanischen Verben: die Wurzeln und ihre Primärstammbildungen*. 1. Aufl. Wiesbaden: Reichert.

LIV[2] = Rix, Helmut & Martin Joachim Kümmel (eds.). 2001. *Lexikon der indogermanischen Verben: die Wurzeln und ihre Primärstammbildungen*. 2., erw. und verb. Aufl. Wiesbaden: Reichert.

LIV[3] = Kümmel, Martin Joachim. 2024a. Addenda und Corrigenda zu LIV[2] [Letzte Änderung: 09.05.2024 14:06]. *Online self publication*. http://www.martinkuemmel.de/liv2add.html (26 August, 2024).

LIV[3] = Kümmel, Martin Joachim. 2025b. Addenda und Corrigenda zu LIV[2] [Letzte Änderung: 15.03.2025 14:56]. https://www.gw.uni-jena.de/phifakmedia/42185/kuemmel-liv2-add.pdf (20 March, 2025).

LSJ = Liddell, Henry George & Robert Scott. 1996. *A Greek-English lexicon*. (Ed.) Henry Stuart Jones & Roderick McKenzie. 9th ed. Oxford: Clarendon Press.

Lubotsky, Alexander. 1980. Gr. *pḗgnumi* : Skt. *pajrá-* : and loss of lanryngeal before mediae in Indo-Iranian. *Münchener Studien zur Sprachwissenschaft* 40. 133–138.

Lubotsky, Alexander. 1997a. Review of *Zu Brugmanns Gesetz im Altindischen* Marianne Volkart (Bern: Universität Bern, Institut für Sprachwissenschaft, 1994). *Kratylos* 42. 55–59.

Lubotsky, Alexander. 1997b. The lndo-lranian reflexes of PIE *CRHUV. In Alexander Lubotsky (ed.), *Sound law and analogy: Papers in honor of Robert S.P. Beekes on the occasion of his 60th birthday*, 139–154. (Leiden Studies in Indo-European 9). Amsterdam & Atlanta (GA): Rodopi.

Lubotsky, Alexander. 2001. Reflexes of **sk* in Indo-Iranian. *Incontri Linguistici* 24. 24–57.

Lubotsky, Alexander. 2020. What Language Was Spoken by the People of the Bactria-Margiana Archaeological Complex? In Paul W. Kroll & Jonathan A. Silk (eds.), *"At the Shores of the Sky": Asian Studies for Albert Hoffstädt*, 5–11. (Sinica Leidensia 151). Brill. https://brill.com/display/book/9789004438200/BP000002.xml (29 January, 2025).

Lubotsky, Alexander. 2023. Indo-European and Indo-Iranian wagon terminology and the date of the Indo-Iranian split. In Eske Willerslev, Guus Kroonen & Kristian Kristiansen (eds.), *The Indo-European puzzle revisited: Integrating archaeology, genetics, and linguistics*, 257–262. Cambridge: Cambridge University Press. doi:10.1017/9781009261753.021. https://www.cambridge.org/core/books/indoeuropean-puzzle-revisited/indoeuropean-and-indoiranian-wagon-terminology-and-the-date-of-the-indoiranian-split/ADBF07BCD6447A00E1B5E3EE4E128FA7 (23 November, 2023).

Lubotsky, Alexander & Alwin Kloekhorst. 2021. Indo-Aryan °(a)u̯artanna in the Kikkuli treatise. In Hannes A. Fellner, Melanie Malzahn & Michaël Peyrot (eds.), lyuke wmer ra*: Indo-European studies in honor of Georges-Jean Pinault*, 331–336. Ann Arbor & New York (NY): Beech Stave Press.

Lubotsky, Alexander M. 1988. *The system of nominal accentuation in Sanskrit and Proto-Indo-European*. (Memoirs of the Kern Institute). Leiden, New York (NY), København & Köln: Brill.

Lubotsky, Alexander M. 1990. La loi de Brugmann et *H_3e-. In Jean Kellens (ed.), *La reconstruction des laryngales*, 129–136. (Bibliothèque de La Faculté de Philosophie et Lettres de l'Université de Liège). Liège & Paris: Les Belles Lettres.

Lubotsky, Alexander M. 2001. The Indo-Iranian substratum. *Early contacts between Uralic and Indo-European: Linguistic and archeological considerations: Papers presented at an international symposium held at the Tvärminne*

Research Station of the University of Helsinki, 8–10 January, 1999, 301–309. (Suomalais-Ugrilaisen Seuran Toimituksia / Mémoires de La Société Finno-Ougrienne 242). Helsinki: Suomalais-Ugrilainen Seura.

Lubotsky, Alexander M. 2018. The phonology of Proto-Indo-Iranian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, 1875–1888. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 3). Berlin: De Gruyter Mouton.

Lundquist, Jesse & Anthony D. Yates. 2018. The morphology of Proto-Indo-European. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 3, 2079–2195. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 41). De Gruyter Mouton.

Mallory, James P. 1975. The Indo-European homeland problem: The logic of the inquiry. UCLA.

Mallory, James P. 1989. *In search of the Indo-Europeans: Language, archaeology, and myth*. London: Thames & Hudson.

Mallory, James P. 1996. The Indo-European homeland problem: A matter of time. In Karlene Jones-Bley & Martin E. Huld (eds.), *The Indo-Europeanization of northern Europe*, 1–22. (Journal of Indo-European Studies, Monograph Series). Washington, DC: Institute for the Study of Man.

Mallory, James P. & Douglas Q. Adams. 1997. *Encyclopedia of Indo-European culture*. London & Chicago: Fitzroy Dearborn.

Mallory, James P. & Douglas Q. Adams. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford & New York: Oxford University Press.

Mallory, James Patrick. 2013. Twenty-first century clouds over Indo-European homelands. *Journal of Language Relationship • Вопросы языкового родства* 9(2013). 145–154.

Malzahn, Melanie. 2010. *The Tocharian verbal system*. (Brill's Studies in Indo-European Languages & Linguistics). Leiden & Boston: Brill.

Mann, Stuart E. 1984. *An Indo-European comparative dictionary*. Hamburg: Buske.

Martínez, Javier & Michiel de Vaan. 2014. *Introduction to Avestan*. (Trans.) 2001 English translation of {Martínez and de Vaan. (Brill Introductions to Indo-European Languages). Leiden & Boston: Brill.

Martirosyan, Hrach. 2013. The place of Armenian in the Indo-European language family: the relationship with Greek and Indo-Iranian. *Journal of Language Relationship • Вопросы языкового родства* 10(1). 85–138. doi:10.31826/jlr-2013-100107.

Martirosyan, Hrach K. 2010. *Etymological dictionary of the Armenian inherited lexicon*. (Leiden Indo-European Etymological Dictionary Series 8). Leiden & Boston (MA): Brill.

Matras, Yaron & Peter Bakker. 2003. *The mixed language debate: Theoretical and empirical advances*. (Trends in Linguistics 145). Berlin & New York (NY): Mouton de Gruyter.

Matras, Yaron & Anton Tenser (eds.). 2020. *The Palgrave Handbook of Romani Language and Linguistics*. Cham: Springer International Publishing. doi:10.1007/978-3-030-28105-2. https://link.springer.com/10.1007/978-3-030-28105-2 (19 March, 2025).

Matzinger, Joachim. 2006. *Der altalbanische Text Mbsuame e Krështerë (Dottrina cristiana) des Lekë Matrënga von 1592: Eine Einführung in die albanische Sprachwissenschaft*. (Jenaer Indogermanistische Textbearbeitung 3). Dettelbach: J.H. Röll.

Mayrhofer, Manfred. 1982. Welches Material aus dem Indo-Arischen von Mitanni verbleibt für eine selektive Darstellung? In Erich Neu (ed.), *Investigationes philologicae et comparativae: Gedenkschrift für Heinz Kronasser*, 72–90. Wiesbaden: Harrassowitz.

Mayrhofer, Manfred. 1984. Lassen sich Vorstufen des Uriranischen nachweisen? *Anzeiger der Phil.-Hist. Klasse der Österreichischen Akademie der Wissenschaften* 1983(120). 249–255.

Mayrhofer, Manfred. 1986. *Etymologisches Wörterbuch des Altindoarischen*. Vol. 1. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Carl Winter.

Mayrhofer, Manfred. 1996. *Etymologisches Wörterbuch des Altindoarischen*. Vol. 2. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Carl Winter.

Mayrhofer, Manfred. 2004. Zur Vertretung der Indogermanischen Liquiden in den Indo-Iranischen. *Indologia Tauriniensia* 28. 149–161.

McMahon, April M. S. & Robert McMahon. 2006. Why linguists don’t do dates: Evidence from Indo-European and Australian languages. In Peter Forster &

Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 153–160. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Meid, Wolfgang. 1975. Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen. In Helmut Rix (ed.), *Flexion und Wortbildung: Akten der V. Fachtagung der Indogermanischen Gesellschaft, Regensburg, 9.–14. September 1973*, 204–219. Wiesbaden: Reichert.

Melchert, H. Craig. 2003. PIE "thorn" in Cuneiform Luvian? In Karlene Jones-Bley, Martin E. Huld, Angela Della Volpe & Miriam Robbins Dexter (eds.), *Proceedings of the Fourteenth Annual UCLA Indo-European Conference, Los Angeles, November 8-9, 2002*, 145–61. (Journal Oflndo-European Monograph Series 47). Washington (DC): Institute for the Study of Man.

Melchert, H. Craig & Norbert Oettinger. 2009. Ablativ und Instrumental im Hethitischen und Indogermanischen: Ein Beitrag zur relativen Chronologie. *Incontri linguistici* 32. 53–73.

Melchert, Harold Craig. 1987. PIE velars in Luvian. In Calvert Watkins (ed.), *Studies in memory of Warren Cowgill (1929-1985): Papers from the Fourth East Coast Indo-European Conference. Cornell University, June 6-9, 1985*, 181–204. (Untersuchungen Zur Indogermanischen Sprach- Und Kulturwissenschaft 3). Berlin & New York (NY): Walter de Gruyter.

Melchert, Harold Craig. 1989. New Luvo-Lycian isoglosses. *Historische Sprachforschung* 102(1). 23–45.

Melchert, Harold Craig. 1994. *Anatolian historical phonology*. (Leiden Studies in Indo-European 3). Amsterdam: Rodopi.

Melchert, Harold Craig. 2012. Luvo-Lycian dorsal stops revisited. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphophonemics*, 206–218. (LINCOM Studies in Indo-European Linguistics 41). München: Lincom Europa.

Miller, Amanda. 2013. 3.1 Northern Khoesan Phonetics and Phonology. In Rainer Vossen (ed.), *The Khoesan languages*, 45–50. (Routledge Language Family Series). London & New York (NY): Routledge.

Miller, D. Gary. 1976. Pure velars and palatals in Indo-European: A rejoinder to Magnusson. *Linguistics* 14(178). 47–64. doi:10.1515/ling.1976.14.178.47.

Monier-Williams, Monier. 1899. *A Sanskrit-English dictionary: Etymologically and philologically arranged with special reference to cognate Indo-European languages*. 2nd ed. Oxford: Clarendon Press.

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.

Nakhleh, Luay, Tandy Warnow, Don Ringe & Steven N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Transactions of the Philological Society* 103(2). 171–192.

Narasimhan, Vagheesh M., Rebecca Bernardos, Swapan Mallick, et al. 2019. The formation of human populations in South and Central Asia. *Science (American Association for the Advancement of Science)* 365(6457). 999-. doi:10.1126/science.aat7487.

Nelson, David Niles. 1986. *The Historical Development of the Nuristani Languages*. University of Minnesota. Google-Books-ID: o1RMrQhB2jMC.

Nichols, Johanna & Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5). 760–820. doi:10.1111/j.1749-818X.2008.00082.x.

Nielsen Whitehead, Benedicte. 2018. PIE *g^wh_3-éu̯- “cow.” In Lucien Christiaan van Beek, Guus Jan Kroonen, Tijmen Pronk, et al. (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*, 217–231. Ann Arbor (MI) & New York (NY): Beech Stave Press.

NIL = Wodtko, Dagmar S., Britta Sofie Irslinger & Carolin Schneider. 2008. *Nomina im Indogermanischen Lexikon*. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Winter.

Norbruis, Stefan. 2023. On the -*θ*- in Greek *οἶσθα* ‘you know’ < *uoid-th$_2$e. *Glotta* 99. 227–239.

Nordqvist, Kerkko & Volker Heyd. 2020. The forgotten child of the wider Corded Ware family: Russian Fatyanovo Culture in context. *Proceedings of the Prehistoric Society* 86. 65–93. doi:10.1017/ppr.2020.9.

Normier, Rudolf. 1977. Idg. Konsonantismus, germ. “Lautverschiebung” und Vernerisches Gesetz. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 91(2). 171–218.

Normier, Rudolf. 1980. Tocharisch *ñkät/ñakte* "Gott." *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 94. 251–281.

Obrador-Cursach, Bartomeu. 2019. On the place of Phrygian among the Indo-European languages. *Journal of Language Relationship • Вопросы языкового родства* 17(3). 233–245.

Olander, Thomas. 2009. *Balto-Slavic accentual mobility*. (Trends in Linguistics. Studies and Monographs 199). Berlin & New York (NY): Mouton de Gruyter.

Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. (Brill's Studies in Indo-European Languages and Linguistics 14). Leiden & Boston (MA): Brill.

Olander, Thomas. 2018. Connecting the dots: The Indo-European family tree as a heuristic device. In David M. Goldstein, Stephanie W. Jamison & Brent Vine (eds.), *Proceedings of the 29th annual UCLA Indo-European conference*, 181–202. (Journal of Indo-European Studies Monograph Series). Bremen: Hempen.

Olander, Thomas. 2019a. Indo-European cladistic nomenclature. *Indogermanische Forschungen* 124. 231–244.

Olander, Thomas. 2019b. A complicated relationship: Balto-Slavic accentual mobility as a non-trivial shared innovation. In Iván Igartua, Brian D. Joseph, Lars Heltoft, et al. (eds.), *Perspectives on language structure and language change*, 359–380. (Current Issues in Linguistic Theory). Amsterdam & Philadelphia: Benjamins.

Olander, Thomas. 2019c. The Indo-European homeland: Introducing the problem. In Birgit Anette Olsen, Thomas Olander & Kristian Kristiansen (eds.), *Tracing the Indo-Europeans: New evidence from archaeology and historical linguistics*, 7–34. Philadelphia (PA): Oxbow.

Olander, Thomas. 2022a. To **b* or not to **b*: Proto-Indo-European **b* in a phylogenetic perspective. *Historische Sprachforschung* 133(1). 182–208.

Olander, Thomas. 2022b. West, east, home is best? The phylogenetic relationship between Balto-Slavic and Indo-Iranian. Paper presented at the New Shoots from the Roots, Copenhagen, 2 June 2022, University of Copenhagen.

Olander, Thomas. 2022c. Introduction. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 1–17. Cambridge: Cambridge University Press. (18 August, 2022).

Olander, Thomas (ed.). 2022d. *The Indo-European Language Family: A Phylogenetic Perspective*. Cambridge: Cambridge University Press. (18 August, 2022).

Olander, Thomas. 2023. Linguistic phylogenetics and words for metals in Indo-European. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European puzzle revisited: Integrating archaeology, genetics, and linguistics*, 93–104. Cambridge: Cambridge University Press.

Olander, Thomas. 2024a. Reconsidering the Indo-European centum–satəm isogloss. Paper presented at the "The Speakers of Indo-European and their World: 150 years of Indo-European studies in Basel". 17. Fachtagung der indogermanischen Gesellschaft, 9.–12.09.2024, Universität Basel, Basel. unter Mitwirkung von Simon Poulsen.

Olander, Thomas. 2024b. The linguistic perspective. Paper presented at the Reconstructing Indo-European Prehistory, Scuola Superiore Sant'Anna, Pisa.

Olander, Thomas & Simon Poulsen. 2022a. Indo-Iranian and Balto-Slavic. Conference presentation. Paper presented at the Secondary Homelands of the Indo-European Languages: Arbeitstagung der Indogermaniwschen Gesellschaft, Universiteit Leiden, Leiden, 7. September 2002, Leiden.

Olander, Thomas & Simon Poulsen. 2022b. Phylogenetic analysis of *Lexikon der indogermanischen Verben*. Research presentation. University of Pennsylvania, Philadelphia (PA).

Olivier, Jean-Pierre. 2024. Syllabic scripts in the Aegean and Cyprus in the second and first millennium. In John Killen (ed.), *The New Documents in Mycenaean Greek*, vol. 1. Introductory Essays, 49–94. Cambridge & New York (NY): Cambridge University Press. doi:10.1017/9781139029049.

Olsen, Birgit Anette. 1988. *The Proto-Indo-European instrument noun suffix* *-tlom *and its variants*. (Det Kongelige Danske Videnskabernes Selskabs Historisk-Filologiske Meddelser 55). København: Munksgaard.

Olsen, Birgit Anette. 1992. The development of high vowel plus laryngeal in Armenian. *Proceedings of the fourth international conference on Armenian linguistics, Cleveland's State University, Cleveland, Ohio, September 14-18, 1991*, 129–146. Delmar (NY): Caravan Books.

Olsen, Birgit Anette. 1999. *The noun in Biblical Armenian: Origin and word formation: With special emphasis on the Indo-European heritage*. (Trends in Linguistics. Studies and Monographs 119). Berlin & New York (NY): Mouton de Gruyter.

Olsen, Birgit Anette. 2009. The conditioning of laryngeal breaking in Greek. In Rosemarie Lühr & Sabine Ziegler (eds.), *Protolanguage and prehistory: Akten der XII. Fachtagung der Indogermanischen Gesellschaft, vom 11. bis 15. Oktober 2004 in Krakau*, 348–365. Wiesbaden: Reichert.

Olsen, Birgit Anette. 2011. Review of Martirosyan (2010). *Kratylos* 56. 14–30.

Olsen, Birgit Anette. 2017. Armenian textile terminology. In Salvatore Gaspa, Cécile Michel & Marie-Louise Nosch (eds.), *Textile terminologies from the Orient to the Mediterranean and Europe, 1000 BC to 1000 AD*, 188–201. Lincoln (NE): Zea Books.

Olsen, Birgit Anette. 2018a. The development of interconsonantal laryngeals in Indo-Iranian and Old Avestan *ząϑā ptā*. In Lucien Christiaan van Beek, Guus Jan Kroonen, Tijmen Pronk, et al. (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*, 256–269. Ann Arbor (MI) & New York (NY): Beech Stave Press.

Olsen, Birgit Anette. 2018b. Notes on the Indo-European vocabulary of sheep, wool and textile production. In Rune Iversen & Guus Kroonen (eds.), *Digging for words: Archaeolinguistic case studies from the XV Nordic TAG conference held at the University of Copenhagen, 16–18 April 2015*, 69–77. Oxford: BAR.

Olsen, Birgit Anette. 2023. An archaeolinguistic approach to Indo-European wool terminology. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European Puzzle Revisited*, 282–286. 1st ed. Cambridge University Press. doi:10.1017/9781009261753.024. https://www.cambridge.org/core/product/identifier/9781009261753%23CN-bp-18/type/book_part (23 November, 2023).

Olsen, Birgit Anette & Rasmus Thorsø. 2022. Armenian. In Thomas Olander (ed.), *The Indo-European language family. A phylogenetic perspective*.

Paškevič, Halyna Oleksandrivna & Myxajlo Jurijovyč Videjko. 2006. *Ril'nyctvo plemen trypilskoji kultury*. Kyjiv: Instytut Arxeolohiji Nacionalnoji Akademiji Nauk Ukrajiny.

Pasquali, Giorgio. 1994. *Procli Diadochi In Platonis Cratylum commentaria*. Ed. stereotypa ed. primae (1908). (Bibliotheca scriptorum Graecorum et Romanorum Teubneriana). Stutgardiae Lipsiae: Teubner.

Pedersen, Holger. 1895. Das indogermanische *s* im Slavischen. *Indogermanische Forschungen* 5(1). 33–87. doi:10.1515/9783110242478.33.

Pellard, Thomas, Robin J. Ryder & Guillaume Jacques. forthc. The Family Tree model. In Ledgeway (ed.), *The Wiley Blackwell companion to diachronic linguistics*. Wiley-Blackwell. https://www.academia.edu/101656989/The_Family_Tree_model (22 May, 2023). Preprint.

Pereltsvaig, Asya & Martin W. Lewis. 2015. *The Indo-European controversy: Facts and fallacies in historical linguistics*. Cambridge: Cambridge University Press.

Peyrot, Michaël. 2019. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European Linguistics* 7(1). Brill. 72–121. doi:10.1163/22125892-00701007.

Peyrot, Michaël. 2022. Tocharian. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 83–101. Cambridge, UK, & New York: Cambridge University Press.

Pinault, Georges-Jean. 2016. Hittite and Indo-European wool. In Henning Marquardt, Silvio Reichmuth & José Virgilio García Trabazo (eds.), *Anatolica et Indogermanica: Studia linguistica in honorem Johannis Tischler septuagenarii dedicata*, 241–255. (Innsbrucker Beiträge Zur Sprachwissenschaft). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Piwowarczyk, Dariusz R. 2022a. *Rekonstrukcja praindoeuropejskiego systemu fonologicznego*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Piwowarczyk, Dariusz R. 2022b. Computational Approaches to Linguistic Chronology and Subgrouping. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 33–51. 1st ed. Cambridge University Press. doi:10.1017/9781108758666.003. https://www.cambridge.org/core/product/identifier/9781108758666%23CN-bp-3/type/book_part (10 January, 2025).

Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. (Indogermanische Bibliothek. Reihe 3. Untersuchungen). Heidelberg: Carl Winter.

Poulsen, Simon. 2020. Ek […] wraitalaþo: The Proto-Norse strong preterite 1sg ending in light of the Trollhättan II bracteate. *NOWELE. North-Western European Language Evolution* 73(1). 21–43. doi:10.1075/nowele.00032.pou.

Poulsen, Simon. 2025a. Loanwords and Linguistic Phylogenetics: **pelek̑u-* 'axe' and **(H)a(i̯)ĝ-* 'goat.' *Transactions of the Philological Society* 123(1). 116–136. doi:10.1111/1467-968X.12308.

Poulsen, Simon. 2025b. Addenda et corrigenda to S. Poulsen (2025) "Loanwords and Linguistic Phylogenetics: **pelek̑u-* 'axe' and **(H)a(i̯)ĝ-* 'goat'", *Transactions of the Philological Society* 123(1). 116-136. (See also https://zenodo.org/records/15075920 [zenodo.org]). *Transactions of the Philological Society* journal homepage. https://philsoc.org.uk/transactions. doi.org/10.5281/zenodo.15079561.

Poulsen, Simon & Thomas Olander. 2021. Intriguing trees: Indo-European linguistic subgrouping. Dissiminating research presentation. Paper presented at the Center for Studies in Indo-European Language and Culture, Stockholm University, Swedish Academy for Advanced Study, Uppsala.

Pozza, Marianna. 2019. On the original formulation and on the resonance over time of Grassmann's Law: remarks on a still open issue. *Lingua Posnaniensis* 61(1). 107–130. doi:10.2478/linpo-2019-0007.

Pronk, Tijmen. 2021. Indo-European secondary products terminology and the dating of Proto-Indo-Anatolian. *Journal of Indo-European Studies* 49(1 & 2). 141–170.

Pronk, Tijmen. 2022. Balto-Slavic. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 269–292. Cambridge, UK, & New York: Cambridge University Press.

Radivojević, Miljana & Benjamin Roberts. 2021. Balkan metallurgy and society, 6200–3700 BC. In Miljana Radivojević, Benjamin Roberts, Miroslav Marić, et al. (eds.), *The Rise of Metallurgy in Eurasia: Evolution, Organisation and Consumption of Early Metal in the Balkans*, 11–37. Archaeopress Archaeology. doi:10.32028/9781803270425.

https://archaeopress.com/Archaeopress/Products/9781803270425 (22 March, 2025).

Rasmussen, Jens Elmegård. 1989a. *Studien zur Morphophonemik der indogermanischen Grundsprache*. (Innsbrucker Beiträge Zur Sprachwissenschaft). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Rasmussen, Jens Elmegård. 1989b. Die Tenues Aspiratae: Dreiteilung oder Vierteilung des Indogermanischen Plosivsystems und die Konsequenzen dieser Frage für die Chronologie einer Glottalreihe. In Theo Vennemann (ed.), *The new sound of Indo-European: Essays in phonological reconstruction*. (Trends in Linguistics Studies and Monographs 41). Berlin: Mouton de Gruyter.

Rasmussen, Jens Elmegård. 1993. Review of Schrijver (1991). *Acta Linguistica Hafniensia* 26(1). 175–205.

Rastorgueva, Vera Sergeevna & Džoj Iosifovna Ėdel'man. 2003. *Ėtimologičeskij slovar' iranskix jazykov*. Vol. 2. b-d. Moskva: Vostočnaja literatura.

Reichelt, Hans. 1900. Der Frahang i Oīm: Zand Pahlavi glossary. *Wiener Zeitschrift für die Kunde des Morgenlandes* 14. 177–213.

Reichelt, Hans. 1901. Der Frahang i Oīm: Zand Pahlavi glossary. *Wiener Zeitschrift für die Kunde des Morgenlandes* 15. 117–186.

Renfrew, Colin. 1987. *Archaeology and language: The puzzle of Indo-European origins*. London: Penguin.

Renfrew, Colin. 2001. The Antolian [*sic*] origins of Proto-Indo-European and the autochthony of the Hittites. In Robert Drews (ed.), *Greater Anatolia and the Indo-Hittite language family: Papers presented at a colloquium hosted by the University of Richmond, March 18–19, 2000*, 36–63. (Journal of Indo-European Studies. Monograph Series). Washington, DC: Institute for the Study of Man.

Rexová, Kateřina, Daniel Frynta & Jan Zrzavý. 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19. 120–127. doi:10.1111/j.1096-0031.2003.tb00299.x.

Richter, Thomas. 2012. *Bibliographisches Glossar des Hurritischen*. Wiesbaden: Harrassowitz.

Rieken, Elisabeth. 2007. Hieroglyphen-luwisch *i-zi-ia-*: Ein Beitrag zur Rekonstruktion der urindogermanischen Kulturgeschichte. *Darъ slovesьny:*

Festschrift für Christoph Koch zum 65. Geburtstag, 263–275. (Specimina Philologiae Slavicae 146). München: Otto Sagner.

Rieken, Elisabeth & Ilya Sergeyevich 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. (Monograph Series of the Institute of Archaeology, Tel Aviv University 28). Tel Aviv: Institute of Archaeology University Tel Aviv.

Ringe, Don. 2010. 'Thorn' clusters and Indo-European subgrouping. In Ron Kim (ed.), *Ex Anatolia Lux: Anatolian and Indo-European studies in honor of H. Craig Melchert on the occasion on his sixty-fifth birthday*, 330–38. Ann Arbor (MI): Beech Stave Press.

Ringe, Don. 2024. *The Linguistic Roots of Ancient Greek*. 1st ed. Oxford: Oxford University Press. https://academic.oup.com/book/56348 (23 July, 2024).

Ringe, Don, Tandy Warnow, Ann A. Taylor, et al. 1998. Computational Cladistics and the Position of Tocharian. In V. Mair (ed.), *Lexical Archaisms in the Tocharian Languages. The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*, vol. 1. Archeology, Migration and Nomadism, Linguistics, 391–414. Washington (DC): Institute for the Study of Man.

Ringe, Donald A. 1998. Some consequences of a new proposal for subgrouping the IE family. In Benjamin K. Bergen, Madelaine C. Plauché & Ashlee C. Bailey (eds.), *Proceedings of the Twenty-Fourth Annual Meeting of the Berkeley Linguistics Society, February 14–16, 1998: Special session on Indo-European subgrouping and internal relations*, 32–46. Berkeley: Berkeley Linguistics Society.

Ringe, Donald A. 2017a. Indo-European dialectology. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, 62–75. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 1). [Berlin]: De Gruyter Mouton.

Ringe, Donald A. 2017b. *From Proto-Indo-European to Proto-Germanic*. 2nd ed. (A Linguistic History of English 1). Oxford & New York (NY): Oxford University Press.

Ringe, Donald A. 2022. What We Can (and Can't) Learn from Computational Cladistics. In Thomas Olander (ed.), *The Indo-European Language Family:*

A Phylogenetic Perspective, 52–62. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.004 (27 September, 2022).

Ringe, Donald A. & Ann A. Taylor. 2003. Screened dataset. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/~nakhleh/CPHL/IEDATA_112603 (1 June, 2022).

Ringe, Donald A. & Ann A. Taylor. 2004. Unscreened dataset. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/~nakhleh/CPHL/IEDATA_050704 (1 June, 2022).

Ringe, Donald A. & Ann A. Taylor. 2007a. Phonological characters. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/~nakhleh/CPHL/code-p-07.pdf (1 June, 2022).

Ringe, Donald A. & Ann A. Taylor. 2007b. Morphological characters. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/~nakhleh/CPHL/code-m-07.pdf (1 June, 2022).

Ringe, Donald A. & Ann A. Taylor. 2007c. The Indo-European wordlists. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/~nakhleh/CPHL/ie-wordlist-07.pdf (1 June, 2022).

Ringe, Donald A. & Ann A. Taylor. 2012. Cognations of lexical characters. Supplementary material of Ringe, Warnow & Taylor 2002. http://tandy.cs.illinois.edu/cognations-2k-revised.pdf (1 June, 2022).

Ringe, Donald A., Tandy Warnow & Ann A. Taylor. 2002. Indo-European and computational linguistics. *Transactions of the Philological Society* 100(1). 59–129.

Rix, Helmut. 1976. *Historische Grammatik des Griechischen. Laut- und Formenlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Rix, Helmut. 1985. *Sūdor* und *sīdus*. In Hermann M. Ölberg, Gernot Schmidt & Heinz Bothien (eds.), *Sprachwissenschaftliche Forschungen: Festschrift für Johann Knobloch zum Geburtstag am 5. Januar 1984*, 339–350. (Innsbrucker Beiträge Zur Kulturwissenschaft 23). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Rix, Helmut. 2009. *Historische Grammatik des Griechischen: Laut- und Formenlehre*. 2. kor. Aufl. (WBG-Bibliothek). Darmstadt: Wissenschaftliche Buchgesellschaft.

Ross, Malcolm. 1997. Social networks and kinds of speech-community event. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language*, 209–

261. (One World Archaeology 1. Theoretical and methodological orientations). London & New York: Routledge.

Sands, Bonny, Edward D. Elderkin, Amanda Miller, et al. 2013. Phonetics and Phonology. In Rainer Vossen (ed.), *The Khoesan languages*, 38–87. (Routledge Language Family Series). London & New York (NY): Routledge.

Sasseville, David. 2021. *Anatolian verbal stem formation: Luwian, Lycian and Lydian*. (Brill's Studies in Indo-European Languages & Linguistics 21). Leiden & Boston (MA): Brill.

Sayeed, Ollie. 2019. Hauchumsprung and the historical phonology of Greek *h. *Indo-European Linguistics* 7(1). 164–175.

Scarborough, Matthew Joseph Charles. 2020. Cognacy and computational cladistics: Issues in determining lexical cognacy for Indo-European cladistic research. In Thomas Olander & Matilde Serangeli (eds.), *Dispersals and diversification: Linguistic and archaeological perspectives on the early stages of Indo-European*, 179–208. (Brill's Studies in Indo-European Languages and Linguistics). Leiden & Boston (MA): Brill.

Scarborough, Matthew Joseph Charles. 2023. *The Aeolic Dialects of Ancient Greek: A Study in historical dialectology and linguistic classification*. (Brill's Studies in Indo-European Languages and Linguistics 26). Leiden & Boston (MA): Brill. https://brill.com/display/title/57781 (15 February, 2023).

Schindler, Jochem. 1972. Das Wurzelnomen im Arischen und Griechischen. Würzburg: Julius-Maximilians-Universität Würzburg Inauguraldissertation.

Schindler, Jochem. 1975. Armensich erkn, griechisch ὀδύνη, irisch idu. *Zeitschrift für vergleichende Sprachforschung* 89. 53–65.

Schindler, Jochem. 1977. A thorny problem. *Die Sprache* 23(1). 25–35.

Schmidt, Klaus T. 1992. Archaismen des Tocharischen und ihre Bedeutung für Fragen der Rekonstruktion der Ausgliederung. In Robert Stephen Paul Beekes, Alexander Lubotsky & Joseph Johannes Sicco Weitenberg (eds.), *Rekonstruktion und relative Chronologie: Akten der VIII. Fachtagung der Indogermanischen Gesellschaft, Leiden, 31. August - 4. September 1987*, 101–114. (Innsbrucker Beiträge zur Sprachwissenschaft 65). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Schmitt, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. (Innsbrucker Beiträge Zur

Sprachwissenschaft 32). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Schoubben, Niels. 2019. DIAL/RINGUISTICS: The opposition between /l/ and /r/ in the diasystem of Vedic Sanskrit. A study in diachronic linguistics, dialectology and sociolinguistics. Leiden: Universiteit Leiden MA thesis.

Schrijver, Peter. 2006. Review of *Veni, vidi, vici: die Vorgeschichte des lateinischen Perfektsystems* by Gerhard Meiser (München: C. H. Beck, 2003). *Kratylos* 51. 46–64.

Schumacher, Stefan & Joachim Matzinger. 2013. *Die Verben des Altalbanischen: Belegwörterbuch, Vorgeschichte und Etymologie*. (Albanische Forschungen). Wiesbaden: Harrassowitz.

Schürr, Dieter. 2023. Tochter in anatolischen Sprachen. *Wékʷos: Revue d'études indo-européennes* 7. 239–245.

Schwyzer, Eduard. 1939. *Griechische Grammatik: Auf der Grundlage von Karul Brugmanns griechischer Grammatik*. Vol. 1. Allgemeiner Teil. Lautlehre. Wortbildung. Flexion. (Handbuch Der Altertumswissenschaft. 2. Abteilung, 1. Teil 1). München: C. H. Beck.

Scott, Ashley, Sabine Reinhold, Taylor Hermes, et al. 2022. Emergence and intensification of dairying in the Caucasus and Eurasian steppes. *Nature Ecology & Evolution* 6(6). Nature Publishing Group. 813–822. doi:10.1038/s41559-022-01701-6.

Seebold, Elmar. 1970. *Vergleichendes und etymologisches Wörterbuch der germanischen starken Verben*. (Janua Linguarum. Series Practica 85). The Hague & Paris: Mouton.

Seebold, Elmar. 1999. Review of *LIV*[1]. *Indogermanische Forschungen* 104. 287–295.

Sherratt, Andrew G. 1983. The secondary exploitation of animals in the Old World. *World Archaeology* 15. 90–104.

Shishlina, N. I., O. V. Orfinskaya, P. Hommel, et al. 2020. Bronze Age Wool Textile of the Northern Eurasia: New Radiocarbon Data. *Nanotechnologies in Russia* 15(9–10). 629–638. doi:10.1134/S1995078020050146.

Shishlina, Natalia I., Polina S. Ankusheva, Olga V. Orfinskaya & Daria V. Kiseleva. 2023. Wool Fibers of the Northern Eurasian Bronze Age: The Cultural and Geographical Contexts. In Eske Willerslev, Guus Kroonen & Kristian Kristiansen (eds.), *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics*, 275–281. Cambridge: Cambridge

University Press. doi:10.1017/9781009261753.023. https://www.cambridge.org/core/books/indoeuropean-puzzle-revisited/wool-fibers-of-the-northern-eurasian-bronze-age-the-cultural-and-geographical-contexts/C12A9BDC134416E38A13DB04DAE55A76 (23 March, 2025).

Sihler, Andrew Littleton. 1995. *New Comparative Grammar of Greek and Latin*. Oxford & New York (NY): Oxford University Press.

Simon, Zsolt. 2013. Once again on the Hieroglyphic Luwian sign *19 〈á〉. 118(2013). 1–22. doi:10.1515/indo.2013.118.2013.1.

Sims-Williams, Nicholas. 1989. Sogdian. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 173–192. Wiesbaden: Reichert.

Sims-Williams, Patrick. 1998. Genetics, linguistics, and prehistory: Thinking big and thinking straight. *Antiquity* 72(277). 505–527.

Skjærvø, Prods O. 2022. Khotan 4. The Khotanese language. In Elton L. Daniel (ed.), *Encyclopaedia Iranica*, vol. 17, 117–133. Leiden: Brill. https://brill.com/edcollbook/title/62996 (4 February, 2025).

Skjærvø, Prods Oktor. 2007. Avestan and Old Persian Morphology. In Alan S. Kaye (ed.), *Morphologies of Asia and Africa*, vol. 2, 853–940. Winona Lake (IN): Eisenbrauns.

Skjærvø, Prods Oktor. 2017a. The documentation of Iranian. *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 471–481. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). Berlin & Boston (MA): De Gruyter. doi:10.1515/9783110261288-032.

Skjærvø, Prods Oktor. 2017b. The morphology of Iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 503–549. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter Mouton.

Smoczyński, Wojciech. 2018. *Lithuanian etymological dictionary*. Berlin & New York (NY): Peter Lang.

Søborg, Tobias Mosbæk. 2020. Sigmatic verbal formations in Anatolian and Indo-European. København: Københavns Universitet (University of Copenhagen) PhD-dissertation.

von Soden, Wolfram. 1965. *Akkadisches Handworterbuch*. Vol. 1. A-L. Wiesbaden: Harrassowitz.

Söhrman, Ingmar. 2016. Onomasiological differentiation. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 588–595. (Oxford Guides to the World's Languages). Oxford: Oxford university press.

Solans, Gabriel. 2024. Paléogénétique : que peut-elle dire aux linguistes sur les langues indo-européennes ? *Wék^{w}os: Revue d'études indo-européennes* 7. 247–281.

Steensland, Lars. 1973. *Die Distribution der urindogermanischen sogenannten Gutturale*. (Studia Slavica Upsaliensia 12). Stockholm: Almqvist & Wiksell.

Strand, Richard F. 2023. Kâmbojâs and Sakas in the Holly-Oak Mountains: On the origins of the Nûristânîs. In Alberto M. Cacopardo & Augusto S. Cacopardo (eds.), *Roots of Peristan: the pre-islamic cultures of the Hindukush/Karakorum: proceedings of the International Interdisciplinary Conference ISMEO, Rome, Palazzo Baleani, 5-7 October, 2022*, 781–808. (Serie Orientale Roma. Nuova Serie 37). Roma: Ismeo.

Stüber, Karin. 2000. Zur Herkunft der altindischen Infinitive auf <i>-sáni</>. *Münchener Studien zur Sprachwissenschaft* 60. 135–167.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). John Benjamins Publishing Company. 226–246. doi:10.1075/dia.27.2.04tad.

Taylor, Ann A., Tandy Warnow & Don Ringe. 2000. Character-based reconstruction of a linguistic cladogram. In John Charles Smith & Delia Bentley (eds.), *Historical Linguistics 1995: Selected papers from the 12th International Conference on Historical Linguistics, Manchester, August 1995*, vol. 1. General issues and non-Germanic Languages.. Selected papers from the 12th International Conference on 1995, 393–408. (Amsterdam Studies in the Theory and History of Linguistic Science. Series 4. Current Issues in Linguistic Theory 161). Amsterdam & Philadelphia (PA): John Benjamins.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley, Los Angeles & London: University of California Press.

Thompson, Rubert. 1996. Dialects in Mycenaean and Mycenaean among the dialects. *Minos* 31–33. 313–33.

Thompson, Rubert & Torsten Meißner. 2024. The Mycenaean writing system. In John Killen (ed.), *The New Documents in Mycenaean Greek*, vol. 1. Introductory Essays, 95–136. Cambridge & New York (NY): Cambridge University Press. doi:10.1017/9781139029049.

Thorsø, Rasmus. 2020. Two Balkan Indo-European Loanwords. In Matilde Serangeli & Thomas Olander (eds.), *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*, 251–262. Leiden & Boston (MA): Brill.

Thorsø, Rasmus, Andrew Wigman, Anthony Jakob, et al. 2023. Word mining: metal names and the Indo-European dispersal. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European puzzle revisited*, 105–126. Cambridge University Press. doi:10.1017/9781009261753.012.

Tischler, Johann. 1973. *Glottochronologie und Lexikostatistik*. (Innsbrucker Beiträge zur Sprachwissenschaft 11). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Trager, George L. & Henry Lee Smith. 1950. A chronology of Indo-Hittite. *Studies in Linguistics* 8(3). 61–70.

Tremblay, Xavier. 2008. Iranian historical linguistics in the twentieth century: Part 2. *Friends and Alumni of Indo-European Studies Bulletin* 13(1). 1–51.

Trofimov, Artem A. 2016. Annotated Swadesh wordlists for the Indo-Aryan group (Indo-European family). https://starlingdb.org/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\ier\iar&limit=-1.

Uesugi, Hendio (ed.). 2024. *Old Avestan Dictionary*. (Asian and African Lexicon 67). Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

de Vaan, Michiel Arnoud Cor. 2008. *Etymological dictionary of Latin and the other Italic languages*. (Leiden Indo-European Etymological Dictionary Series 7). Leiden & Boston (MA): Brill.

de Vaan, Michiel. 2003. *The Avestan vowels*. (Leiden Studies in Indo-European). Amsterdam & Atlanta, GA: Rodopi.

Vaillant, André. 1974. *Grammaire comparée des langues slaves*. Vol. 4. La formation des noms. Paris: Éditions Klincksieck.

Van Willigen, Samuel, Sylvain Ozainne, Michel Guélat, et al. 2024. New evidence for prehistoric ploughing in Europe. *Humanities and Social Sciences Communications* 11(1). 372. doi:10.1057/s41599-024-02837-5.

Vasmer, Max. 1955. *Russisches etymologisches Wörterbuch*. Vol. 2. L-Ssuda. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Winter.

Vennemann, Theo. 1974. Sanskrit *ruki* and the concept of a natural class. *Linguistics* 12(130). doi:10.1515/ling.1974.12.130.91.

Villanueva Svensson, Miguel. 1999. The Italic Simple *R*-Endings. *Glotta* 75(3/4). 252–266.

Villanueva Svensson, Miguel. 2014. The distribution of the 3 rd sg. endings of the Indo-European middle present. *Historische Sprachforschung* 127. 68–100.

Viñas-Caron, Laura C., Mikkel Nørtoft, Peder Flemestad, et al. 2024. From Fleece to Thread. Interdisciplinary Evidence for the Origins of Sheep Wool. In Ulla Mannering, Marie-Louise Nosch & Anne Drewsen (eds.), *The Common Thread*, 33–60. Turnhout: Brepols Publishers. doi:10.1484/M.NAA-EB.5.141753. https://www.brepolsonline.net/doi/10.1484/M.NAA-EB.5.141753 (26 September, 2024).

Vincent, Nigel. 2016. Continuity and change from Latin to Romance. In J. N. Adams & Nigel Vincent (eds.), *Early and Late Latin: Continuity or Change?*, 1–13. Cambridge: Cambridge University Press. doi:10.1017/CBO9781316450826.002. https://www.cambridge.org/core/books/early-and-late-latin/continuity-and-change-from-latin-to-romance/0774BF2CF307871BDCB65991557D2118 (15 January, 2025).

Vine, Brent. 1999. On 'Cowgill's law' in Greek. In Christian Luschützky & Heiner Eichner (eds.), *Compositiones Indogermanicae in memoriam Jochem Schindler*, 555–600. Prague & Vienna: enigma.

Volkart, Marianne. 1994. *Zu Brugmanns Gesetz im Altindischen*. (Universität Bern, Institut Für Sprachwissenschaft, Arbeitspapier). Bern: Universität Bern, Institut für Sprachwissenschaft.

Wackernagel, Jakob & Albert Debrunner. 1975. *Altindische Grammatik*. 2. Ausg. Vol. 3. Nomialflexion, Zahlawort, Pronomen. Göttingen: Vandenhoeck & Ruprecht.

Wandl, Florian & Thilo H. K. Thelitz. 2024. RelChronVis: An interactive web application for visualizing the relative chronology of language changes. *International Journal of Digital Humanities*. doi:10.1007/s42803-024-00086-1. https://doi.org/10.1007/s42803-024-00086-1 (26 March, 2025).

Warnow, Tandy, Don Ringe & Ann A. Taylor. 1996. Reconstructing the evolutionary history of natural languages. In Éva Tardos (ed.), *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms: Atlanta, Georgia, January 28-30, 1996*, 314–321. New York (NY) & Philadelphia (PA): Society for Industrial and Applied Mathematics.

von Wartburg, Walther. 1962. *Französisches etymologisches Wörterbuch: eine Darstellung des galloromanischen Sprachschatzes*. Vol. 10. R. Basel: Zbinden.

Weiss, Michael. 2009. The Cao Bang theory.

Weiss, Michael. 2022a. Italic. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 114–134. Cambridge: Cambridge University Press. doi:10.1017/9781108758666. https://www.cambridge.org/core/books/indoeuropean-language-family/4B44B5ACF0D3BBA89B9408050F112A52 (23 September, 2022).

Weiss, Michael. 2022b. Italo-Celtic. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 102–113. Cambridge: Cambridge University Press. doi:10.1017/9781108758666. https://www.cambridge.org/core/books/indoeuropean-language-family/4B44B5ACF0D3BBA89B9408050F112A52 (23 September, 2022).

Weiss, Michael L. 1995. Life everlasting. Latin *iūgis* "everflowing", Greek *ὑγιής* "healthy", Gothic *ajukdūþs* "eternity" and Avestan *yauuaējī* living forever'. *Münchener Studien zur Sprachwissenschaft* 1994(55). 131–156.

Weiss, Michael L. 2020. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor (MI) & New York (NY): Beech Stave Press.

Wendtland, Antje. 2009. The Position of the Pamir Languages within East Iranian. (Ed.) Agnes Korn. *Orientalia Suecana* 58. 172–188.

West, Martin Litchfield (ed.). 2017. *Homerus Odyssea*. (Bibliotheca Scriptorvm Graecorvm et Romanorvm Tevbneriana BT 2026). Berlin & Boston (MA): De Gruyter.

Whitney, William Dwight. 1885. *The roots, verb-forms, and primary derivatives of the Sanskrit language. A supplement to his Sanskrit grammar*. Leipzig & London: Trübner.

Whittle, Alasdair. 2015. Early agricultural society in Europe. In Candice Goucher & Graeme Barker (eds.), *The Cambridge World History: Volume 2: A World with Agriculture, 12,000 BCE–500 CE*, vol. 2, 555–588. (The Cambridge World History). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978807.023. https://www.cambridge.org/core/books/cambridge-world-history/early-agricultural-society-in-europe/988FFC421D340BB440A3D388934E027E (22 March, 2025).

Winter, Werner. 1997. Lexical archaisms of the Tokharian languages. In Hans Henrich Hock (ed.), *Historical, Indo-European, and lexicographical studies: A Festschrift for Ladislav Zgusta on the occasion of his 70th birthday*, 183–193. (Trends in Linguistics. Studies and Monographs 90). Berlin & New York (NY): Mouton de Gruyter.

Winter, Werner. 1998. Lexical archaisms of the Tokharian languages. In V. Mair (ed.), *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*, vol. 1. Archeology, Migration and Nomadism, Linguistics, 347–357. Washington (DC): Institute for the Study of Man.

Woodhouse, Robert. 2015. Some observations on the putative dual reflexes of PIE **CRHC* in Greek and Armenian, Francis' law and Greek *αὐχήν* 'neck, etc.' *Studia etymologica Cracoviensia* 20(4). 257–272. doi:10.4467/20843836SE.15.017.2803.

Wright, Roger. 2011. Romance languages as a source for Spoken Latin. In James Clackson (ed.), *A companion to the Latin language*, 59–79. (Blackwell Companions to the Ancient World). Chichester & Malden (MA): Wiley-Blackwell.

Yakubovich, Ilya Sergeyevich. 2020. Doing things reverently among the Luwians. *Les Études Classiques* 88. 465–489.

Yamazaki, Yoko. 2014. Monosyllabic circumflexion or shortening? The treatment of the long vowels in the 3rd person future forms in Lithuanian. *Indogermanische Forschungen* 119. 339–354.

Yamazaki, Yoko. 2016. Monosyllabic Circumflexion in Lithuanian. Stockholm: Stockholm University PhD-dissertation.

Yediay, Fulya Eylem, Guus Kroonen, Serena Sabatini, et al. 2024. Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages. bioRxiv. doi:10.1101/2024.12.02.626332. https://www.biorxiv.org/content/10.1101/2024.12.02.626332v1 (21 March, 2025).

Yoshida, Kazuhiko. 1990. *The Hittite mediopassive endings in -</>ri<i>*. (Untersuchungen zur indogermanischen Sprach- und Kulturwissenschaft. Neue Folge 5). Berlin & New York (NY): de Gruyter.

Yoshida, Kazuhiko. 2011. Proto-Anatolian as a mora-based language. *Transactions of the Philological Society* 109(1). 92–108. doi:10.1111/j.1467-968X.2011.01252.x.

Yoshida, Kazuhiko. 2013. The mirage of apparent morphological correspondence: A case from Indo-European. In Ritsuko Kikusawa & Lawrence A. Reid (eds.), *Historical Linguistics 2011: Selected papers from the 20th International Conference on Historical Linguistics, Osaka, 25-30 July 2011*, 153–172. Amsterdam & Philadelphia (PA): John Benjamins.

Zair, Nicholas. 2018. On the relative sonority of PIE /m/. *Indo-European Linguistics* 6(1). 271–303. doi:10.1163/22125892-00601005.

Zgusta, Ladislav. 1957. Review of Birwé (1956). *Archív Orientální* 25(2). 306–307.

9. Appendix

9.1. Lexical isoglosses: The 44 unique verbal roots in *LIV*^{2}

1. b^{h}ag- ‘als Anteil bekommen’
2. 2.*deh$_1$- ‘aufspüren, (an)treffen’
3. dei̯h$_2$- ‘aufleuchten’
4. 2.*dei̯k̂- ‘werfen’
5. dens- ‘kundig werden, kunstfertig werden’
6. 1.*dreh$_2$- ‘(weg)laufen’
7. 1.*drem- ‘(wohin) laufen’
8. *d^{h}eu̯d^h- ‘erschüttern’
9. gres- ‘fressen, verschlingen’
10. g^{wh}reh$_1$- ‘etw. riechen’
11. Hei̯ĝ^h- ‘begehren’
12. Hi̯aĝ- ‘verehren’
13. h$_1$ei̯sh$_2$- ‘kräftigen; antreiben’
14. h$_1$ger- ‘erwachen’
15. h$_1$lei̯ĝ- ‘ins Zittern/Beben geraten’
16. h$_2$ei̯d^h- ‘schwellen’
17. h$_2$eld^h- ‘glücklich erreichen’
18. h$_2$elg^{wh}- ‘einbringen (als Erlös)’
19. h$_2$leks- ‘abwehren, schützen’
20. h$_2$merĝ- ‘abstreifen, (ab)wischen’
21. h$_2$res- ‘sich verweigern’
22. h$_3$ek^w- ‘ins Auge fassen, erblicken’
23. *h$_3$peu̯s- ‘sich mehren, reich werden an’
24. k̂emh$_2$- ‘müde werden; sich abmühen’
25. 2.*k̂erh$_2$- ‘brechen, zerbrechen (intr.)’
26. k̂rei̯H- ‘sich auszeichnen, vortrefflich sein’
27. k̂remh$_2$- ‘schlaff werden’
28. kamp- ‘krümmen, biegen’
29. keu̯d^h- ‘verbergen’
30. ksen- ‘Wolle krempeln’
31. melk^w- ‘behindern, schädigen, zerstören’
32. peth$_1$- ‘fallen’

33. (h$_1$)rek̂s- ‘schädigen’
34. seĝ^h- ‘überwältigen, in den Griff bekommen’
35. sep- ‘(richtig) behandeln, (in Ehre) halten’
36. sk̂^{h}eh$_2$(i̯)- ‘aufschneiden, schinden’
37. sneu̯- ‘(Flüssigkeit) hervorquellen lassen, tropfen’
38. sperĝ^h- ‘sich beeilen’
39. (s)reg/ĝ- ‘(sich) färben’
40. *tag- ‘ordnen, anordnen, aufstellen’
41. ti̯eg^w- ‘sich zurückziehen’
42. tu̯ei̯s- ‘erregen, erschüttern’
43. u̯ei̯k̂- ‘eingehen in, eintreten’
44. 1.*u̯el- ‘einschließen, verhüllen’

9.2. Lexical isoglosses: The 87 unique verbal stem formations in *LIV²*

9.2.1. The 22 verbal formations in *LIV²* unique to Indic, Iranian and Greek

1.	*dek̂-	acrodynamic root present	*dḗk̂-/dék̂-{3}
2.	*h$_2$merĝ-	acrodynamic root present	*h$_2$mḗrĝ-/h$_2$mérĝ-
3.	*b^{h}er-	amphidynamic root-present	?*b^{h}ér-/b^{h}r-
4.	*g^{w}ĝ^{h}er-	amphidynamic root-present	*g^{w}ĝ^{h}ér-/g^{w}ĝ^{h}r-
5.	*sep-	amphidynamic root-present	*sép-/sp-
6.	*k^{w}i̯eu̯-	éi̯e-causative–iterative	*k^{w}i̯ou̯-éi̯e-
7.	*h$_1$u̯eg^{wh}-	e-reduplicated athematic present	*h$_1$é-h$_1$u̯og^{wh}/h$_1$ug^{wh}-
8.	1.*dei̯k̂-	intensive	*déi̯-doi̯k̂/dik̂-
9.	*Hei̯ĝ^h-	i-reduplicated athematic present	*Hi-Héi̯ĝ^h/Hiĝ^h-
10.	*h$_2$merĝ-	nasal-infix present	*h$_2$mr̥-né/n-ĝ-
11.	*h$_1$ger-	perfect	*h$_1$ge-h$_1$gór/h$_1$gr-
12.	*sh$_2$ei̯-	perfect	*se-sh$_2$ói̯/sh$_2$i-
13.	*derk̂-	root aorist	*dérk̂-/dr̥k̂-
14.	*d^{h}er-	root aorist	*d^{h}ér-/d^{h}r-
15.	*h$_2$eld^h-	root aorist	*h$_2$éld^h-/h$_2$l̥d^h-
16.	1.*h$_2$er-	root aorist	*h$_2$ér-/h$_2$r-
17.	*i̯eu̯g-	root aorist	*i̯éu̯g-/i̯ug-
18.	*k̂u̯eh$_1$-	root aorist	?*k̂u̯éh$_1$-/k̂uh$_1$-
19.	1.*sek^w-	root aorist	*sék^w-/sk^w-
20.	1.*sek^w-	s-desiderative	*sék^w/sk^w-s-

21. *sk̂^{h}eh$_{2}$(i̯)-	zero-grade é-present	*sk̂^{h}h$_{2}$i̯-é-
22. *d^{h}g^{wh}ei̯-	zero-grade i̯é-present	?*d^{h}g^{wh}i-i̯é-

9.2.2. The 52 verbal formations in *LIV*^{2} unique to Indic and Greek

1.	*gres-	amphidynamic root-present	*grés-/gr̥s-
2.	*ksen-	amphidynamic root-present	*ksén-/ksn-
3.	*h$_{2}$u̯ers-	éi̯e-causative–iterative	*h$_{2}$u̯ors-éi̯e-
4.	*k̂emh$_{2}$-	éi̯e-causative–iterative	*k̂omh$_{2}$-éi̯e-
5.	*neu̯-	éi̯e-causative–iterative	?*nou̯-éi̯e-
6.	*(s)g^{w}esh$_{2}$-	éi̯e-causative–iterative	*(s)g^{w}osh$_{2}$-éi̯e-
7.	*h$_{2}$ei̯d^{h}-	full-grade e-present	*h$_{2}$éi̯d^{h}-e-
8.	*h$_{2}$leks-	full-grade e-present	*h$_{2}$léks-e-
9.	*seĝ^{h}-	full-grade e-present	*séĝ^{h}-e-
10.	2.*i̯eh$_{2}$-	i-reduplicated athematic present	*i̯i-i̯éh$_{2}$/i̯h$_{2}$-
11.	*nes-	i-reduplicated athematic present	*ni-nés/ns-
12.	*pleh$_{1}$-	i-reduplicated athematic present	*pi-pléh$_{1}$/pl̥h$_{1}$-
13.	*kel-	iterative	?*kol-éi̯e-
14.	*h$_{1}$ei̯sh$_{2}$-	nasal-infix present	*h$_{1}$is-né/n̥-h$_{2}$-
15.	*k̂emh$_{2}$-	nasal-infix present	*k̂m̥-né/n-h$_{2}$-
16.	*dek̂-	néu̯-present	*dek̂-néu̯/nu-
17.	*d^{h}enh$_{2}$-	perfect	*d^{h}e-d^{h}ónh$_{2}$/d^{h}n̥h$_{2}$-
18.	*d^{h}eu̯g^{h}-	perfect	?*d^{h}e-d^{h}óu̯g^{h}/d^{h}ug^{h}-
19.	*g^{w}eh$_{2}$-	perfect	*g^{w}e-g^{w}óh$_{2}$/g^{w}h$_{2}$-
20.	*g^{w}erh$_{3}$-	perfect	*g^{w}e-g^{w}órh$_{3}$/g^{w}r̥h$_{3}$-
21.	*k̂emh$_{2}$-	perfect	*k̂e-k̂ómh$_{2}$/k̂m̥h$_{2}$-
22.	*k̂lei̯-	perfect	*k̂e-k̂lói̯/k̂li-
23.	1.*k^{w}ei̯-	perfect	*k^{w}e-k^{w}ói̯/k^{w}i-
24.	*peh$_{3}$(i̯)-	perfect	*pe-póh$_{3}$/ph$_{3}$-
25.	*peth$_{1}$-	perfect	*pe-póth$_{1}$/pth$_{1}$-
26.	*b^{h}eu̯d^{h}-	root aorist	*b^{h}éu̯d^{h}-/b^{h}ud^{h}-
27.	*denk̂-	root aorist	*dénk̂-/dn̥k̂-
28.	1.*dreh$_{2}$-	root aorist	*dréh$_{2}$-/dr̥h$_{2}$-
29.	*d^{h}g^{wh}ei̯-	root aorist	*d^{h}g^{wh}éi̯-/d^{h}g^{wh}i-
30.	*h$_{3}$slei̯d^{h}-	root aorist	*h$_{3}$sléi̯d^{h}-/h$_{3}$slid^{h}-
31.	*k̂emh$_{2}$-	root aorist	*k̂émh$_{2}$-/k̂m̥h$_{2}$-

32. *k̑remh$_2$-	root aorist	*k̑rémh$_2$-/k̑rm̥h$_2$-
33. *k^{w}i̯eu̯-	root aorist	*k^{w}i̯éu̯-/k^{w}i̯u-
34. *nes-	root aorist	*nés-/n̥s-
35. *peh$_3$(i̯)-	root aorist	*péh$_3$-/pih$_3$-
36. *(h$_1$)rei̯k̑-	root aorist	*réi̯k̑-/rik̑-
37. *seĝ^h-	root aorist	*séĝ^h-/sĝ^h-
38. ?2.*deh$_1$-	s-desiderative	*déh$_1$/dh$_1$-s-
39. 1.*dreh$_2$-	s-desiderative	?*dréh$_2$/dr̥h$_2$-s-
40. 1.*k̑ei̯-	s-desiderative	*k̑éi̯/k̑i-s-
41. *mei̯k̑-	s-desiderative	?*méi̯k̑/mik̑-s-
42. *seĝ^h-	s-desiderative	*séĝ^h/sĝ^h-s-
43. *g^{w}ĝ^{h}er-	sigmatic aorist	*g^{w}ĝ^{h}ḗr/g^{w}ĝ^{h}ér-s-
44. *h$_1$u̯eh$_2$-	sigmatic aorist	*h$_1$u̯ḗh$_2$-/h$_1$u̯éh$_2$-s-{2a}
45. *h$_2$merĝ-	sigmatic aorist	?*h$_2$mḗrĝ/h$_2$mérĝ-s-
46. *mneh$_2$-	sigmatic aorist	*mnḗh$_2$/mnéh$_2$-s-
47. *h$_2$teu̯g/ĝ-	zero-grade i̯é-present	*h$_2$tug/ĝ-i̯é-
48. ?*h$_3$peu̯s-	zero-grade i̯é-present	*h$_3$pus-i̯é-
49. *melk^w-	zero-grade i̯é-present	*ml̥k^w-i̯é-
50. *meu̯k-	zero-grade i̯é-present	*muk-i̯é-
51. 1.*pek^w-	zero-grade i̯é-present	*pek^w-i̯é-
52. *sk̑^{h}ei̯d-	zero-grade i̯é-present	*sk̑^{h}id-i̯é-

9.2.3. The 12 verbal formations in *LIV*^{2} unique to Iranian and Greek

1. *h$_3$er-	éi̯e-causative–iterative	*h$_3$or-éi̯e-
2. *dens-	i-reduplicated athematic present	*di-déns/dn̥s-
3. *h$_1$ei̯sh$_2$-	i-reduplicated athematic present	*h$_1$i-h$_1$éi̯sh$_2$/h$_1$ish$_2$-
4. 1.*d^{h}reu̯-	iterative	*d^{h}rou̯-éi̯e-
5. *keu̯d^h-	nasal-infix present	*ku-né/n-d^h-
6. *h$_1$lei̯ĝ-	perfect	*h$_1$e-h$_1$lói̯ĝ/h$_1$liĝ-
7. *h$_3$rei̯H-	perfect	?*h$_3$e-h$_3$rói̯H/h$_3$riH-
8. *g^{wh}en-	reduplicated thematic aorist	*g^{wh}é-g^{wh}n-e-{9a}
9. *h$_1$nek̑-	reduplicated thematic aorist	*h$_1$é-h$_1$n̥k̑-e-
10. *perd-	root aorist	*pérd-/pr̥d-
11. *h$_3$ek^w-	s-desiderative	*h$_3$ék^w/h$_3$k^w-s-
12. 1.*med-	sigmatic aorist	*mḗd/méd-s-

Part 2

Article 1: “Indo-Greek” between terminology and topology

The following paper has been submitted in March 2025.

“Indo-Greek” between terminology and topology

“Rapporti ario-greci esistono, seppure non sempre sicuri e chiari”

(Bonfante 1976: 71)

Abstract

Across the literature, terms like *Indo-Greek* and *Graeco-Aryan* are widely used to describe an almost mythical relationship between the Indo-Iranian and Greek branches. This paper aims to shed more light on this relationship, but most importantly, on the terminology surrounding it. After an investigation of the use of the terms *Indo-Greek* and *Graeco-Aryan*, it is concluded that they are used almost interchangeably, but with various connotations. Frequently, they simply reflect a relationship of unspecified nature with no further justification, but when they are used as a label for a dialect group or a clade in a phylogenetic tree, they never signify an exclusive group consisting of just the two branches. This situation is compared to a survey of binary-branching Indo-European family trees (topologies), which finds no tree depicting an exclusive “Indo-Greek” clade. The paper concludes that while the terms are often used to describe a perceived paradigm, neither the terminology nor the topology backs the existence of such a school of thought.

1. Introduction

1.1. Background

It does not come as a surprise to any Indo-Europeanist that Greek and Indo-Iranian share peculiar and intriguing features, nor is it provoking to state that many scholars claim in vague terms that they share “numerous late isoglosses”, or mention in passing that we know as an established fact that they might go back to a subgroup younger than Proto-Indo-European. The widespread allegation of a close but undefined relationship is unhelpful and may be a hinderance for obtaining a more accurate understanding the actual disintegration of the family in linguistic and geographical reality.

The present paper covers and compares two main issues:

1. Terminology: What relationship is implied by the terminology surrounding Indo-Iranian and Greek (“Indo-Greek”, “Graeco-Aryan” and the like)

2. Topology: In published tree-like depictions of the Indo-European language family, what is the latest common ancestor of Greek and Indo-Iranian, and how many other branches descend from this proto-language?

1.2. Significance

Whatever the historical background of the many similarities between Greek and Indo-Iranian is, their relationship is of great importance to the reconstruction of Proto-Indo-European. Euler (1980: 174) operated with three scenarios of how the seemingly many unique Indo-Iranian and Greek similarities came about:

1. shared archaisms from PIE lost elsewhere
2. shared innovations in a common pre-stage of the two branches following the break-up of Proto-Indo-European
3. independent but parallel innovations

In a phylogenetic framework, Euler's second possibility is worth exploring because it has consequences for the reconstruction of Proto-Indo-European itself. Unresolved trees call for action: Reconstructions should only be projected as far back as the latest shared ancestor in order not to contaminate the reconstructed parent language with late or areal innovations, until the topology is established (Olander 2018; Goldstein 2022).

At branch level, all inherited features may be considered archaisms, but they could potentially be innovations at any preceding intermediate stage in the development of the language family. Thus, when Greek and Indo-Iranian share a great deal of features, these features need not – but can certainly – be inherited from Proto-Indo-European. Positive evidence from the two branches alone does not allow us to determine how far back the innovation occurred. The task of determining the nature of the relationship between these two branches is complicated by their early and abundant attestation because many other and more innovative branches do not allow us to determine when a feature was lost or innovated. This overrepresentation (data bias) may lead to an overly positive assessment of this relationship.

However, if the qualitative assessment of the shared isoglosses turns out not to find them insignificant for a subgrouping, any feature they share would have to be projected all the way back into the earliest parent language. If the last common ancestor of Greek and Indo-Iranian is indeed the root of the family tree, any shared feature would consequently have to have been lost independently several times on the way to all other branches. Such a negative conclusion would obviously be the final nail in the coffin to any "Indo-Greek" theory, yet it would be intriguing in itself as it would follow from it that the Brugmannian-style heavily Sanskrit and Greek biased reconstruction of Proto-Indo-European is not as far off as it has been assumed in recent years.

1.3. Delimitation

This paper cannot cover all proposed grammatical and lexical isoglosses as well as shared features in mythology and poetics, and it does it discuss the metalinguistic context, that is, the archaeological and genetic evidence complementary to the linguistic, possibly shared, history of the two branches. The paper also offers no complete research history on the linguistic relationship between Greek and Indo-Iranian, most of which is covered in other overviews and useful surveys (West 2007: 6–7; De Decker 2016a: 150–1). The older views and discussions are covered by Euler (1979: 18–23).[1] Marcantonio (2009: 19–21) also gives a short discussion but only quotes Pisani's (1933; 1940; 1974) works and Mallory & Adams (2006: [455]). Another line of research, in Spanish, follows Tovar (1977; 1979) and is visible in much of Adrados' work (see references below). While these works are very important for research in the "Graeco-Aryan" relationship, they do not necessarily employ the terminology under investigation here.

2. The terminological survey

2.1. What's in a name?

The terms *Indo-Greek*, *Graeco-Aryan*, and *Helleno-Indo-Iranian* as well as their mirror image equivalents are used to refer to the special relationship between Greek and Indo-Iranian. However, the nature of this special relationship varies greatly, and the terms are thus employed in many different functions. This is not a problem in itself. Further, it does not constitute a problem that different subdisciplines within Indo-European linguistics have different traditions of the exact nuance of the terms' meanings. It would, however, be problematic if conclusions get imported

[1] Although it will not be a complete research history, I here repeat the literature given by the beforementioned scholars (Kern 1858: 272–4; Grassmann 1863a: 85, 94, 109; 1863b: 119; Sonne 1863: 273; Kretschmer 1896: 168–70; Bonfante 1976: 92; Durante 1976: 18–30; Euler 1979; 1980; Clackson 1994; Gamkrelidze & Ivanov 1995: 794–5).

from one framework into the other without the realisation that they might rest on rather different premises.

For example, Fries (2021: 87 & n. 1), who writes on comparative mythology, refers to West, whose "chronology of the formation of the Indo-European language family is based on the most widely accepted views about early migrations into Greece, Anatolia, the Levant and towards the north of India", but he does add that "[c]omparison between Greek and Old Indian sources alone is usually assumed to lead back to the period of Graeco-Aryan (i.e. Greek and Indo-Iranian) unity before ca. 2300 BC". When consulting the reference, West's (2007: 6-7, n. 12 & 14) "most widely accepted" "Greco-Aryan unity" is only based on researchers subscribing to dialect geography (Kretschmer 1896: 168–170; Durante 1976: 18–30; Euler 1979)[2] and backed by a reference to Clackson (1994) whose focus is arguably elsewhere. This stratification is by no means universally accepted, as is evident in the reviews of West 2007, e.g. "das (freilich keineswegs allgemein akzeptierte) 'Graeco-Aryan'" (Janda 2012: 481).

There is no overall system behind the use of the different terms, and in many cases, they are completely synonymous (see 2.4 below). Therefore, I have searched for them all and sorted them according to meaning and context, not the terms themselves. That being said, some authors do distinguish between *Graeco-Aryan* and *Indo-Greek*, so that *Graeco-Aryan* labels a model for the reconstruction of the Proto-Indo-European verbal system (the tense-aspect system otherwise known as the "Cowgill-Rix" verb – as opposed to the "Jasanoff" or "Indo-Hittite" verb), and *Indo-Greek* labels a model of Indo-European phylogeny in which a relatively close relationship between Indo-Iranian and Greek is preferred over others (e.g. an *Indo-Slavic* model (Kroonen et al. 2022: 2)). Such a distinction can be found in the works of Olander, who uses "Graeco-Aryan" about the traditionally reconstructed Proto-Indo-European verbal tense-aspect system (Olander 2015: 296), and who suggests "Indo-Greek" as a term for the latest common ancestor of Indo-Iranian and Greek – as well as Armenian, Albanian and Balto-Slavic (Olander 2019; 2022: 191). Similarly, Marcantonio (2009: 20, 42) uses "Greco-Aryan" of the verbal system and "Indo-Greek" of an alleged close affiliation (Marcantonio 2009: 21).

[2] See also section 2.11

However, Ringe consciously and consistently uses *Indo-Greek* in both senses, and with clear reasoning: "I therefore follow Clackson 2007: 115 in naming the model after the most conservative daughters on which its reconstruction is based, altering his term 'Greco-Aryan' to eliminate the obsolete 'Aryan'." (Ringe 2017: 6, n. 1). Conversely, Martin West (2007: 6) consistently uses *Greco-Aryan* to describe a phylogenetic group. His influence is the reason behind the relative popularity of the phylogenetic use of the term *Gr(a)eco-Aryan* in comparative mythology and poetics – most scholars simply refer to him and leave the matter untouched (Allen 2007: 53; Fries 2021: 87; Elbourne 2012: 9, n. 9; Vergados 2013: 413; West 2014: 170).

2.2. What's not in a name

Most of the occurrences of the term *Indo-Greek* are entirely irrelevant for this investigation because they concern the Hellenistic kingdoms – and especially their monetary systems – in Southeast Asia. Likewise, in the discussion of the nature of the similarities between the Homeric Greek epics and the classical Sanskrit Mahābhārata in comparative literature, there is a debate between adherents of two different *Indo-Greek* theories: inheritance from a common ancestor vs. Hellenistic influence. They never seem to be used by the same authors or in the same text. This debate can be found in, for example West (2007: e.g. 133) and Sheldon (2009: 527–8) as well as between Wulff Alonso (2014) and his critics (Allen 2015; Phillips-Rodriguez 2016; Pisano 2015).

2.3. Methodology of the terminological survey

Other than gathering all relevant references I have come across over the past few years of research, I have systematically searched for the terms in English and their equivalents in German – with unsystematic supplements in other languages when the references occurred in these texts. I have searched for the terms using various interlinked library databases (chiefly The Royal Library, Copenhagen; which links among others to the databases of De Gruyter, Brill, Oxford Academic and JSTOR), Google Scholar and Google Books. In order to get as much relevant material as possible, I searched for various spelling variations, including unorthodox oscillation between *o* and *a* expected from the quality of the OCR of some resources, with and without hyphens and disregarding accents and differences in upper and lower casing, for the following variants as first members of the compounds: *graeco-*, *græco-*, *gräco-*, *greco-*; *helleno-*; *indo-, indo-ario-, indo-ariano-, indo-aryano-; ario-*, *aryo-*, *ariano-*, *aryano-*, combined with the second members *-aryan, -arian, -arisch; -indo-iranian, -indo-iranisch, -indo-arisch; -greek*, *-griechisch, -grec-* and *-hellen-*[3]. Combinations of these, variations of *Indo-Greek*, *Graeco-Aryan* and *Helleno-Indo-Iranian*, are what I refer to as "the terms". I purposely excluded the true dvandvaic collocations (e.g. *Greek-Indo-Iranian* and *arisch-griechisch*) since these in their nature need not imply anything further.

This search yields an impossible amount of hits, and it is neither possible nor necessary to exhaust them all here. It also leads to many doublets. I collected relevant quotes and sorted them according to meaning and scientific context. It is indeed possible that I have missed valuable publications or other definitions, but for the purpose of investigating the use of terminology across the field, the examples presented below should be adequate and representative.

2.4. The meanings of "Indo-Greek" and "Graeco-Aryan"

In my search for the definitions and use of the terms, I have come across the following different uses of the near-synonymous terms:

a. A label for a model of the reconstructed Proto-Indo-European verbal system
b. A *dvandva*: Indo-Iranian (or Indic) as well as Greek, just like a sweet-and-sour sauce
c. A purposefully vague denominator for an underspecified connection open for interpretation, typically about isoglosses or in poetics and mythology
d. A label for the last shared state of Indo-Iranian and Greek, which is probably but not necessarily different from Proto-Indo-European proper
e. A defined phylogenetic or dialectal group consisting of Indo-Iranian and Greek

The most isolated among these is (a). The items (b), (c), (d) and (e) are rather similar, but I have distinguished them according to these following principles. Uses of the terms stripped of connotations are grouped under (b). The groups (c) and (d) encompass the broad range of (intentionally) vague uses, where a close relationship is insinuated, but not elaborated upon. I have distinguished (d) from (c) by whether the underspecified use of the terms is overtly recognised or not: Under (c), it is not, and under (d) the terms are accompanied by an "at least" (or

[3] To include Romance and English at once.

similar). Finally, when used about a phylogenetic subgroup or dialectal or areal unity with named member branches, they have been treated under (e). Those will be the most important for the comparison with tree topologies in the second part of this paper.

2.5. (a) A label for a model of the reconstructed Proto-Indo-European verbal system

The terms, and especially *Graeco-Aryan*, are frequently used as a label for a model of the reconstructed Proto-Indo-European verbal system. Apparently, only Ringe (2017: 6, n. 1) consequently uses Indo-Greek in this context. Although very frequent, this use is not strictly relevant for this paper.[4] This meaning of the term is more relevant to the earlier stages of the Indo-European verbal system, and it says less about how close Indo-Iranian and Greek are to each other in the diversifying language family. Quite obviously, they both descend from this system no matter its age. However, in discussion of innovations and archaisms of the Indo-European verb, the distinction between a Graeco-Aryan model of the verb and an innovative Indo-Greek subgroup dwindles, e.g.:

- "The 'Central' subgroup includes Germanic, Balto-Slavic, Indo-Iranian, Armenian, Greek, and probably Albanian; its internal subgrouping is still very unclear, though it seems likely that Indo-Iranian, Balto-Slavic, and Germanic were parts of a dialect chain at a very early date. Note the implications of this phylogeny for the reconstruction of the PIE verb. The Indo-Greek verb is a reasonable reconstruction of the system for 'Proto-Core IE', and can even account for much of the 'Proto-Nuclear IE' system; it is only for the ancestor of the whole family that it is seriously inadequate." (Ringe 2017: 6–7)

and, but perhaps more polemic,

- "Los mas conservadores la rechazan, empeñados en acomodar a las lenguas anatolias en el lecho de Procrustes del indo-griego" (Villar 1996: 303) ["The most conservative [researchers] refute it [Indo-Hittite], insisting on accommodating the Anatolian languages in the Procrustes' bed of Indo-Greek"]

Similarly, the Graeco-Aryan model is frequently invoked in discussions of the prehistory of the verbal systems of "non-Indo-Greek" branches, which makes the

[4] Thorough discussions can be found in Clackson (2007: 115–51) and Willi (2018: 24–57)

distinction between model of the reconstructed verb and model of reconstructed family tree fade. The following examples are on Celtic, where the title of Rix's (1977) paper, "Das keltische Verbalsystem auf dem Hintergrund des indo-iranisch-griechischen Rekonstruktionsmodells", is indeed telling, and Germanic:

- "Es ist aber bisher nicht gelungen, von anderen idg. Sprachen aus ein alternatives Rekonstruktionsmodell zu entwerfen, das die einzelsprachlichen Verbalsysteme im ganzen besser verstehen läßt als das graeco-arische Modell." (Rix 1977: 132)
- "Daraus folgt, daß die graeco-arische Grundlage des germanischen Verbalsystems einer Ergänzung durch die Berücksichtigung der Verhältnisse in den dem Germanischen benachbarten idg. Dialekten bedarf." (Udolph 1995: 301)

This use is not just restricted to the verbal system, though it has its origin and is by far most popular here. Other aspects of grammar can be described using the terms; e.g. syntax (Koch 1991: 2), the nom. sg. ending of the h_2-stems (Kortlandt 2019: 144) and accentology (Kapović 2016: 56).

2.6. (b) A dvandva: Indo-Iranian (or Indic) as well as Greek

Rarely, *Indo-Greek* can be used stripped from any additional assumptions, simply meaning *Indo-Iranian* (or *Indic*) *as well as Greek*, like in the following discussion of a definition of Romani:

- "'Indo-Greek origin' appears to refer to the recognition that the core of the Romani language contains, alongside a majority of Indic lexical roots and grammatical inflections, also a strong element of Byzantine Greek origin" (Matras 2015: 308)

The term can also be used about more conventionally *Indo-Iranian as well as Greek* topics like:

- "[…] Indo-Greek accentuation of privative syllables […]"(Gray 1925: 122, n. 22)

In the latter quote, it could be argued that Gray means to imply a specific situation in the latest shared ancestor of the two branches, but given that these branches preserve the clearest evidence on the matter, it should rather be seen as an agnostic statement of the facts.

Similarly, in the following, only the wider context makes it clear that “Indo-Greek” does in fact mean *Indo-Iranian as well as Greek*, yet the entire paper does leave an aftertaste of something closer being insinuated:

- “[…] some of these common Indo-Greek pictorial motifs are present also on the most ancient anthropomorphic stelae of the North Pontic region, as well as on the stelae of similar types from the other regions of Eurasia.” (Vassilkov 2011: 204)

In Indo-European poetics, it is common to speak of “Indo-Greek comparison” as a simple matter of fact (e.g. Andrianne 2013: 310; Allen 2015: 242). This can leave it unclear if the authors presumes any special cultural, dialectal or even phylogenetic affiliation between the two, or if the compared features are ultimately inherited, universal or coincidentally parallel unless it is followed by a caveat:

- “However, we must not forget that Indo-Greek comparativism will one day have to take its place in a much fuller IE comparativism” (Allen 2019: 227–8, n. 23)
- “Our picture, therefore, will be largely based on the Indo-Iranian and Greek evidence, with sporadic contributions from elsewhere. If it were not for these latter, our conclusions would be valid only for the Graeco-Aryan level.” (West 2007: 305)

Comparative mythology, religion and poetics are very well represented in the survey of the use of the “Indo-Greek” terminology for exactly this reason.[5]

2.7. (c) A purposefully vague denominator for an underspecified connection

2.7.1. An alleged “model” in Indo-European phylogenetics

Although the issue of the “Indo-Greek” data bias in comparative mythology is very clear and acknowledged by Allen and West above, the exact phylogenetic approach to reconstruction should have much a wider application. The terms are occasionally

[5] This is not the place to discuss shared motifs, poetic language and phraseology in detail, but the role of Indo-Iranian and Greek in traditional Indo-European reconstruction is very symptomatic, cf. the prime example of Kuhn’s (1853) unforgettable _*k̂léu̯os ń̥-d^{h}g^{wh}itom_ ‘imperishable fame’ (Schmitt 1967: 62–4; West 1988: 153; Watkins 1995: 173–8; Jackson 2006: 14).

used to describe a model of the dispersal of the Indo-European languages. While it can be useful to distinguish between different competing hypotheses, there is a certain risk of writing something into existing by assigning a name to it. Noticeably, none the "Indo-Greek" hypotheses in the following examples are accompanied by references, framing the "common knowledge" status, although the details remain unclear:

- "Within core Indo-European, various rival models exist, including primarily those prioritizing a Graeco-Indo-Iranian ("Graeco-Aryan") subnode versus a Balto-Slavo-Indo-Iranian ("Indo-Slavic") subnode, with Albanian and Armenian as their satellites. Without a generally established phylogeny, the identification of suitable archaeological and genetic proxies for the prehistoric locations and movements of the various Indo-European speech communities, itself a highly challenging endeavor, is all the more treacherous." (Kroonen et al. 2022: 2)
- "[...] Greek and Indo-Iranian share so many traits that the term Greco-Aryan (or Indo-Greek) is sometimes used to describe their relationship" (Anthony & Ringe 2015: 207)
- "The idea of the existence of some sort of 'Indo-Greek branch / unity' is not totally new, having already been proposed, for example, by Pisani (1933, 1940 & 1974). As is known, this idea was mostly rejected at the time, since there are several other grammatical (as well as phonological) isoglosses that cut across such an assumed branch" (Marcantonio 2009: 21)

In the final example, the references to Pisani are do not actually deal with phylogenetic grouping but with various areal isoglosses.

2.7.2. On ways of innovating

More frequently, the terms are not used about a hypothesis of phylogenetic subgrouping but are applied to traits or innovations in a much more subtle context in which the reader is left with the impression that the author has a specific node or dialect group in mind, but its configuration remains unmentioned. Below are some representative examples:

- "Elements of the all-IE (compact) vocabulary, marginal elements and Graeco-Aryan innovations; these can be associated with the Pelasgic stratum." (Szemerényi 1964)

While this use is by far the most common use across the literature and perhaps rarely causes confusion, notice how Tocharian must be included in Graeco-Aryan in the first example below, but excluded from it in the second one:

- "This has to be taken into consideration when we now turn to the Genus femininum in Graeco-Aryan." (Lühr 2012: 188)
- "Accordingly, the huge number of verbal stem classes found in Tocharian may be seen as the consequence of a conspiracy of two different tendencies to be found in the verbal system of this branch, actually a tendency to innovate the Graeco-Aryan way and a tendency to innovate the Anatolian way." (Malzahn 2012: 239)

2.7.3. On specific isoglosses

The underspecified use of the terms is very common in connection with specific grammatical isoglosses – a topic I will explore separately elsewhere. A few examples of the application of the vague "Indo-Greek" terminology to specific isoglosses include: Grassmann's law (discussion in De Decker 2015; Kiparsky 1973; Miller 1977: 381), "die Erscheinung dass *-meno* und *-to* zugleich und beisammen als Exponenten eines part. perf. pass. gelten" (Ascoli 1893: 555), the perfect (Pisani 1933: 570, 637) or obligatory reduplication of the perfect (Bonfante 1976: 71), the genitive in *-*osi̯o* (Bonfante 1976: 73 with Arm.), the augment (Hartmann 1979: 56; Zahn 2014: 119; discussion in Goldstein 2022: 81), adjectival as well as nominal "Ableitungen auf *-tero-* und *-tm̥*^{m}*o-*" (Euler 1979: 252), the element *-*me* of the plural personal pronouns (Kloekhorst 2008: 116), "so-called "passive aorists"" (Pooth 2009: 241), "die Verallgemeinerung von [rel. pron.] **(H)i̯o-*" and extension of "der Ausgang *-*o-Hom* von den Thematica auf andere Klassen" (Hajnal 2011: 21, n. 16; also Sommer 2016), the ablaut pattern of the *s*-stems in **CéC-os*, **CéC-es-os* (Kloekhorst 2013: 121, n. 35), the reduplicated aorist (Birwé 1956: 29–30; Bendahman 1993: 245; Willi 2018: 104); and finally the loss of "the long-vowel preterite (aorist)" and "morphologically middle perfects" (Willi 2018: 114, 219).

Such isoglosses may of course also be lexical. Many unique lexical correspondences have been treated by Porzig (1954: 157–61) and Euler (1979); but the application of the terminology is much wider than that. Below are a few examples of proposed "Indo-Greek" lexical correspondences: The word for 'thousand' containing *-*ĝ*^{h}*esl-* (Bonfante 1976: 78), "[t]he name of the 'quail'" (Ved. *vártikā*, Gr. (*ϝ*)*ὄρτυξ*) (Polomé 1989: 220); the word family *κρατερός* and *śrathayati* (De Decker 2011: 92–2 & n.

9), **mak*^{h}- ‘fight’ (De Decker 2016b: 86), *h_2*u̯ed*- ‘to sing’, (West 2007: 33), the use of **mr̥to*- ‘mortal’ for ‘man’ (West 2004: 54, 65; 2007: 127–8), “a Graeco-Aryan isogloss where Grk *despótēs* master, lord’ and Indo-Iranian (e.g. Skt *dámpati*- ‘master’ derive from a compound **dems-pot*- ‘master of the house’”(Mallory & Adams 2006: 209), “a reduplicated deponent present **i̯i-i̯*$_{ə/i}$$h_2$*-toi̯* ‘seeks, requests’ at least for late inner-Indo-European/pre-Graeco-Aryan” (Grestenberger 2016: 122), “ancient derivatives in *-yo- from the root *ek̂^{h}w- [i.e. *h_1*ék̂u̯-i̯o*-]” (Gamkrelidze & Ivanov 1995: 463), the word “*ték̂tsō [i.e. **tétk̂on*-] ‘carpenter’” (Ringe 2010: 336) and many others. However practical it is to describe these lexical correspondences as “Indo-Greek” in the dvaindvaic sense, they tell very little about the relationship between the two branches without further analysis. Should they be interpreted as genetic innovations, coincidental archaisms or even shared or parallel borrowings? Notably, Armenian at least partakes in the semantic shift of **mr̥to*- from ‘mortal’ to ‘man’, cf. *mard* ‘man’; but also Arm. *tēr* ‘master’ < **ti-ayr*- is probably a substation of **dems-potis* with **dems*-h_2*nḗr* (Olsen 1999: 671), making this isogloss, too, non-exclusive.

2.8. (d) A label for the last shared state of Indo-Iranian and Greek, which is probably but not necessarily different from Proto-Indo-European proper

In cases where the evidence – or rather the lack thereof – does not allow us to judge how far back a feature should be projected, the very uncertainty about the relationship can be acknowledged by means of modifications of the “Indo-Greek” terminology like “at least”, “zumindest” or “mindestens”. This latest common ancestor could be the same as Proto-Indo-European, from which all other branches also descend, or it could be a later node only ancestral to a subsection of the branches. The point is summed up nicely here:

- “[…] Graeco-Aryan, which does not indicate a special branch of Indo-European but a pattern of isoglosses that we may feel cautious about assigning to full Proto-Indo-European antiquity without additional evidence.” (Mallory & Adams 2006: 110)[6]

[6] Though the term “Graeco-Aryan” is reserved for a relationship of this sort, a “linguistic continuum” based on “a number of innovations” is mentioned elsewhere (Mallory & Adams 2006: 455).

Such modification of the terms is very frequent in comparative mythology where the overrepresentation of Greek and Indo-Iranian data by comparison to many of the other Indo-European branches which do not preserve the relevant linguistic material let alone the rich mythologies. Therefore, any mythological comparison runs the risk of resting on *argumenta ex silentio* or may be skewed:

- "Eine graeco-arische Isoglosse gibt noch keinen Aufschluss über eine gemein-indogermanische Grundlage, sondern höchstens über eine graeco-arische Gemeinsamkeit. In diesen Fällen ist also auch Sprachmaterial außerhalb dieser geografischen Zone zu orten und genau zu differenzieren, in welchen Traditionen bestimmte Vergleichspunkte belegt sind." (Meusel 2019: 58)

It is by no means reserved for the realm of comparative mythology; it is only that the problem of data bias is so much clearer there. Specific isoglosses like the ones quoted above may also be presented with such caution:

- "Aufgrund der indoiran.-griech. Gleichung *εἶπε* : *vocat* : *vaocaṯ* muß **u̯e-ukʷ-(e/o)-* zumindest für die graecoarische Spracheinheit des Uridg. postuliert werden." (Bendahman 1993: 41)
- "The question whether the late (at least Graeco-Aryan) proto-language already possessed a pluperfect has been controversially discussed by many" (Willi 2018: 221).

2.9. (e) A defined phylogenetic or dialectal group consisting of Indo-Iranian and Greek

Finally, and most importantly for the comparison with family tree analyses, the terms Indo-Greek and Graeco-Aryan can be used to describe a phylogenetic subgroup or node in a family tree, oftentimes in practice indistinguishable from a "late dialect group" or the result of prehistoric convergence. This node is, of course, the latest shared ancestor of Indo-Iranian and Greek but very often also of other – named – branches. In fact, I have not found anyone explicitly claiming that the "Indo-Greek" or "Graeco-Aryan" clade consists of these two branches only. It should be duly noted that this lack of findings need not be contributed to more than my methodological choices or categorisation efforts: I have intentionally only mentioned those that overtly recognise the descending branches of the "Indo-Greek" node here and treated less obvious connections elsewhere. Quite possibly, some of the terminology quoted above may intend to cover a straightforward definition of "Indo-Greek" as Indo-Iranian and Greek. Another reason for it could be that the special relationship is described without the terms surveyed here (e.g. Mallory & Adams 1997: 555; Gąsiorowski 1999: 55–6; Yakubovich 2011: 227; Elbourne 2012: 6, 9 n. 9).

2.9.1. Indo-Graeco-Armenian

The closest instances of an application of a narrow and exclusive Indo-Greek subgroup are cases like the following:

- "some scholars (e.g., Olander 2019) would even agree that, in view of the numerous innovations shared by Ancient Greek and Indo-Iranian, the latter may be traced back to a later common sub-ancestor (aka "Graeco-Arian" or "Indo-Greek")" (Ginevra 2022: 19)

Ginevra seems to imply that Olander is among the scholars who recognise such an exclusive subgroup of Greek and Indo-Iranian based on their shared innovations, however, Olander's "Indo-Greek" is the ancestor of Balto-Slavic, Armenian and Albanian as well. Similarly, in the following instance, there is reason to dive into the references:

- "since Greek and Sanskrit are often traced back to a single sub-group (Fortson 2010: 203), lower sonority of /m/ is as likely to have developed in Graeco-Indo-Iranian as in PIE." (Zair 2018: 298–9).

The refence to the alleged "Graeco-Indo-Iranian" subgroup is very much taken out of context, and what Fortson actually states is by no means as straightforward as the impression one could get from Zair's wording: "It is widely thought that Indo-Iranian forms a subgroup with Greek, Armenian, and Phrygian" (Fortson 2010: 203). This is quite a common definition. In most other instances where a named Indo-Greek node is defined, Armenian is included with quite a bit of certainty as a descendent of the last shared ancestral language of Greek and Indo-Iranian (Adrados 1979: 272; McCone 1991: 62; Bendahman 1993: 245; Grigoriev 2002: 342; Parpola 2005: 3; Kallio 2012: 226; Martirosyan 2013: 105; Mallory 2013: 146), even by Euler who is so often referred to for the special relationship between Greek and Indo-Iranian (though without the exact terms under investigation here): "Das Armen. wird zweifellos zu Recht als Nächstverwandter des Gr. angesehen" (Euler

1979: 26).[7] Sometimes, the scarcely attested Phrygian is included alongside Armenian in the Indo-Greek clade (West 1994: 807; 2007: 6–7, 9; Fortson 2010: 203). In most of the authorship of Adrados, Armenian, Phrygian and Thracian belong to the Indo-Greek group (Adrados 1974: 857–9; 1979: 271–5; 1991: 15; 1980), but occasionally, also Macedonian and Illyrian "most certainly" belong to it (Adrados 1981: 15). Balto-Slavic (Ringe 2010: 336; Adrados 1980: 7) or even Balto-Slavic and Albanian are cautiously said to descend from this "Indo-Greek" proto-language (Olander 2018: 240–1; 2022: 191).

2.10. Archaism, innovation and the lack of evidence

While overt and narrow definitions of "Indo-Greek" are hard to come by, the discussion of the issue and the consequences of our interpretations are, of course, not. Not only is the question of the relationship between Greek and Indo-Iranian complicated by the lack of decisive evidence in other branches; it is also complicated by the fact that so much Indo-European reconstruction is based on these two branches.

Many of the shared traits are often seen as archaisms – and thus not constituting any evidence in favour of a subgroup – by some authors, whereas others judge the very same evidence as significant shared innovations. For instance, Rix (2009: 8) rejects Greek descending from a subgroup with other members, whereas Sihler (1995: 485) argues for innovations shared among the descendants of an intermediate proto-language. Similarly, van Beek (2022: 196) sees most "Indo-Greek" similarities as archaisms, whereas Bičovský (2017: 15, n. 6) claims that the "modern view" is that the very same similarities are innovations of a shared "Graeco-Aryan" (i.e. non-Anatolian) group.

Even Birwé (1956) and Euler (1979: e.g. 171, 179), who are very important figures in systematic Indo-Greek comparison and are often cited for Indo-Greek parallels, are not overly optimistic:

- "Die sekundär ausgebildeten griechisch-arischen Übereinstimmungen rechtfertigen keine griechisch-arische Verwandtschaft. Sie finden vielmehr ihre

[7] It is impossible to cover the hotly debated topic of the position of Armenian here, but there is plenty of literature on the matter (e.g. Bonfante 1976; Leroy & Mawet 1986; Clackson 1994; Martirosyan 2013; Kim 2018; Olsen & Thorsø 2022).

> Erklärung durch eine Nachbarschaft der Vorfahren der Griechen und Arier in vorhistorischer Zeit." (Birwé 1956: 70)

Naturally, this conclusion raises the question of when, how and where this contact took place and what evidence we have in favour of it. After all, contact is a real linguistic and archaeological phenomenon that leaves traces, and such a claim cannot be accepted without further evidence and argumentation. I hope to return to this matter elsewhere.

Assumptions of phylogenetics are deeply interconnected with the reconstruction of Proto-Indo-European. This is very summed up very well by Ringe in his discussion of **tétḱon-* mentioned above:

- "It seems to follow that 'carpenter' either was created just before the ancestors of IIr. and Greek lost contact with each other, or that it was created even later and borrowed from pre-IIr. into pre-Greek [i.e. pre-proto-Greek, not the alleged non-Indo-European substrate language] before any further relevant sound changes had occurred. In the former case the word is potential evidence for a relatively small clade including Greek and IIr. – an "Indo-Greek" clade, though of course it might include other daughters that have simply lost the relevant lexemes (most plausibly Balto-Slavic and Armenian). In the latter case the word is evidence for dialect contact at an early post-PIE period." (Ringe 2010: 336)

Distinguishing between archaism and innovation in the "Indo-Greek" branches is impossible based on the available evidence of these well-attested branches, as we very rarely have positive and unique evidence in favor of a narrow Indo-Greek clade *to the exclusion* of other branches. We would need to know that these were not included, preferably by sharing innovations with other branches. When traits are lost, so is the possibility for complete arguments. Ironically, the matter of determining whether the two best attested Indo-European branches descend from a common node hinges on the analyses of some of the latest attested or most innovative branches.

2.11. Waves, trees, groups and nodes

While I have tried to isolate definitions of an alleged "Indo-Greek" clade, most overt uses of the terms are not directly concerned with subgrouping in a strict – phylogenetic – sense. This is not the place for a discussion on models of linguistic diversification, but the different approaches do have implications for the definition of the very concept of a "group", and thus the nature of "Indo-Greek".

Birwé, Porzig, Pisani and Bonfante all wrote in a tradition of dialectology inspired by Schmidt's (1872) wave theory and Krahe's dialect geography; Euler and Adrados further subscribe to Meid's (1975) famous time-space-model. The groups of the Krahe–Meid frameworks are not directly transferable to phylogenetic argumentation. The same evidence that may be analysed as significant innovations underlying what some researchers interpret as bifurcating splits, need not be explained at the expense of other, later or just as significant innovations across the linguistic area under the Meid model. This further allows the for the incorporation of much more material that do not meet the somewhat stricter criteria of a potential shared innovation in a phylogenetic framework. I do not wish to imply that there is no such thing as horizontal, areal spread of features or borrowing of lexicon (and to a lesser extent also grammar), I will simply reiterate that such traits reveal groupings in time and geography, not necessarily in topology. Accordingly, their nomenclature cannot be transferred directly.

This discrepancy need not be as grave as one could think at first glance. It is widely accepted that the wave and tree models are in fact not incompatible, but that they focus on different aspects of linguistic prehistory (Drinka 2013). The models might even complement each other within the same scenario, but this is not always recognised, e.g.:

- "This term [inner Indo-European] should be understood as a loose cover term and may ultimately not stand for a separate node in the family tree but a 'dialect continuum' (Anthony and Ringe 2015)" (Grestenberger 2016: 99, n. 1).

From this explanation, it is clear that Grestenberger views nodes and dialect continua as inherently different and incompatible under the same model. This assumption is backed by the reference to Anthony and Ringe, who do indeed work within a phylogenetic framework, but Ringe does not consider the two types mutually exclusive: A dialect continuum may very well *descend* from a node within a larger tree topology. This is exactly how Ringe (2017: 7) depicts "Inner-Indo-European": as a node whose internal relationship may behave like a dialect continuum within a binary-branching tree structure. While there may not be a recoverable underlying tree-structure beneath the diversification of this dialect continuum, its members all descend from a distinct intermediate ancestor which is not the ultimate parent language of the entire family.

Such a "hybrid" family tree need not be more than a graphic depiction of uncertainty. It could signify that the internal subgrouping of the descendants of such a note has not (yet) been demonstrated with our current methods or the available material. Alternatively, but harder to prove, such a hybrid tree model can may be the result close contact following a binary split forming a dialect continuum whose internal structure is so interwoven that imposing a tree-structure on it will be meaningless.

2.12. Summing up

While it is necessary to disclaim that the quotes presented above are of course given outside of their wider context, the above survey is sufficient to reveal some trends: Although the application of the "Indo-Greek" terminology is very frequent and often implies a special relationship between the branches, the use of the terms is not uniform and never exclusive. Indo-Greek and Graeco-Aryan are used (often interchangeably) to denote (a) a reconstructed model of the Proto-Indo-European verbal system, (b) a dvandvaic collocation of the two branches, (c) a more or less purposefully vague description of an unclear but special relationship, (d) the last shared ancestor of the two branches no matter how many others descend from it, and finally (e) a specified subgroup. Most importantly, I have found *no examples* of authors using the terms explicitly to denote a narrow subgroup consisting of only Indo-Iranian and Greek. When the terms are used to describe a group – whatever its nature – Armenian is almost always included, sometimes also Albanian, and even Balto-Slavic. This paradox stems from the fact that much of the comparative work on the branches which has been undertaken in a Schmidt-Krahe-Meid framework is freely quoted without further comments in more phylogenetically oriented works.

3. The topological survey

3.1. Why collect trees

The previous sections have been concerned with nomenclature and definitions of terms. It has proven difficult to find clear-cut uses of the terms investigated to describe an exclusive or even well-defined subgroup. In the following, I will contrast the use of the terms in prose with graphical depictions of family trees. While the approaches may vary greatly, and many of the authors quoted above certainly have no wish for their work to be reduced to a comparison with trees, the contrast is nevertheless interesting: Prose does leave more room for interpretation and conflicting conclusions than graphical depictions of family trees. Some hold that model is an oversimplification of the complex reality of linguistic diversification, but it does have the valuable advantage that one solution must be preferred over the others. At any rate, the widespread claims of a somehow close or special relationship described in words could indeed benefit from the clear-cut clarity of the family trees.

3.2. Seeing the forest for trees

The number of mathematically possible trees is not just far greater than the number of trees examined in the survey below, it is also astronomical. In mathematical terms, we can only draw a single rooted binary-branching family tree if the number of branches is two: Staying with the "Indo-Greek" example, Indo-Iranian and Greek can only form one binary-branching tree with a root. Graphically, this could, of course, be depicted in two ways: One with Greek on the left, and one with Indo-Iranian on the left; but topologically, these trees are identical. Taking one extra branch into account gives us three possible trees: In our example, this other branch could be Armenian, which could be a sister of Greek, a sister of Indo-Iranian or a sister of a branch of "Indo-Greek" proper. A fourth branch, e.g. Albanian, would yield 15 possible trees, and a fifth 105, a sixth 945 and so forth (Felsenstein 1978: 31). For baffled linguists, the formula to find the number of possible rooted bifurcating trees is $(2n-3)!!$ where n is the total number of leaves (i.e. branches or languages).

For this reason, Greenhill and Gray (2021: 228) argue for the superiority of Bayesian methods over traditional assessments of family trees:

- "[T]raditional historical linguists do not use an explicit optimality criterion to select the best tree, nor do they use an efficient computer algorithm to search for the best tree. This is *surprising* given that the task of finding the best set of trees is inherently a combinatorial optimisation problem of considerable computational difficulty. In abstract mathematical terms, for just five languages there are in theory 105 ways of subgrouping them in a rooted bifurcating tree. For ten languages this number grows to over 34 million possibilities." (my emphasis)

While this is of course true in a strictly theoretical or epistemological sense, it is rather misleading. I find it unnecessarily polemic to miscredit traditional methodologies by insinuating that each and every mathematically possible tree is equally likely and that they all have to be evaluated on equal terms. This would only hold true if there were no systematic phylogeny in the first place. I would argue that is *not* "surprising" that traditional historical linguists do not give as much time and energy the millions of trees that are not backed by any qualitative linguistic arguments. Among the 34.459.425 possible rooted bifurcating trees when the number of taxa is the ten widely recognized Indo-European branches, [8] most will depict rather absurd topologies that would never be proposed on the basis of qualitative evidence. At the very least millions of trees could be excluded with the same single argument: If a close relationship between two branches, e.g. Italic and Celtic, is accepted, the total number of possible trees drops as dramatically as it rose in the first place.

That said, the quantitative methods do have some clear advantages: First, the overarching approach of treating all subgrouping issues of the entire family as interconnected makes sure that no grouping is based on positive evidence alone, and that possible counterevidence is clear. Second, all possible trees can be directly compared through numerical scores on how well they fit the input data. Whether or not a quantitative score is always better than a qualitative argument is irrelevant for now, but the equal treatment of data from all input branches should serve as a model. Following the plea of the same authors (Greenhill & Gray 2012: 534; Greenhill, Heggarty & Gray 2021: 228) to view quantitative methods as a supplement to existing methods, I find it justified to compare trees that are the products of many different methods.

3.3. Collecting and comparing Indo-European family trees

In the following, I will present a survey of the relative position and composition of the latest common ancestor of Indo-Iranian and Greek, labelled "Indo-Greek" for short. These trees will be analysed and grouped according to the number of other branches also descending from this latest "Indo-Greek" node. Naturally, this ties into many ongoing debates, such as the existence of an "Armeno-Greek" clade, an "Albano-Greek" clade, the questions on a "Balkanic" subgroup (Greek, Armenian

[8] $(2 \cdot 10-3)!! = 34.459.425$

and Albanian), the reality of the Satəm languages as a genetic group (Indo-Iranian, Balto-Slavic, Armenian and Albanian), and an exclusive "Indo-Slavic" group (Indo-Iranian and Balto-Slavic). Such a direct comparison of family trees produced under many different frameworks can tell us little about the actual relationship between the branches, but it can indeed present the state of the scholarship.

I gathered as many Indo-European language family topologies as possible, including anything from tabulations to drawings of trees with branch names. I excluded models that are uninformative about the relative position of Indo-Iranian and Greek. These were chiefly true "starburst" models, where all branches disperse from Proto-Indo-European at the same time, and derivatives of such, in which Anatolian – sometimes also Tocharian – branch off before a starburst of the "core" branches. I removed duplicated trees by the same author and reprinted trees with clear references to already included works. This amounted to a compilation of 41 more-or-less binary branching trees, which is, of course, a ridiculously small subset (0,001‰) of the mathematical possibilities, but nevertheless a representative sample of the propositions across the literature.

3.4. Pruning and grafting

The trees collected were not produced with direct comparison in mind. Not only do the methods and purposes vary, but so does the number of branches. Some trees were published before the discovery of major branches. Others were simply presented to give a brief overview of the language family and may deliberately leave out the branches that are the most difficult to place or cause the most debate.

To present the following overview, I have thus had to make some rather bold choices. First, I have ignored Phrygian, Messapic, Dacian, Thracian and other *Trümmersprachen*. Second, I have consistently counted Baltic and Slavic as two branches to enable direct comparison with topologies in which they do not descent from an exclusive common ancestor. Third, one tree does not place Baltic. In view of the uncontroversial acceptance of a Balto-Slavic clade, I have counted Baltic as a sister of Slavic. Fourth, and more drastic, I have deliberately not counted "incomplete" trees as evidence in favour of a slimmer grouping. Presenting the number of trees in which only one other branch descends from the common ancestor of Indo-Iranian and Greek at face value would make a closer relationship seem more popular that it actually is. For instance, it would be misleading to present trees depicting Armenian as the only other descendant of Indo-Greek if Albanian is left unplaced. For this reason, such incomplete trees are not taken at face value in the summations below, but they are marked accordingly and counted as though they depicted the missing branch as descending from "Indo-Greek" as well. This is only reasonable because this paper aims at illustrating the discrepancy between "Indo-Greek" in terminology and on trees and does *not* discuss the consensus about a possible "Balkanic" subgroup within "Indo-Greek" or Indo-European. Fifth, and in direct contrast with the approach just described, I have not counted Anatolian and Tocharian as descending from the latest common ancestor of Indo-Iranian and Greek when these languages were unplaced. I find this inconsistency justified by the growing consensus that Anatolian and Tocharian were branches split off first.

3.5. The Indo-European Arboretum

Below are the results of the survey. Figure 1 presents the references to the 41 more-or-less binary branching Indo-European language family trees organised in groups of composition of branches and ordered according to the number of branches descending from the latest common ancestor of Indo-Iranian and Greek (indicated with a + and the number in question). The different compositions of branches – and in the case of Albanian, Armenian, Baltic and Slavic also the internal topologies of the same compositions – are marked with the letters *a*–*u*. These letters correspond to the blocks in Figure 2.

Figure 1: The composition of branches (*a–u*), ordered according to the number of branches descending from the latest shared ancestor of Indo-Iranian and Greek ("Indo-Greek")

+0 others	—
+1: AL	(Bouckaert et al. 2012: 959)
+1: AR	(Kallio 2012: 226; Nakhleh, Ringe & Warnow 2005: E)
+[2]: [AL[1]], AR	(Gamkrelidze & Ivanov 1995: 363; Mateescu & Salomaa 1997: 6; Fitch 2007: 1; Watkins 2001: 57; West 2007: 20†; Schindler *apud* Matzinger 2012: 143)
+3: ar, ba-sl	(Nakhleh, Ringe & Warnow 2005: A*, D; Campbell 2020: 232–3)
+[3]: [AL[1], AR[2]], IT	(Schleicher 1853: 787‡)
+4: AL, AR, AN, TO	(Holm 2008: 635)
+4: an, ge, it-ce	(Kozincev 2018: 176)
+4: AL, AR, BA-SL	
= "Balkan" & "Indo-Slavic"	(Kortlandt 2016: 81)
= Greek & "Satəm"	(Olander 2019: 241; Kortlandt 2020: 1)
+4: AR, BA, SL, GE	(Nakhleh, Ringe & Warnow 2005: B, C)
+[4]: AL, [AR[2]], IT-CE	(Schleicher 1860: 81‡)
+5: ba-sl, ge, it-ce	(Cavalli-Sforza 2000: 164‡)
+6: AL, AR, BA, GE, IT, SL	(Starostin *apud* Blažek 2007: 84)
+6: al, ba-sl, ce, ge, it	(Rexová, Frynta & Zrzavý 2003: a†, b†)
+6: ar, ba-sl, it, ce, ge	(Ryder & Nicholls 2010: 89, 90; Rexová, Frynta & Zrzavý 2003: c†)
+7: al, ar, ba-sl, ge, it-ce	(Gippert 1994: 458[12]; Chang et al. 2015: 200[23]; Bouckaert et al. 2013: 1446; Kassian et al. 2021: 956)
+7: al, ba-sl, ge, it-ce, to	(Trager & Smith 1950: 64)
+7: [AL[1], AR[2]], BA-SL, GE, CE, IT	(Fick 1870: 1051‡[12])
+8: all except AN	(Hamp 1989, 1990, 2012 in Hamp 2013: 4–8; Gray & Atkinson 2003: 437; Chang et al. 2015: 200; Heggarty et al. 2023)
+9: AN, AL, AR, BA-SL, GE, IT, CE, TO	(Stifter 2006: 1)

(Centum-Satəm divide)

[1] Albanian unplaced, counted as descending from "Indo-Greek"

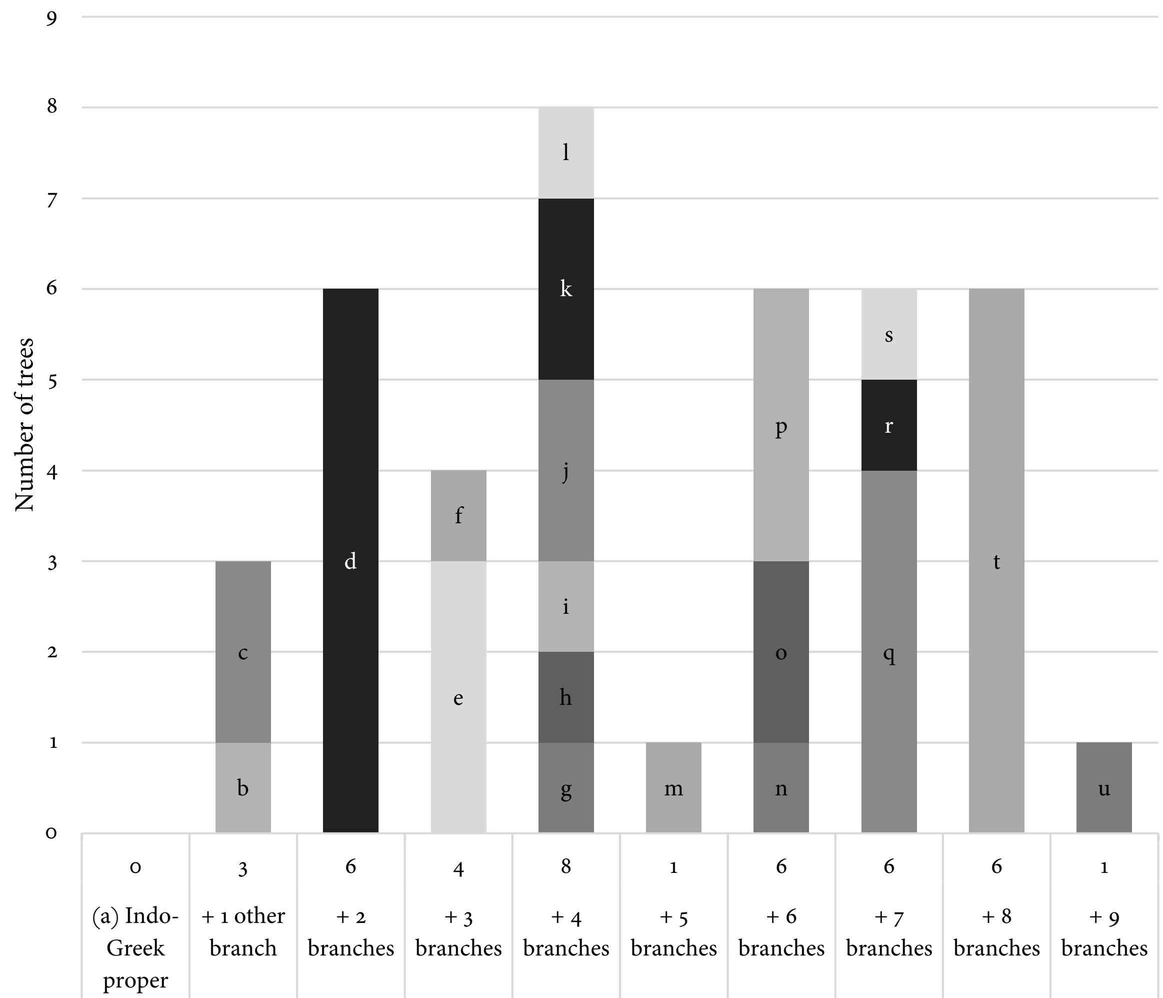
[2] Armenian unplaced, counted as descending from "Indo-Greek"

[3] Baltic unplaced, counted as descending from Balto-Slavic and consequently from "Indo-Greek"

† Tocharian or ‡ Anatolian and Tocharian unplaced, not counted as descendants of "Indo-Greek"

* Tree A is the same as in Ringe, Wanow & Taylor (2002: 87), and it is only included once

Figure 2: The “Indo-Greek” Arboretum depicting how many and which branches descend from the latest shared ancestor of Indo-Iranian and Greek in published Indo-European family trees.



Each column represents the *number of family trees* (summed up in the first row of the x-axis legend) depicting a certain *number of branches* as descending from the last shared ancestor of Indo-Iranian and Greek (“Indo-Greek”) found in the survey. This number of branches is labelled in the second row of the x-axis legend. Every shaded block marked with a letter corresponds to a composition (and in the cases of *i* and *j* also topology) of branches presented in Figure 1.

3.6. Picking the fruits of the phylogenetic trees

To my surprise, I found no trees grouping only Indo-Iranian and Greek together (Scenario *a* in Figure 1). There are, however, a few trees grouping Greek as the closest relative of Albanian (*b*) and Armenian (*c*), and this clade as the sister of Indo-Iranian. Strikingly, none of the six trees grouped under (*d*) overtly group Indo-Iranian, Greek, Armenian and Albanian together since they collectively leave Albanian unmentioned. This is of course due to the disputed placement of Albanian in the Indo-European family, but for the purpose of addressing the latest shared "Indo-Greek" ancestor, such a (graphic) depiction muddles how narrow this group might be.

The trees that include three other branches mainly represent the inclusion of Balto-Slavic and Armenian as descendants of "Indo-Greek" (*e*) – as well as an early tree by Schleicher (*f*), in which Italic is a close relative of Greek. This tree does not place Albanian and Armenian, the former probably the for the same reasons as the modern trees that leave it out, and the latter due to the fact that it predates Hübschmann's (1877a; 1877b) classification of Armenian as a separate branch.

Given the nature of the problem investigated, we should not expect any normal distribution. Nevertheless, the most popular number of other descendants is close to the median. There are eight publications that depict four other branches descending from the last common ancestor of Indo-Iranian and Greek, and they fall into six different topologies (*g*-*l*). Although more included languages yield more possible topologies, it is rather unexpected that the trees with four other descendants are so diverse. I decided to separate *i*, representing "Indo-Slavic" and "Balkanic" as sisters, from *j*, representing Greek as a sister of a "Satəm" clade because of their prominence, but obviously not because of their (independent) popularity among published trees.

It is possible to find claims of all other branches descending from "Indo-Greek" as well (*m*–*r*). If Greek or Indo-Iranian is seen as the second branch to split-off after Anatolian (*s*), "Indo-Greek" consequently comes to mean nothing but non-Anatolian Indo-European. Finally, if the Centum-Satəm isogloss is treated as the first split of the language family (*t*), "Indo-Greek" loses the last bit of meaning as the latest common ancestor turns out to be the same as Proto-Indo-European.

Since duplicates and trees with clear references to previously published works were not included in this survey, these numbers are not representative of the relative popularity in the scholarship. In fact, one true tree accepted and referenced by all scholars but the ones quoted above, would appear rather isolated in this presentation.

However, the *Arboretum* above does reveal the lack of consensus and the many competing hypotheses on the evolution of the Indo-European language family; and more importantly, the complete absence of an exclusive "Indo-Greek" node across all trees. There is a vague tendency for tree topologies based on grammatical feature to group the two branches somewhat closer than those based on cognacy or lexicon; but this is not clear-cut, and the works of prominent scholars go against it. The relationship between Greek, Armenian and/or Albanian on the one hand, and between Indo-Iranian and Balto-Slavic (and by extension the status of the Satəm languages as a subgroup) on the other hand are deeply connected to the question of an "Indo-Greek" unity.

4. Conclusion

In this paper, I have assessed two issues of comparison of Indo-Iranian and Greek: the vague use of the terminology surrounding the relationship between the two branches, and their position in family trees across the literature. Neither the application of the suggestive terminology nor the tree topologies reveal any particularly close relationship across the scholarship.

The terms *Graeco-Aryan* and *Indo-Greek* are often used interchangeably. Some scholars prefer reserving the former as a label for a model of the reconstructed Indo-European verbal system and the latter for a suggested phylogenetic or dialectal group, whereas others do not distinguish and employ only one term in both functions.

The terms are popular in prose to describe a rather undefined special – but to my knowledge never exclusive – relationship between the two branches; most often Armenian and/or Albanian and sometimes Balto-Slavic descend from their latest common ancestor as well.

I have not been able to find a single bifurcating phylogenetic tree which places Greek and Indo-Iranian as exclusive descendants of a common ancestor across a survey of 41 (not reprinted) family trees. The closest being sharing their latest ancestor with one other branch (Armenian or Albanian as a sister of Greek).

The origin of the terminological confusion instead lies in the spread of terms from a dialect geographical model of language change into works in linguistic phylogenetics.

Of course, neither terms nor trees define linguistic subgroups – only exclusive shared innovations do; a topic I will explore further elsewhere.

5. Bibliography

Adrados, Francisco Rodríguez. 1974. *Evolución y estructura del verbo indoeuropeo*. 2ª ed. (Consejo Superior de Investigaciones Cientificas. Manuales y Anejos de Emérita 21). Madrid: Instituto Antonio de Nebrija.

Adrados, Francisco Rodríguez. 1979. Arqueología y diferenciación del Indoeuropeo. *Emerita* 47(2). 261–282.

Adrados, Francisco Rodríguez. 1980. Les langues slaves dans le contexte des langues indoeuropéennes. *Sapostavitelno ezikoznanie* 5(4). 3–15.

Adrados, Francisco Rodríguez. 1981. Towards a new Stratigraphy of the Homeric Dialect. *Glotta* 59(1/2). 13–27.

Adrados, Francisco Rodríguez. 1991. Dictionaries of Indo-European and Their Problems. *Lexicographica* 7. 11–24.

Allen, N. J. 2007. Review of West (2007). *Bryn Mawr Classical Review*(10). 53–53.

Allen, N. J. 2019. *Arjuna-Odysseus: Shared heritage in Indian and Greek epic*. London & New York (NY): Routledge & Taylor & Francis Group.

Allen, Nick. 2015. Review of Wulff Alonso (2014). *Rivista degli studi orientali* 88(1/4). 241–247.

Andrianne, Gilles. 2013. Tendre l'arc: Comparaisons entre le *Rāmāyaṇa* et l'*Odyssée*. *Arhiv Orientální* 81(2). 307–319.

Anthony, David & Don Ringe. 2015. The Indo-European Homeland from Linguistic and Archaeological Perspectives. *Annual Review of Linguistics* 1. 199–219. doi:10.1146/annurev-linguist-030514-124812.

Ascoli, Gradziadio. 1893. Ein Wort über die Verwandtschaftsverhältnisse innerhalb der indogermanischen Familie. In E. Delmar Morgan (ed.), *Transactions of the Ninth International Congress of Orientalists (held in London, 5th to 12th September 1892)*, vol. 1. Indian and Aryan Sections, 554–556. London & Edinburgh: Ballantyne, Hanson & Co.

van Beek, Lucien Christiaan. 2022. Greek. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 173–201.

Cambridge: Cambridge University Press. doi:10.1017/9781108758666.011 (23 September, 2022).

Bendahman, Jadwiga. 1993. *Der reduplizierte Aorist in den indogermanischen Sprachen*. (Deutsche Hochschulschriften 642). Egelsbach: Hänsel-Hohenhausen.

Bičovský, Jan. 2017. *Praindoevropština. 2. Dodatky*. (Varia 62). Praha: Filozofická fakulta Univerzity Karlovy.

Birwé, Robert. 1956. *Griechisch-Arische Sprachbeziehungen im Verbalsystem*. (Beiträge zur Sprach- und Kulturgeschichte des Orients 8). Walldorf Hessen: Verlag für Orientkunde Dr. H. Vorndran.

Blažek, Václav. 2007. From August Schleicher to Sergei Starostin: On the development of the tree-diagram models of the Indo-European languages. *Journal of Indo-European Studies* 35(1 & 2). 82–109.

Bonfante, Giulano. 1976. *I dialetti indoeuropei*. (Studi Grammaticali e Linguistici 12). Brescia: Paidaia.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, et al. 2012. Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337(6097). 957–960. doi:10.1126/science.1219669.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, et al. 2013. Corrections and Clarifications. *Science* 342. 1446. doi:10.1126/science.1219669.

Campbell, Lyle. 2020. *Historical linguistics: An introduction*. 4th ed. Edinburgh: Edinburgh University Press.

Cavalli-Sforza, Luigi Luca. 2000. *Genes, peoples, and languages*. New York (NY): North Point 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. doi:10.1353/lan.2015.0005.

Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. (Publications of the Philological Society 30). Oxford & Cambridge (MA): Blackwell.

Clackson, James. 2007. *Indo-European linguistics: An introduction*. (Cambridge Textbooks in Linguistics). Cambridge & New York: Cambridge University Press.

De Decker, Filip. 2011. Evidence for laryngeal aspiration in Greek? Part I: The “recent” evidence. *Indogermanische Forschungen* 116. 87–109.

De Decker, Filip. 2015. What is the Greek counterpart of (Proto-)Indo-Iranian (*)th? *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 12. 89–164.

De Decker, Filip. 2016a. Etymological and methodological observations on the ‹pg› and ‹pg?› vocabulary in Robert Beekes's etymological dictionary of Greek: *N. Studia Linguistica Universitatis Iagellonicae Cracoviensis* 133(3). 133–149.

De Decker, Filip. 2016b. Etymological and methodological observations on the ‹pg› and ‹pg?› vocabulary in Robert Beekes's etymological dictionary of Greek: *M. Studia Linguistica Universitatis Iagellonicae Cracoviensis* 133(2). 75–96.

Drinka, Bridget. 2013. Phylogenetic and areal models of Indo-European relatedness: The role of contact in reconstruction. *Journal of Language Relationship • Вопросы языкового родства* 6(2). 379–410. doi:10.1163/19552629-00602009.

Durante, Marcello. 1976. *Sulla preistoria della tradizione poetica greca*. Vol. 2. Risultanze della comparazione indoeuropea. (Incunabula Graeca 64). Roma: Edizioni dell'Ateneo.

Elbourne, Paul. 2012. A rule of deaspiration in ancient Greek. In Philomen Probert & Andreas Willi (eds.), *Laws and Rules in Indo-European*, 125–133. Oxford University Press. https://academic.oup.com/book/12348/chapter/161913377 (23 March, 2023).

Euler, Wolfram. 1979. *Indoiranisch-griechische Gemeinsamkeiten der Nominalbildung und deren indogermanische Grundlagen*. (Innsbrucker Beiträge Zur Sprachwissenschaft B30). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Euler, Wolfram. 1980. Gemeinsamkeiten der Nominalbildung im Indoiranischen und Griechischen. In Manfred Mayrhofer, Oskar E. Pfeiffer & Martin Peters (eds.), *Lautgeschichte und Etymologie: Akten der IV. Fachtagung der Indogermanischen Gesellschaft Wien, 1978*, 173–179. Wiesbaden.

Felsenstein, Joseph. 1978. The Number of Evolutionary Trees. *Systematic Zoology* 27(1). 27. doi:10.2307/2412810.

Fick, August. 1870. *Vergleichendes Wörterbuch der indogermanischen Sprachen: Ein sprachgeschichtlicher Versuch*. 2. Umgearb. Aufl. Göttingen: Vandenhoeck & Ruprecht. http://archive.org/details/bub_gb_uhQJAAAAQAAJ (2 February, 2024).

Fitch, W. Tecumseh. 2007. An invisible hand. *Nature* 449(7163). 665–667. doi:10.1038/449665a.

Fortson, Benjamin Wynn, IV. 2010. *Indo-European language and culture: An introduction*. 2nd ed. (Blackwell Textbooks in Linguistics 19). Chichester & Malden (MA): Wiley-Blackwell.

Fries, Almut. 2021. The Mountain Feet of the Ship: An Indo-European poetic image? *Quaderni urbinati di cultura classica*(3). doi:10.19272/202106403005.

Gamkrelidze, Tʻamaz & Vjačeslav V. Ivanov. 1995. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture*. (Trends in Linguistics. Studies and Monographs 80). Berlin & New York (NY): Mouton de Gruyter.

Gąsiorowski, Piotr. 1999. The tree of language: A cladistic look at the genetic classification of languages. *Dialectologia et Geolinguistica* 7. 1–20. doi:10.1515/dig.1999.1999.7.39.

Ginevra, Riccardo. 2022. Review of *Kentauren und Gandharven* by Michael Janda (Münster: Thomas Kubo, 2022) & *Apollon und Dionysos* by Michael Janda (Münster: Thomas Kubo, 2022). *Kratylos* 67(1). 12–31. doi:10.29091/KRATYLOS/2022/1/4.

Gippert. 1994. Who were the laryngeals. In Jens Elmegård Rasmussen & Benedicte Nielsen Whitehead (eds.), *In honorem Holger Pedersen: Kolloquium der Indogermanischen Gesellschaft vom 26. bis 28. März 1993 in Kopenhagen*, 455–466. Wiesbaden: Reichert.

Goldstein, David M. 2022. There's no escaping phylogenetics. In Laura Grestenberger, Charles Reiss, Hannes A. Fellner & Gabriel Z. Pantillon (eds.), *Ha! Linguistic studies in honor of Mark R. Hale*, 71–91. Wiesbaden: Reichert.

Grassmann, H. 1863a. Ueber die aspiraten und ihr gleichzeitiges vorhandensein im an- und auslaute der wurzeln. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 12(2). 81–110.

Grassmann, H. 1863b. Ueber das ursprüngliche Vorhandensein von wurzeln, deren anlaut und auslaut eine aspirate enthielt. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 12(2). 110–138.

Gray, Louis H. 1925. The Indo-European Negative Prefix in *N. Language* 1(4). 119–129. doi:10.2307/409539.

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–9.

Greenhill, Simon J. & Russell D. Gray. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica* 29(4). 523–537. doi:10.1075/dia.29.4.05gre.

Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2021. Bayesian Phylolinguistics. In Richard D. Janda, Joseph, Brian D. & Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, 226–253. John Wiley & Sons. doi:10.1002/9781118732168.ch11 (28 September, 2021).

Grestenberger, Laura. 2016. Reconstructing Proto-Indo-European Deponents. *Indo-European Linguistics* 4(1). 98–149. doi:10.1163/22125892-00401001.

Grigoriev, Stanislav A. (ed.). 2002. *Ancient Indo-Europeans*. (Eurasian Ancient History 1). Chelyabinsk: RIFEI.

Hajnal, Ivo. 2011. *Studien zum mykenischen Kasussystem*. De Gruyter. doi:10.1515/9783110888324.

Hamp, Eric P. 2013. The expansion of the Indo-European languages: An Indo-Europeanist's evolving view (with annotation and comments by Douglas Q. Adams). *Sino-Platonic Papers* 239. (1–14).

Hartmann, Hans. 1979. Synckronische und diachronische Studien zur Syntax des Irischen. 37(1). 10–157. doi:10.1515/zcph.1979.37.1.10.

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, et al. 2023. Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* 381(6656). eabg0818. doi:10.1126/science.abg0818.

Holm, Hans J. 2008. The distribution of data in word lists and its impact on the subgrouping of languages. In Christine Preisach, Hans Burkhardt, Lars Schmidt-Thieme & Reinhold Decker (eds.), *Data analysis, machine learning and applications: Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7–9, 2007*, 629–636. Berlin & Heidelberg: Springer.

Hübschmann, H. 1877a. Ueber die stellung des armenischen im kreise der indogermanischen sprachen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 23(1). 5–49.

Hübschmann, H. 1877b. Armeniaca. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 23(4). 400–407.

Jackson, Peter. 2006. *The Transformations of Helen: Indo-European Myth and the Roots of the Trojan Cycle*. J.H. Röll.

Janda, Michael. 2012. Review of West (2007). *Gnomon* 84(6). 481–487.

Kallio, Petri. 2012. The prehistoric Germanic loanword strata in Finnic. *A Linguistic Map of Prehistoric Northern Europe*, 225–238. (Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de La Société Finno-Ougrienne 266).

Kapović, Mate. 2016. Proto-Indo-European phonology. In Mate Kapović, Anna Giacalone Ramat & Paolo Ramat (eds.), *The Indo-European Languages*, 13–60. 2nd ed. London: Routledge. doi:10.4324/9781315678559.

Kassian, Alexei S., Mikhail Zhivlov, George Starostin, et al. 2021. Rapid radiation of the inner Indo-European languages: an advanced approach to Indo-European lexicostatistics. *Linguistics* 59(4). 949–979. doi:10.1515/ling-2020-0060.

Kern, Heinrich. 1858. 1) Ueber die Italer; 2) Foemina, foetus. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 7(4). 272–275.

Kim, Ronald I. 2018. Greco-Armenian: The persistence of a myth. *Indogermanische Forschungen* 123. 247–272. doi:10.1515/if-2018-0009.

Kiparsky, Paul. 1973. On comparative linguistics: The case of Grassmann's law. In Thomas Sebeok (ed.), *Current trends in linguistics*, 115–134. The Hague: Mouton.

Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. (Leiden Indo-European Etymological Dictionary Series 5). Leiden & Boston (MA): Brill.

Kloekhorst, Alwin. 2013. Indo-European nominal ablaut patterns: The Anatolian evidence. In Götz Keydana, Paul Widmer & Thomas Olander (eds.), *Indo-European accent and ablaut*, 107–128. (Copenhagen Studies in Indo-European 5). Copenhagen: Museum Tusculanum Press.

Koch, John T. 1991. On the prehistory of Brittonic syntax. *Studies in Brythonic word order*, 1–43. Amsterdam: John Benjamins. (21 March, 2023).

Kortlandt, Frederik Herman Henri. 2016. Baltic, Slavic, Germanic. *Baltistica* 51(1). 81–86.

Kortlandt, Frederik Herman Henri. 2019. Indo-European and its Neighbors. *Journal of Indo-European Studies* 47(1/2). 143–147.

Kortlandt, Frederik Herman Henri. 2020. The disintegration of the Indo-European language family. In Harald Bichlmeier, Ondřej Šefčik & Roman Sukač (eds.), *Etymologus: Festschrift for Václav Blažek. Čas pro sladkou tečku*, 267–276. (Studien Zur Historisch-Vergleichenden Sprachwissenschaft 14). Hamburg: Baar.

Kozincev, Alexander. 2018. On certain aspects of distance-based models of language relationships, with reference to the position of Indo-European among other language families. *Journal of Indo-European Studies* 46. 173–205.

Kretschmer, Paul. 1896. *Einleitung in die Geschichte der Grieehisehen Sprache*. Göttingen: Vandenhoeck & Ruprecht.

Kroonen, Guus, Anthony Jakob, Axel I. Palmér, et al. 2022. Indo-European cereal terminology suggests a Northwest Pontic homeland for the core Indo-

European languages. *PLOS ONE* 17(10). e0275744. doi:10.1371/journal.pone.0275744.

Kuhn, Adalbert. 1853. Ueber die durch nasale erweiterten verbalstämme. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 2(6). 455–471.

Leroy, Maurice & Francis Mawet (eds.). 1986. *La place de l'Arménien dans les langues indo-européenes*. (Fonds René Draguet 3). Louvain: Peeters.

Lühr, Rosemarie. 2012. The structure of nominal paradigms in Indo-European languages. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphophonemics*, 182–196. (LINCOM Studies in Indo-European Linguistics 41). München: Lincom Europa.

Mallory, James Patrick. 2013. Twenty-first century clouds over Indo-European homelands. *Journal of Language Relationship • Вопросы языкового родства* 9(2013). 145–154.

Mallory, James Patrick & Douglas Quentin Adams (eds.). 1997. *Encyclopedia of Indo-European culture*. London & Chicago (IL): Fitzroy Dearborn.

Mallory, James Patrick & Douglas Quentin Adams. 2006. *The Oxford introduction to Proto Indo European and the Proto Indo European world*. Oxford & New York (NY): Oxford University Press.

Malzahn, Melanie. 2012. Archaism and Innovation in the Tocharian Verbal System: The Case of Valency and the Case for a Conspiracy Theory. In Harold Craig Melchert (ed.), *The Indo-European verb: Proceedings of the Conference of the Society for Indo-European Studies, Los Angeles 13–15 September 2010*, 235–240. Wiesbaden: Reichert.

Marcantonio, Angela. 2009. Introduction. In Angela Marcantonio (ed.), *The Indo-European language family: Questions about its status*, 1–71. (Journal of Indo-European Studies Monograph Series 55). Washington (DC): Institute for the Study of Man.

Martirosyan, Hrach. 2013. The place of Armenian in the Indo-European language family: the relationship with Greek and Indo-Iranian. *Journal of Language Relationship • Вопросы языкового родства* 10(1). 85–138. doi:10.31826/jlr-2013-100107.

Mateescu, Alexandru & Arto Salomaa. 1997. Formal languages: An introduction and a synopsis. In Grzegorz Rozenberg & Arto Salomaa (eds.), *Handbook of Formal Languages*, vol. 1. Word, Language, Grammar. Berlin & Heidelberg: Springer.

Matras, Yaron. 2015. Transnational policy and ‘authenticity’ discourses on Romani language and identity. *Language in society* 44(3). 295–316. doi:10.1017/S0047404515000202.

Matzinger, Joachim. 2012. “Zwischensprachen” – areallinguistische Bemerkungen aus dem Bereich des Balkanindogermanischen. In Velizar Sadovski & David Stifter (eds.), *Iranistische und indogermanistische Beiträge in memoriam Jochem Schindler (1944–1994)*, 137–160. (Veröffentlichungen Des Instituts Für Iranistik 51). Wien: Verlag der Österreichischen Akademie der Wissenschaften.

McCone, Kim. 1991. *The Indo-European Origins of the Old Irish Nasal Presents, Subjunctives and Futures*. (Innsbrucker Beiträge Zur Sprachwissenschaft 66). Institut für Sprachwissenschaft der Universität Innsbruck.

Meid, Wolfgang. 1975. Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen. In Helmut Rix (ed.), *Flexion und Wortbildung: Akten der V. Fachtagung der Indogermanischen Gesellschaft. Regensburg, 9.-14- September 1973*, 204–219. Wiesbaden: Reichert.

Meusel, Eduard. 2019. Pindarus Indogermanicus: Untersuchungen zum Erbe dichtersprachlicher Phraseologie bei Pindar. *Pindarus Indogermanicus*. De Gruyter. doi:10.1515/9783110666496 (28 April, 2023).

Miller, D. Gary. 1977. Bartholomae’s Law and an IE root structure constraint. *Studies in descriptive and historical linguistics: Festschrift for Winfred P. Lehmann*, 365–392. Amsterdam: Hopper.

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.

Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. (Brill’s Studies in Indo-European Languages and Linguistics 14). Leiden & Boston (MA): Brill.

Olander, Thomas. 2018. Connecting the dots: The Indo-European family tree as a heuristic device. In David M. Goldstein, Stephanie W. Jamison & Brent Vine (eds.), *Proceedings of the 29th annual UCLA Indo-European conference*, 181–202. (Journal of Indo-European Studies Monograph Series). Bremen: Hempen.

Olander, Thomas. 2019. Indo-European cladistic nomenclature. *Indogermanische Forschungen* 124. 231–244.

Olander, Thomas. 2022. To **b* or not to **b*: Proto-Indo-European **b* in a phylogenetic perspective. *Historische Sprachforschung* 133(1). 182–208.

Olsen, Birgit Anette. 1999. *The noun in Biblical Armenian: Origin and word formation: With special emphasis on the Indo-European heritage*. (Trends in Linguistics. Studies and Monographs 119). Berlin & New York (NY): Mouton de Gruyter.

Olsen, Birgit Anette & Rasmus Thorsø. 2022. Armenian. In Thomas Olander (ed.), *The Indo-European language family. A phylogenetic perspective*.

Parpola, Asko. 2005. The Nāsatyas, the Chariot and Proto–Aryan Religion. *Journal of Indological Studies* 16–17. 1–63.

Phillips-Rodriguez, Wendy. 2016. Review of Wulff Alonso (2014). *Saṃbhāṣā: Nagoya Studies in Indian Culture and Buddhism* 33(4). 73–74.

Pisani, Vittore. 1933. *Studi sulla preistoria delle lingue indeuropee: Memoria*. Vol. 6. (Reale Accademia Nazionale Dei Lincei. Serie 6 4). Roma: Bardi.

Pisani, Vittore. 1940. Geolinguistica e indeuropeo. *Memorie della R. Accademia Nazionale dei Lincei. Classe di Scienze Morali, Storiche e Filologiche. Anno 336. Serie 6* 9(2). 113–379.

Pisani, Vittore. 1974. *Indogermanisch und Europa*. (Ars grammatica 2). München: Wilhelm Fink.

Pisano, Carmine. 2015. India and Greece: Methods and Models of Comparison in "the Mahābhārata and Greek Mythology" by Fernando Wulff Alonso. *Rivista degli Studi Orientali* 88(1/4). 215–227.

Polomé, Edgar C. 1989. A few notes on Indo-Aryan-Hellenic isoglosses. In Karin Heller, Oswald Panagl & Johann Tischler (eds.), *Indogermanica Europaea: Festschrift für Wolfgang Meid zum 60. Geburtstag am 12. 11. 1989*. (Grazer Linguistische Monographien 4). Graz: Institut für Sprachwissenschaft der Universität Graz.

Pooth, Roland A. 2009. Proto-Indo-European ablaut and root inflection: An internal reconstruction and inner-PIE morphological analysis. In Jens Elmegård Rasmussen & Thomas Olander (eds.), *Internal reconstruction in Indo-European: Methods, results, and problem: section papers from the XVI International Conference on Historical Linguistics University of Copenhagen, 11th-15th August, 2003*, 229–254. (Copenhagen Studies in Indo-European 3). Copenhagen: Museum Tusculanum.

Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. (Indogermanische Bibliothek. Reihe 3. Untersuchungen). Heidelberg: Carl Winter.

Rexová, Kateřina, Daniel Frynta & Jan Zrzavý. 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19. 120–127. doi:10.1111/j.1096-0031.2003.tb00299.x.

Ringe, Don. 2010. 'Thorn' clusters and Indo-European subgrouping. In Ron Kim (ed.), *Ex Anatolia Lux: Anatolian and Indo-European studies in honor of H. Craig Melchert on the occasion on his sixty-fifth birthday*, 330–38. Ann Arbor (MI): Beech Stave Press.

Ringe, Donald A. 2017. *From Proto-Indo-European to Proto-Germanic*. 2nd ed. (A Linguistic History of English 1). Oxford & New York (NY): Oxford University Press.

Ringe, Donald A., Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1). 59–129. doi:10.1111/1467-968X.00091.

Rix, Helmut. 1977. Das keltische Verbalsystem auf dem Hintergrund des indo-iranisch-griechischen Rekonstruktionsmodells. In Karl Horst Schmidt (ed.), *Indogermanisch und Keltisch: Kolloquium der Indogermanischen Gesellschaft am 16. und 17. Februar 1976 in Bonn*, 132–158. Wiesbaden: Reichert.

Rix, Helmut. 2009. *Historische Grammatik des Griechischen: Laut- und Formenlehre*. 2. kor. Aufl. (WBG-Bibliothek). Darmstadt: Wissenschaftliche Buchgesellschaft.

Ryder, Robin & Geoff K. Nicholls. 2010. Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. *Applied Statistics* 60. 71–92.

Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3. 786–787.

Schleicher, August. 1860. *Die Deutsche Sprache*. Stuttgart: J. C. Cotta'scher Verlag.

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

Schmitt, Rüdiger. 1967. *Dichtung und Dichtersprache in indogermanischer Zeit*. Wiesbaden: Harrassowitz.

Sheldon, J. S. 2009. Review of West (2007). The Wider Context: Common Elements in Indo-Iranian, Greek and other Poetic Traditions and Mythologies. *Journal of the Royal Asiatic Society* 19(4). 509–528. doi:10.1017/S1356186309990083.

Sihler, Andrew Littleton. 1995. *New Comparative Grammar of Greek and Latin*. Oxford & New York (NY): Oxford University Press.

Sommer, Florian. 2016. Diachronie und areale Effekte: Zur Entstehung der bestimmten Adjektive im Baltischen und Slavischen. *Die Sprache* 52(2). 202–255.

Sonne, Wilhelm Günther Heinrich. 1863. Sprachliche und mythologische Untersuchungen, angeknüpft an Rigveda I, 50. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 12(4). 267–298.

Stifter, David. 2006. *Sengoídelc: Old Irish for beginners*. (Irish Studies). Syracuse (NY): Syracuse University Press.

Szemerényi, Oswald John Louis. 1964. Review of Devoto 1962. *The Journal of Hellenic Studies* 84. 175–177. doi:10.2307/627726.

Tovar, Antonio. 1977. *Krahes alteuropäische Hydronymie und die westindogermanischen Sprachen: Vorgetragen am 7. Februar 1976*. Vol. 2. (Sitzungsberichte Der Heidelberger Akademie Der Wissenschaften, Philosophisch-Historische Klasse 1977). Heidelberg: Winter.

Tovar, Antonio. 1979. Notas lingüísticas sobre monedas ibéricas. In Antonio Tovar (ed.), *Actas del II Coloquio sobre Lenguas y Culturas Prerromanas de la Península Ibérica (Tübingen, 17-19 junio 1976)*. (Acta Salmanticensia 113). Salamanca: Ediciones Universidad de Salamanca.

Trager, George L. & Henry Lee Smith. Jr. Smith. 1950. A chronology of Indo-Hittite. *Studies in Linguistics* 8(3). 61–70.

Udolph, Jürgen. 1995. Review of Bammesberger (1986). *Indogermanische Forschungen* 100. 298–301. doi:10.1515/9783110243413.268.

Vassilkov, Yaroslav. 2011. Indian "hero-stones" and the earliest anthropomorphic stelae of the bronze age. *Journal of Indo-European Studies* 39(1–2). 194–229.

Vergados, Athanassios. 2013. *The "Homeric hymn to Hermes": Introduction, text and commentary*. (Texte Und Kommentare: Eine Altertumswissenschaftliche Reihe 41). Berlin & Boston: De Gruyter.

Villar, Francisco. 1996. *Los indoeuropeos y los orígenes de Europa: Lenguaje e historia*. 2a ed. corrigida y muy aumentada. Madrid: Gredos.

Watkins, Calvert. 1995. *How to kill a dragon: Aspects of Indo-European poetics*. New York (NY): Oxford University Press.

Watkins, Calvert. 2001. An Indo-European linguistic area and its characteristics: Ancient Anatolia. Areal diffusion as a challenge to the comparative method. In Alexandra Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 44–63. Oxford: Oxford University Press.

West, Emily. 2014. Circe, Calypso, Hiḍimbā: The Odyssey and Graeco-Aryan Proto-Epic. *Journal of Indo-European Studies* 42(1–2). 144–174.

West, Martin Litchfield. 1988. The Rise of the Greek Epic. *The Journal of Hellenic Studies* 108. 151–172. doi:10.2307/632637.

West, Martin Litchfield. 1994. Review of *Greek Mythology and Poetics* by Nagy, Gregory (Ithaca: Cornell University Press, 1990). *History of European Ideas* 18(5). 807–808.
West, Martin Litchfield. 2004. Man, fate, and death in Indo-European tradition. *Letras Clássicas* 8. 49–66. doi:10.11606/issn.2358-3150.v0i8p49-66.
West, Martin Litchfield. 2007. *Indo-European Poetry and Myth*. Oxford: Oxford University Press.
Willi, Andreas. 2018. *Origins of the Greek Verb*. Cambridge: Cambridge University Press. doi:10.1017/9781108164207 (21 March, 2023).
Wulff Alonso, Fernando. 2014. *The Mahābhārata and Greek Mythology*. (Trans.) Andrew Morrow. Delhi: Motilal Banarsidass.
Yakubovich, Ilya Sergeyevich. 2011. Review of Marcantonio (ed.) (2009). *Journal of Language Relationship • Вопросы языкового родства* 6. 227–233.
Zahn, Ingo. 2014. *Vergleichende indogermanische Formenlehre*. (Schriftenreihe Schriften zur vergleichenden Sprachwissenschaft 10). Hamburg: Kovač.
Zair, Nicholas. 2018. On the relative sonority of PIE /m/. *Indo-European Linguistics* 6(1). 271–303. doi:10.1163/22125892-00601005.

Article 2: Loanwords and Linguistic Phylogenetics: **pelek̂u-* 'axe' and **(H)a(i̯)ĝ-* 'goat'

The following paper appeared as Poulsen (2025a) in *Transactions of the Philological Society*.

Immediately after the initial online publication, I received valuable comments in correspondences. These are summarised in the *Addenda et corrigenda*, which are now published on the homepage of the journal (Poulsen 2025b).

Immediately after *that* publication, I was made aware (J. Torpy, p.c., March 26, 2025) that I have misrepresented the date of the domestication of goats (2025a: 131). While wild goats were widespread after the last ice age, domesticated goats were not widespread before the neolithic. For the linguistic angle, the point still stands: "it is difficult to attribute its lexical diversity to a "late introduction" to speakers of IE languages" (Mallory & Adams 1997: 230). Goat (and cattle) dairying does not occur before 2800 BCE (Scott et al. 2022: 815–7).

Transactions of the Philological Society

PhilSoc The Philological Society

Transactions of the Philological Society Volume 00 (2025) 1–21 doi: 10.1111/1467-968X.12308

LOANWORDS AND LINGUISTIC PHYLOGENETICS: **pelek̑u-* 'AXE' AND **(H)a(i̯)ĝ-* 'GOAT'[1]

By Simon Poulsen
University of Copenhagen

(Submitted: 8 March, 2024; Accepted: 9 November, 2024)

Abstract

This paper assesses the role of borrowings in two different approaches to linguistic phylogenetics: Traditional qualitative analyses of lexemes, and quantitative computational analysis of cognacy. It problematises the assumption that loanwords can be excluded altogether from datasets of lexical cognacy. It discusses two exemplary lexemes with a limited regional or topological distribution, which have been argued to be borrowings into intermediate proto-languages between the dissolution of Proto-Indo-European and the protolanguages reconstructible for the daughter branches: **pelek̑u-* 'axe' (limited to Indo-Iranian and Greek; allegedly from a Semitic language) and **(H)a(i̯)ĝ-* 'goat' (limited to "Balkanic", "Indo-Slavic" or both; allegedly North-East Caucasian). The paper brings to light how diverging analyses of these lexemes have been and may be used as phylogenetic arguments for different subgroupings. It further discusses the problems with the loan word origins of the lexemes: Although there is a Semitic root **p-l-q* 'cut', it is impossible to derive the noun **pelek̑u-* from it in Semitic and in Indo-European; and while there is a reconstructed Proto-North-East-Caucasian form comparable to **(H)a(i̯)ĝ-*, the connection is dependent on idiosyncrasies. The main point of the paper is that loanword judgements and linguistic reconstruction are interdependent. This does not discredit the discipline, but it does call for awareness of the assumptions underlying the linguistic analyses on which the phylogenetic results rest, no matter the approach.

Zusammenfassung

Dieser Aufsatz untersucht die Rolle von Lehnwörtern unter zwei verschiedenen Zugängen zur Untersuchung sprachlich-genetischer Verwandtschaftsverhältnisse, nämlich traditionelle qualitative Analysen von Lexemen und quantitative Analysen von Kognaten. Er problematisiert die Annahme, dass Lehnwörter ganz aus Kognatdateien ausgeschlossen werden können. Die Diskussion wird anhand von zwei Beispielen von Lexemen mit einer geographisch oder topologisch eingeschränkten Verbreitung geführt, die als Lehnwörter in Ursprachen zwischen der Auflösung des Urindogermanischen und den rekonstruierbaren Ursprachen der Tochtersprachen vorgeschlagen worden sind: **pelek̑u-* ‚Axt, Beil' (beschränkt auf das „Indo-Griechische"; angeblich aus einer semitischen Sprache) und

[1]I wish to thank Thomas Olander and Matthew Scarborough for their extensive comments on previous drafts, and two anonymous reviewers for valuable comments, critique and suggestions. I also wish to express gratitude to Benjamin Suchard, Peter Schrijver, Adam Hyllested, Joachim Matzinger, Rasmus Bjørn and Don Ringe for help, answers and fruitful discussions. Any remaining misunderstandings and errors are, of course, solely my own.

*(H)a(i̯)ĝ- ‚Ziege‘ (beschränkt auf die „Balkan-Gruppe“, das „Indo-Slavische“ oder auf beide Gruppen; angeblich aus dem Nordostkaukasichen). Der Aufsatz hebt hervor, wie abweichende Analysen dieser Lexeme als phylogenetische Argumente verschiedener Untergruppierungen verwendet worden sind und verwendet werden können. Es wird im Weiteren auf die konkreten Probleme eingegangen, die der postulierte Lehnwortursprung dieser Lexeme aufwirft: Obwohl es eine semitische Wurzel **p-l-q* ‚schneiden‘ gibt, kann man von ihr das Nomen **peleḱu-* weder im Semitischen noch im Indogermanischen ableiten; und obwohl es eine rekonstruierte Form im Urnordostkaukasischen gibt, die mit *(H)a(i̯)ĝ- verglichen werden kann, hängt dies von spezifischen Eigenheiten ab, die die Verbindung fragwürdig erscheinen lassen. Der Hauptpunkt des Aufsatzes ist, dass Beurteilung von Lehnwörtern und sprachwissenschaftliche Rekonstruktion voneinander abhängig sind. Obwohl dieses Abhängigkeitsverhältnis an sich die Methodik der vergleichenden Sprachwissenschaft nicht grundsätzlich in Frage stellt, verdeutlicht es doch die Notwendigkeit von Klarheit in Bezug auf die Analysen, die phylogenetischen Ergebnissen zugrunde liegen, unabhängig von der spezifischen angewandten Forschungsmethode.

1. INTRODUCTION

While loanwords may make up a substantial subsection of the lexicon, the process of borrowing itself is seen as rather trivial in historical linguistics. Loanwords are often judged insignificant for establishing linguistic phylogenies given that borrowing is a horizontal process (Hoenigswald 1966: 8; Clackson 1994: 7, 23; Warnow et al. 1996: 317; Ringe et al. 2002: 61–3; Gray & Atkinson 2003: 436; Nakhleh et al. 2005: 386–8; Chang et al. 2015: 205; Kassian et al. 2021: 957; Heggarty 2021: 390; Heggarty et al. 2023a: 10; Heggarty et al. 2023b: 37–9). Nevertheless, loanword judgement is a crucial part of determining linguistic ancestry and topology, and thus borrowings play an important role in establishing linguistic phylogenies, albeit less prominently than other types of data.

This article will problematise the assumption that lexical borrowings can be excluded completely from lexical (or, rather, cognacy) datasets compiled for computational studies, and it will discuss some attempts to make use of shared borrowings in phylogenetic arguments.

Once a lexeme has been borrowed and made its way into the lexicon of a language, it is passed on just like the inherited bulk (Nakhleh et al. 2005: 395; Heggarty 2021: 390). If we see protolanguages as real linguistic entities and not purely abstract repositories of projections, the co-occurrence of multiple cognates in the same meaning across related languages logically requires some lexical replacement or semantic innovation after the break-up of the protolanguage. Some scholars have investigated exactly this aspect of uniquely shared vocabulary. For instance, Porzig (1954) examined the regional distribution across the Indo-European continuum, and Thorsø (2020) has substantiated the hypothesis of a Balkanic branch through uniquely shared but ultimately foreign vocabulary.

Since this article is as much about exploring the advantages and limitations of different approaches to phylogenetic data as it is about the individual lexemes and their role as arguments for subgroupings, a robust foundation of the different premises of the methods is necessary. In the first section of the article, I will address the phylogenetic approach and model of language change. In connection to this, I will give a discussion of the data selection for phylogenetics and to different analytical approaches, before turning to two cases of possible early borrowings of regionally distributed etyma. The lexemes **peleḱu-* ‘axe’ and *(*H*)*a*(*i̯*)*ĝ*- ‘goat’ are traditionally posited as going back to at least the latest shared stage of Greek and Indo-Iranian, but the internal Indo-European developments and the possibility of ultimately foreign origins are still up for debate.

2. PRELIMINARIES

2.1. Phylogenetics and the tree model

This article subscribes to the tree model as an adequate model for depicting *genealogical* linguistic descent, because it allows us to generate falsifiable hypotheses (Ringe et al. 2002: 111; Olander 2018: 188; Pellard et al. forthcoming: 4–8). The great strength of the model is that it cannot encompass everything: individual arguments, isoglosses, or data points can be proven *not* to be innovated at a common stage, by for example differing relative chronologies, intermediate innovations, or alternative analyses. This means that hypotheses based on the tree model are generally falsifiable, in contrast to certain other models – especially the wave model with which it is difficult to be incompatible, and a network which can technically encompass everything.

2.2. A phylogenetic approach to reconstruction

Recent research has made it increasingly clear that we need to rethink reconstruction based on unresolved trees or unstratified data. Accepting the reality of a binary-branching tree has consequences for all aspects of linguistic reconstruction (Olander 2018, 2022; Jacques & List 2019; Goldstein 2022). In order not to project features all the way back into the earliest protolanguage, it is necessary to evaluate carefully whether each feature truly belongs to the oldest stage, or if it could be an innovation at a later node in the tree. For each linguistic trait, it is necessary to take an explicit stance on how far back it can be reconstructed. This means that the nodes within the family tree should be treated as linguistically real intermediate protolanguages or tightly-knit speech communities capable of undergoing the same innovations to such an extent that we probably would not be able to recognise the distinction.

2.3. The quality–quantity trade-off of data selection

As a gross oversimplification, different types of linguistic data fall on a scale of their value as arguments when drawing phylogenetic conclusions, bearing in mind of course that many types of data can be important if substantiated and weighted accordingly. Unfortunately, the importance of the data types is somewhat inversely proportionate to their availabilities.

Syntax is very prone to synchronic variation and contact-induced change (Hock 1988: 561). Despite it being the most intuitive example of linguistic change, sound change is surprisingly uninformative; individual phonetic changes are very often universal and trivial, and true irreversible phonemic mergers are rare (Ringe et al. 2002: 66–8; Hoenigswald 1966: 12). Strictly conditioned sound changes and sound shift complexes are, of course, much more telling, but in turn much rarer (Ringe & Eska 2013: 262).

Most significance is traditionally assigned to morphology, especially inflectional morphology because it is the least prone to borrowing (Nakhleh et al. 2005: 387). The number of categories and their formants are in principle unlimited, and the risk of chance resemblance is therefore low. However, the smaller amount of comparable items and their internal complexity leave room for subjectivity or idiosyncrasy in the analyses (Clackson 2022: 23).

Thus, the need of relatively large amounts of homogenous data to run reliable statistical analyses has made the lexicon gain popularity in computationally based phylogenetic studies.

3. LEXICAL DATA IN PHYLOGENY

3.1. Traditional approaches to lexical data

The main criterion for establishing subgroups in traditional methodology is shared non-trivial innovations. Cognates are not significant in themselves, but their semantic shifts and derivations may be.

Traditionalist arguments for subgroupings based on lexical data will therefore need *directionality* to be significant (Clackson 1994: 23). Mere distribution of cognates does not reveal which state is older or ancestral, and thus we cannot determine the innovation (Ringe et al. 2002: 96, 105; Adrados 1991: 16). To combat this issue, word formation might be taken into account. However, in related languages there is a certain risk of the innovations being parallel if they are coined with productive suffixes already present in the parent language (Clackson 1994: 24). Such arguments stand the strongest not just when form, function, meaning and relative chronologies align, but when they form conglomerates that are unlikely to occur anywhere else.

3.2. Directionality

Lexical innovations may be purely semantic, but these arguments are often fragile, since most developments are trivial and could occur independently (Ringe 2022: 54). Occasionally, backformation can be ruled out for semantic innovations; that is, they can appear irreversible, just like phonemic mergers. One such famous example is Schmidt's (1992): 113) analysis of the semantic development of the root *i̯eb^h-. This verb preserves the older meaning in Toch.B. *yäp-* 'to enter', whereas the meaning is purely sexual 'to penetrate' in Sanskrit, Greek, and Slavic (Winter 1997: 185). It is impossible to imagine this semantic development in reverse. It would be attractive to argue that this semantic shift happened after the break-off of a clade. But, as an anonymous reviewer reminds me, the root could very well have been polysemous in the parent language. In addition, we cannot rule out that the semantic shift (or, perhaps, semantic narrowing) happened multiple times independently in the non-Tocharian branches. This point has also been elaborated by Malzahn (2016: 283) and Friis (2024: 24–5, esp. n. 31).

3.3. Archaisms and cherry picking

It is always possible to find such unique shared features in languages that descend from the same ancestor, and while these cognates are important for reconstructing the lexicon of the intermediate protolanguages, they are not significant for establishing them. For a family like Indo-European where there are vast differences in the age and amount of attested material between the different branches, uncritically relying on the distribution of lexical material would lead to wrong groupings of the older or more copiously documented branches based on randomly shared archaisms.

Looking for positive evidence in favour of a hypothesised subgroup might happen 'at the expense of correspondences with other languages, in the belief that there was a special relationship between the two languages' (Clackson 1994: 191–2). Tackling this potential confirmation bias – or avoiding 'cherry picking' (e.g. Greenhill et al. 2021: 236–7) – is one of the goals of computational phylogenetic approaches. While most computational studies rely on lexical data, it would be useful to distinguish this more curated type, cognacy data, from the wider range of possible types of lexical data of the traditionalist approaches.

To compile large and relatively homogenous datasets for reliable and comparable analyses, computational phylogeneticists code *cognacy* of *basic vocabulary*. The entire lexicon is in principle scalable *ad infinitum*, and limiting the input to vocabulary which is supposedly universal in all languages will avoid the issues raised above. Further, the basic vocabulary is recognised to be the least prone to be borrowed (within one of the most borrowable linguistic categories); however, the definition or limitation of such a category is by no means straightforward (Tadmor et al. 2010: 227–8; Scarborough 2020; Kassian et al. 2010).

3.4. *Cognacy data sets*

In datasets of cognacy, the data are organised according to semantic slot and not the etymon. This means that each character is the answer to the question of what etymon is the standard or most 'unmarked' word in a given meaning, and not whether an etymon is preserved (possibly in a different meaning). These datasets cover lexical cognacy of basic vocabulary coded as shared inheritance form a common ancestral form (or, more frequently, root). As not all included etyma are likely to be reconstructible for the ultimate protolanguage, at least not in the same meaning, this translates to coding semantic shifts (but according to the outcomes, not input) and lexical replacement. The original state could very well be lost everywhere and thus not represented, or it could be preserved as an isolated state anywhere in the tree (Warnow et al. 1996: 317).

While the approach to cognacy data solves many of the inconsistencies of the traditional methods, the lack of directionality – and the underspecification of the innovations covered by lexical replacements, which is arguably one type of lexical innovation – still leave room for desirable discussion. Further, it remains open whether the replacement of the inherited lexicon (root cognates) over time is a meaningful or realistic proxy for language change and splits, although this question gets muddled by debates on the chronology and dating of changes, which is way beyond the scope here.

3.5. *The matrix format*

For the computational analyses of lexical cognacy, the data are presented in a matrix. As discussions of items in such matrices play a crucial role when discussing the analyses of borrowed lexemes further on in this article, an overview of some important concepts is in place here.

Usually, the data input in a matrix is constituted by comparable cognates derived from the same ancestor with the same meaning. The strength of this format is that it will not favour branches that have more attested material or are more thoroughly treated. All the analysed languages or branches will be included on equal terms, and the features will be "identifiable" across all analasyed languages (in the words of Peyrot 2022). The analysis will ideally reveal not just arguments in favour of a hypothesised and possibly presupposed subgroup but also the alternative or competing possibilities, thus not judging from positive evidence alone.

In the matrix, the data are organised as *characters* which are coded according to *states*. Characters refer to isoglosses, mostly in the form of lexemes in a given meaning. A state is a number assigned to all languages or branches sharing the given feature. This can be undertaken in different ways depending on what is needed for further analysis: as binary coding expressing the presence of absence of said shared feature, or as multi-state encoding where each individual number will correspond to a given feature.[2] In the case of cognacy coding, the same number will be assigned to all languages sharing the same etymon in the same meaning. The number in multi-state encoding is arbitrary, but it must be so that sharing

[2] Multi-state encodings can easily be converted to sets of binary characters. For an illustration, see Canby et al. (2024: 195–6).

a number corresponds to sharing an isogloss. In most datasets, synchronic synonyms are avoided even though it is very common in natural languages to have synonyms or near-synonyms for 'basic' concepts. The reason for this is that *polymorphism* remains a computational issue though these limitations depend on the specific computational analysis. Polymorphism can be avoided in several ways prior to the computational analysis. Ringe, Warnow and Taylor (2002: 84–5) give in-depth explanations of the coding strategies they employed. These are great illustrations of the processes. First, some characters can be reworked into non-polymorphic sets which creates interconnected duplicates. Second, *non-effective* polymorphism can be ignored, that is if only one of the synonyms is shared, the unique one would mean nothing for the analysis and can be left out safely. Third, if the polymorphism is confined to a 'known branch', it can be recoded as a single state – this heightens the risk of a circular outcome. Fourth, when irresolvable, polymorphic characters can simply be left out during the computational analysis. This final strategy is problematic as it leaves out very regular data because it cannot be coded according to a standard. Luckily, new methods and models can substantially improve this (Canby et al. 2024). *Unique states* are numbers exhibited by only one language and thus not shared anywhere in the dataset.

Further, it is not always enough for the data to fall into two distinct groups. The concepts of *informativity* and *compatibility* as well as the distinction between *compatible* and *forcing* characters in parsimony analyses can be extended to traditional phylogenetic argumentation. A character is *informative* if it can be fitted on to a tree. *Uninformative* characters occur when a character exhibits only one state shared by more than one language, either because only one language innovated uniquely (the rest having a shared state), or because only one grouping can be found, the rest of the languages showing unique states (Ringe et al. 2002: 98). A character is *compatible* if it can be fitted on to a given, usually the best tree, that is, if it can be displayed as having evolved without backformation or parallel innovations. A character which is compatible with any tree is (parsimony-)uninformative. Without known directionality, a compatible character 'does not *force* either subgroup; it is possible to accommodate this distribution of states in a tree in which one or the other of those larger subgroups is not posited' (Ringe et al. 2002: 104). This is the case for lexical cognates where the ancestral language (the root) could have had none, any, or all of the cognates coded. In other words, if we do not know the ancestral state, we cannot distinguish the innovation from the archaism, and thus the character is insufficient for a phylogenetic argument. It can, at best, *validate* a subgroup. Conversely, if the characters have directionality, they can *force* a subgroup, since such a shared character corresponds to sharing an innovation rather than simply observing a compatible distribution of states. Although this terminology is taken from a parsimony framework, it is a benefit of the matrix format we can carry into traditional methodology as well. However attractive a shared feature may seem, judging from positive evidence alone can be misleading. In many – if not most – cases there is no way of knowing if other languages took part in the same innovation but lost it or obscured it again afterwards. The matrix format assures that this possibility is at least considered.

4. LEXICAL REPLACEMENTS

4.1. How do new cognations arise?

We have settled that not all competing states coexisted at the root of the tree. What should be addressed here is *how* such a distribution arises. The matrix is after all just a tool for analysis, but the choices of assigning different numbers to different cognate classes within related languages correspond to lumping together many types of lexical innovations. In a matrix of lexical cognates in related languages, different states will occur because of different types of lexical innovations. Semantic shifts, such as broadening or narrowing of the meaning of near-synonyms, or formation of new words to different lexical roots (with appropriate derivational morphology) may account for many of these. Spontaneous or "iconic" creations (e.g. Campbell 2020: 103–13) are rather unlikely to occur in the basic vocabulary, and thus the cognate sets with a limited distribution and no etymology are quite mysterious.

4.2. Borrowings and common prestages

A very prominent way in which a language obtains new root etyma is through borrowing. And while borrowings do not provide any information on subgrouping at the time of the loan itself, they are not fundamentally different from other lexical innovations. Once they have happened, they are passed on to descendants of the recipient language and thus "work their way into the core vocabulary over time" (Ringe et al. 2002: 69). In this way, a new cognate can emerge and form an innovation in a clade:

> But if reflexes of a word borrowed into language X appear in the daughters of X, they are [...] cognates for the clade including X and its daughters, since within that clade they have been transmitted by genetic descent. (Nakhleh et al. 2005: 395)

When preparing lexical data for computational studies, this means that known borrowings are singled out, since these are compatible with any tree. If we were to assign the same state to the borrowing as the donor language, we would infer the wrong phylogeny – to take a classic example, Armenian does not become an Iranian language by borrowing Iranian lexemes. This has also been implemented systematically in the *IE-CoR* dataset which enables loanwords to have descendants and distinguishes them from parallel events (Heggarty 2021: 390; Heggarty et al. 2023b: 37–9).

Identifying borrowings into prehistorical stages of a language is complicated, and it is inevitable "that some unrecognised loanwords will slip through" (Scarborough 2020: 186). Two types are especially difficult to recognise: borrowings between closely related languages or still diverging dialects and prehistoric borrowings from donor languages, where the donor language is unknown or unattested, or when the context of the contact situation is obscure (Kassian et al. 2021: 957). The latter might indeed play a role in the distribution of competing unexplainable lexical roots or "regional isoglosses" that cause neighbouring branches to group together in some phylogenetic trees (Pereltsvaig & Lewis 2015: 84–8). For instance, the grouping of Italic, Celtic and Germanic into one clade based on lexical cognacy (Bouckaert et al. 2012: 959; Heggarty et al. 2023a: 4) is remarkably similar to early glottochronological studies (Tischler 1973: 96). See, however, Greenhill & Gray (2012: 527–8, 533) for a reply to similar counterarguments.

4.3. Shared borrowings from unknown sources and phylogeny

Some scholars have argued that we could make use of this unknown lexical element in phylogenetic argumentation. While these lexemes may not be easy to determine and define, they might still make up a rather large part of cognacy datasets. In his paper on Balkan Indo-European lexical isoglosses, Thorsø (2020: 251) argues:

> Sometimes, related forms in two or more branches may be traced to a common proto-form which at the same time cannot be reconstructed for the older parent language. When this is the case, and especially when similar, but incomparable forms are found

> among the other branches of the family, we can reasonably infer that the etymon was borrowed at a common pre-stage of these branches.

It is ingenious to combine our knowledge of how difficult it is to determine the reason for an isogloss distribution with the realisation that lexical replacements will be passed on as inherited no matter if their origin is from within the language itself or a borrowing from an outside source. It is crucial that Thorsø considers the methodological delimitation that it takes more than a simple coincidence or superficial agreement in form to be considered a common borrowing. If we are to make use of the unique and foreign-tasting lexicon for phylogenetics, we need a single preform of a suspected substrate word reconstructible for the branches analysed. However, some caution is still necessary since the material may not allow us to distinguish between "lexical correspondences which might have arisen through early borrowings from a third language or from each other" (Clackson 1994: 200).

In the following section, I will address two such instances where a regionally distributed lexeme has been argued to be borrowed into a shared pre-stage. I hope to underline the main argument, namely that loanwords and cognacy judgements always rely inherently on previous phylogenetic analysis.

5. **peleḱu-* '(BATTLE) AXE'

5.1. *A single 'axe'?*

It is relatively uncontroversial that the Greek and Indic words for '(battle) axe', Gr. *πέλεκυς pélekus* and Ved. *paraśú-* are somehow related. Based on the slender material, we could transpose a proto-form **peleḱu-*. At first glance, this is interesting for a number of reasons: the *u*-stem inflection, the palatal **ḱ* and the odd "triconsonantal" root-structure without an Indo-European derivational history. Throughout the Indo-Europeanist literature, there is a persistent claim that the word family is Semitic in origin, and an Akkadian source has often been invoked (see below). The comparison has been interpreted in many ways, often even disregarded, but it has long been suspected that the words are ultimately descendants of one or more borrowings (Wüst 1956: 13–31; Makkay 1998). While it would certainly be an interesting case for many reasons, a lot of the details do not stand closer scrutiny.

5.2. *Indic, Iranian, and Indo-Iranian*

The transposed preform **peleḱu-* is mostly based on the comparison of Greek and Indic. Ved. *paraśú-* shows developments of pre-Proto-Indo-Iranian age (satəmisation of **ḱ*, merger of **l* and **r*, and the merger of **e, a, o* > **a*). From Iranian, on the other hand, there is only circumstantial evidence from alleged loanwords of unattested varieties. Though rarely mentioned, even some Nuristani languages speak for an old age: Ashkun *pōs*, Kati *pεċ* 'large axe' can also go back to the Proto-Indo-Iranian form.

The comment of the coding of the Tocharian forms in the dataset of the Ringe-group[3] is a good starting point for the discussion:

> The Tocharian word was borrowed from some northeast Iranian reflex of [*peleḱus] *before* the Proto-Tocharian period; thus we assign both Tocharian languages the same state [*peretə], but a different state from [*peleḱus] (Ringe & Taylor 2012: 94)

[3] The character was not included in the final screened dataset because of its irresolvable polymorphism.

This means that they accept the analysis of the Tocharian forms as descendants of the same etymon, which is ultimately an Iranian word borrowed into (pre-)Proto-Tocharian: TB *peret*, TA *porat* < PT **peretə* (Ringe & Taylor 2007a: s.v. *306 axe*). Likewise, they view the Greek and Indic forms as descendants of the same etymon. It follows that they must thus recognise its existence in Proto-Iranian and Proto-Indo-Iranian, despite the lack of Old Iranian comparanda. Interestingly, the inclusion of this ultimate loanword into a higher-order subgroup is isolated in the dataset of the Ringe group (Don Ringe, p.c. March 28, 2022). Other loanwords are clearly marked with unique codings (like Proto-Tocharian **peretə* above), but this coding in favour of a relationship between Indic and Greek based on a loanword isogloss might go under the radar.

Middle and Modern Iranian do have some potentially cognate forms, most famously Ossetic *færæt* 'axe', but also Old Khot. *paḍa-* and Khwar. *pdyk*, which cannot go directly back to a Proto-Iranian **paraću*, as this would have given Oss. ***færæs* and Khot. ***parsa* (Abaev 1973: 451; Bernard 2023: 44). It has been argued that the forms look like borrowings from something like an unattested Old Persian **paraϑu-* (Abaev 1973: 451; Cheung 2002: 74).

This hypothesised loanword trajectory has been heavily and thoroughly criticised by Bernard (2023: 43–8). Among other issues, he points to (a) the lack of Western Iranian cognates of the etymon and the isolated distribution of such an Old Persian loanword, to (b) the fact that Khotanese and Khwarezmian had no need to render **ϑ* as **t* since they both preserved **ϑ* at the time, and to (c) the fact that another Old Persian word for 'axe', **tapara-*, is indirectly attested from Elamite *dabara*. Among these (b), the unnecessary sound substitution, is by far the most concerning. By contrast, (a) and (c) need not undermine the analysis. After all, Greek also had two words for axe, *πέλεκυς pélekus* and *ἀξίνη aksínē*, neither of which survived without reborrowing, analogy, and semantic shifts into Mod. Greek (Bampiniōtēs 2019: 46–7, 1604). Although there are of course many more descendants of Western Iranian than of Attic Greek, highly specific terms in the semantic domain of tools and trade might not necessarily be expected to survive millennia. In addition, it would not be an isolated case of an Iranian word only known from external comparanda (cf. Martirosyan 2013: 105 on Arm. *nirh* 'dormancy, slumber' ← Iranian **niδrā-*, cf. Ved. *nidrā́-*).

Bernard (2023: 46–7) also comments on the accent of Gr. *πέλεκυς pélekus*, though he doubts it is a cognate, and that the syncope of the Khotanese and Khwarezmian forms which also seem to go back to initially-stressed preforms. The mismatch of accentuation of the Greek and Vedic forms, at least, does not speak against a common preform as all but one Greek *u*-stem nouns have become accented on the initial syllable (Lubotsky 1988: 123–4).

However problematic the inner-Iranian situation is, Bernard's (2023): 48) final conclusion that Iranian, Indic, Nuristani – and Greek – independently borrowed **páratu-*, **paraćú-* – and **péleku-* – from unknown source languages (with para-Iranian sound laws?) is not necessarily economic either. Especially in the light that he, as many others, considers Semitic a probable donor language for the Greek form (Bernard 2023: 47, n. 19).

5.3. Akkadian pilaḳḳu 'wooden spindle'

A Semitic preform has been a persistent explanation of the origin of the word family. The connection with Akk. *pilakku/pilaḳḳu* has been observed frequently and for a long time (Kretschmer 1896: 107), but it has just as frequently been accompanied by the addition that *pilaḳḳu* does not mean 'axe' (Porzig 1954: 160; Frisk 1966: 497; Mayrhofer 1996: 87; Euler 1979: 145; Tremblay 2005: 16; Beekes 2010: 1167; Fortson 2010: 79). In fact, it has long been acknowledged that *pilaḳḳu* rather means '(wooden) spindle', and that it is most likely itself a borrowing from Sum. *balag* (Falkner 1952). Thus, the connection with Indo-European has indeed been abandoned many times (von Soden 1972: 863; D'iakonov 1985: 125; Verhasselt 2011: 259).

However, the connection is not just impossible for semantic reasons, but also because of morphology. Even within Semitic, the form is no longer accepted as inherited or derivable:

> All in all, while the common origin of the Sumerian, Akkadian and [Western Semitic] words is not in doubt in view of the striking phonetic and semantic similarity between them, no convincing reconstruction of the underlying historical developments can be proposed at present (Kogan et al. 2020: 254)

Nevertheless, Gamkrelidze & Ivanov (1995: 620) revived the connection claiming that there must have existed a homophone *pilakḳu* with a more appropriate meaning, derived from the verbal root **p-l-q*, which they gloss 'chop, split apart'. However, they do not touch upon the word formation of this 'clear correlation'. Similarly, Bjørn (2022: 21) and Bernard (2023: 47, n. 19) cite Watson (2013: 169–70) who mentions the possible Semitic connection between Gr. *πέλεκυς pélekus* and Aramaic *plq* 'axe'. The full form behind *plq* is Syr. Aram. *pelqā* 'hatchet, axe'; and other than being attested millennia later than the Akkadian form, the connection is equally problematic. First, D'iakonov (1985): 126) argues that it might be derived from a different root (**p-l-k* 'separate'). No matter the root, **pelqā* could follow productive patterns of Aramaic, not Proto-Semitic (Kaufman 1974: 82–3; O'Leary 1923: 178). Second, it has even been suggested that *pelqā* could be borrowed from Greek (Sokoloff & Brockelmann 2009: 1203; Tremblay 2005: 16, n. 14), but many remain sceptical (Kogan et al. 2020: 252), as the vowels would differ from other Greek borrowings (Butts 2016: 91). This vocalism should indeed trouble us a great deal. Given that no other nominal forms exist in Semitic, it is frankly impossible to imagine a formal match between Aramaic and Greek, no matter the directionality.

5.4. *Proto-Semitic *p-l-q 'to split, cleave'*

While the Semitic verbal root **p-l-q* 'to split, cleave' (Akk. *palāḳum* (I) 'slaughter', Arab. *falaqa* 'split apart', Tigre *fälḳä* 'split' (Kogan et al. 2020: 258)) does indeed look attractive, there are neither attested nouns nor word formation patterns to justify the connection with Indo-European. A hypothetical form that could underlie **pelek̂u-* would violate many rules of Semitic derivation:

Proto-Semitic does not allow nouns with the vocalic structure of two **i*'s in a triconsonantal root – a noun pattern ***qitil* to Semiticists (O'Leary 1923: 177–9). While other patterns would be allowed, it remains to be answered why the Indo-Europeans would change the vowels. The final *u*-vowel is also difficult: Affirmative (suffixal) *-*ū(t)* derives abstract nouns and does not pair well with the gemination of the third root consonant (Gragg & Hoberman 2012: 184, 175). Some see it as a fossilised accusative ending that got reinterpreted as the stem consonant in the Indo-European recipient languages (Bernard 2023: 47, n. 19), but this is quite ad hoc. Deverbal nouns are mostly derived with an affirmative prefix (Weninger 2011: 164) – in this instance, Akk. *naplaḳtu* 'butchering knife' (Kogan et al. 2020: 257–8) would probably be what a hypothetical Semitic source would have looked like.

Even if we were to accept the formal and semantic match between **pelek̂u-* and **p-l-q* for a moment, we would still have to account for the fact that a Semitic back-velar /q/ was borrowed as a palato-velar **k̂* in (pre-)Indo-Iranian. This is, of course, irrelevant to Greek and does not speak against independent borrowings, especially not if what we note as **k̂* was the unmarked velar which was later fronted in the satəm languages (Sihler 1995: 153–5), or if there never was a plain velar in Proto-Indo-European (Kortlandt 1978). However, the diverging sound changes prove that the borrowing(s) must be relatively ancient, at the very least into Indic. Unless, of course, we prefer to project these phonemic differences onto unattested languages.

5.5. *How many borrowing events?*

The word family of **pelek̂u*- 'axe' looks strangely similar across the "Indo-Greek" area, but the word cannot be inherited or derived from an Indo-European root. It could possibly be a common borrowing at an early shared state:

> Importantly, this Semitic word was not borrowed independently by Indo-Iranian and Greek after the separation of the two branches, but while a Greek-Aryan dialect community still existed. (Gamkrelidze & Ivanov 1995: 620)

Facing the choice between a single borrowing into a probable parent language and two independent borrowings that are so early that the result would be the same anyway, the former option is – all other things being equal – the most parsimonious one. However, all other things are not equal, and especially the internal Iranian evidence requires problematic borrowings of unattested Old Persian-looking forms. Indeed, Bernard argues against an early borrowing and in favour of two or more donor languages: one from which Iranian (and from there Tocharian) took the word, a separate one from which Indic and Nuristani borrowed it, and finally one from which Greek borrowed it. In this scenario, the otherwise understandable consonantism (*ś* : *k* : *t*) is projected on to unattested languages.

The root connection with Semitic **p-l-q* remains a tempting explanation, but there are neither nominal comparanda nor patterns of word formation to qualify it. The usual label 'Wanderwort' or 'culture word' is unhelpful, but I see no point in postulating an unattested source. However, it is difficult at best and tautological at worst to argue in favour of a subgrouping based on a questionable borrowing event.

6. **(H)a(i̯)ĝ*- or *(h_2)*aĝ*- and **ai̯ĝ*- 'goat'

6.1. *A single 'goat'?*

A similar case has been argued for the word for '(she-)goat' in the 'Balkanic' languages: Gr. *αἴξ*, *αἰγός*/*aíks, aigós*; Arm. *ayc*, Alb. *dhi* (Matzinger 2006: 25). Thorsø (2020; Olsen & Thorsø 2022) specifically argues that they all descend from a preform **ai̯ĝ*- which goes back to a single borrowing event into a common Balkan Indo-European subgroup, and that Indo-Iranian and Slavic share the ancestor **aĝ*-, which might ultimately be a borrowing from a similar or related form. Cf. similarly Kortlandt (2016: 361).

This analysis contrasts the stages coded by the Ringe-team (Ringe & Taylor 2007a: 108; 2012: s.v. *344 goat*) who connect Alb. *dhi* (fem.) with Ved. *ajá*- (masc.) 'billy goat', Lith. *ožkà* 'she-goat' and OPr. *wosee* (E) as descendants of **aĝ*-, whereas they view Arm. *ayc* as a descendant of **ai̯ĝ*-. Gr. *αἴξ aíks* is not included in the dataset on purpose (see below).

The reconstruction of the onset of the forms is also unclear. There is no direct evidence for an initial laryngeal. However, the generally accepted rules of PIE phonotactics do not allow a root to begin with a vowel. To some scholars, the otherwise initial "root vowel" **a* would be reason enough to reconstruct an initial *h_2, but the interplay of laryngeal colouring and PIE vocalism is disputed. Reconstructing *h_2*e(i̯)ĝ*- (or *h_2*a(i̯)ĝ*- with laryngeal colouring) is surely possible. Other scholars could also reconstruct *h_1a(i̯)ĝ- (or even *h_3a(i̯)ĝ-, for that matter) with an ablauting root vowel **a*. There may, however, be external evidence for an initial laryngeal if the connection with the Caucasian languages discussed below is to be accepted. Some authors (Porzig 1954: 163; Mann 1984: 6; Mayrhofer 1986: 51; Bird 1993: 9; Clackson 1994: 90, 219) think it possible to connect the two preforms *(*H*)*aĝ*- and *(*H*)*ai̯ĝ*-. If that is the case, influence from or blending with the verbal root *h_2eĝ- 'to lead, drive' could have led to the loss of *-*i̯*- in Indo-Iranian and Balto-Slavic (Mallory & Adams 1997: 229). An anonymous reviewer points out to me that this is perhaps highly likely given the tendency of Vedic to express even basic concepts with synchronically understandable derivatives of roots rather than opaque label nouns. It is indeed rather common to etymologise zoonyms in this way, cf. for instance the discussion of *h_1éḱu̯o- 'horse' as *'having swiftness' (Höfler 2024: 65–75).

6.2. *Filling the gaps*

The analysis of Ringe, Warnow and Taylor is the result of a strict methodology on the phonological, semantic and chronological sides, but quite a relaxed one on the morphological side. It is worth noting that in their dataset, the semantic slot 'goat' seems not to have been limited to a specific sex. In their data, there are generic terms applicable to both sexes (like Goth. *gaits*, though grammatically masculine), but also mostly feminine terms (Alb. *dhi* grammatically fem. 'goat, nanny goat'), exclusively feminine terms (Lith. *ožkà* 'she-goat', Lat. *capra* 'id.', ON *geit*) and even exclusively masculine ones (Osc. acc. *kaprum*; Ved. *ajá*- 'billy goat' – though the corresponding feminine *ajā́*- is attested in R̥V 8,70,15).

Judging from the coding of this character alone, **aĝ*- could be a shared isogloss of Indo-Iranian and Balto-Slavic ("Indo-Slavic") – and Albanian. However, there are several gaps in the dataset. Some of these could be filled if we were to include more material from closely related languages or additional evidence from derivatives. Avestan and Old Persian do not attest the word for goat, and Old Church Slavonic, the only representative of the Slavic branch in the dataset, has the unique *koza*, which has been borrowed in Latvian (as *kaza*). Armenian *ayc* (continuing **ai̯ĝ*-) and Greek (*χίμαιρα khímaira*, see below) also exhibit unique states and could fit anywhere on any tree. The Iranian gap could perhaps have been filled by the same etymon as Indic:. Young Avestan might have a single attestation of *aza*- < **aj́-á*- (Hoffmann 1967), and Middle Persian attests *'z* /az-/ 'goat', both of which would fit Vedic *ajá*- exactly. The Young Avestan adjective *īzaena*- 'leathern' has also been analysed as a derivative showing an otherwise unattested zero-grade *(*H*)*iĝ*- of *(*H*)*ai̯ĝ*- (Mayrhofer 1986: 51), but this is rather doubtful since Indic shows no trace of the *-*i̯*- (Thorsø 2020: 254). Cf, however, Palmér (2024: 43–5).

Slavic attests a similar derivative, but not the noun from which it must be derived: Serb. Church Slav., ORuss. *jazъno* 'leather' could go back to **aĝ-ino*-, mirrored in Ved. *ajína*- 'fur of a black antelope or tiger' (AV+) and Lith. *ožìnis* 'pertaining to a billy-goat' – possibly even Alb. *dhirë/-në*. Hock et al. (*ALEW*[2]: 830) consider an Iranian loan origin in Slavic possible, but this can of course only be true if the etymon was preserved in Iranian. The chronology of an Iranian loan before Winter's law is also peculiar.

Baltic has several other relevant forms, only they are not attested in the specific meaning 'goat'. Lith. *ožỹs* and Latv. *âzis* mean exactly 'billy goat' and continue a masculine *i̯i̯o*-derivative of the same etymon, **aĝ*-. Had morphological subgroupings been taken into account in the coding of this character, these derivatives would have added material in favour of an East Baltic group; just like the cognate PIIr. **aj́-á*- would do for Indo-Iranian.

*6.3. Albanian edh 'goat kid' and Balkanic *ai̯ĝ-*

One of the analyses that separate Thorsø and Ringe, Warnow and Taylor is the treatment of Alb. *dhi*. Ringe & Taylor (2012: 108–9) stated clearly that they 'accepted the usual etymology of the Albanian word', being a continuation of **aĝ*-, but noted that the alternative would have been to connect Albanian with Arm. *ayc (continuing *aĝ-)*. This alternative analysis has been advanced by Thorsø (2020: 253–5), who argues that Alb. *dhi* should be connected not just with Arm. *ayc*, but also Hom. Gr. *αἴξ aíks*.

An advantage of this analysis is that it can connect Alb. *dhi* (fem.) 'goat' with *edh* (masc.) 'goat kid'. Within Albanian, these are difficult to separate, and the connection may be more transparent in Proto-Albanian: *edh* can go back to Proto-Albanian **aidza* (a regular reformation of the root-noun **ai̯ĝ*-), and *dhi* can go back to an *ii̯ā*-derivative of the same 'root': **aidzijā* < **ai̯ĝ-ii̯ā*-. According to some, the connection can best be upheld if the forms go back to **ai̯ĝ*-, and not **aĝ*- (Thorsø 2020: 254; Orel 1998: 85). Unfortunately, the pretonic vowel would also be lost in **adzijā* > *dhi* (Demiraj 1997: 160). The argument thus hinges on whether or not *edh* can go back to **aĝ*-, which is problematic on formal and semantic grounds. A root-noun **aĝ*- or a thematic derivative **aĝ-o*- (the preform of Ved. *ajá*-) would probably have yielded Alb. **adh*. The preform of Lith. *ožỹs*, **aĝ-ii̯o-*, would probably have given **dhi* as well, which would be phonetically identical to *dhi* (fem.) 'goat', but masculine! The exact causes of umlaut are unclear, and it is perhaps possible that a hypothetical **aĝ-ii̯o*- could mirror *i keq* 'bad' < **kakii̯o*- and have become **edh* as well. However, if we postulated an otherwise unknown *i*-stem ***aĝ-i*-, we might expect its outcome to be *edh*. The details of Albanian remain rather opaque, and the forms are not self-evidently forcing a Balkanic isogloss **aiĝ*- to the exclusion of **âg*-.[4] Finally, *edh* could be a borrowing from Latin *haedus* 'goat kid' in which case this entire exercise is pointless (Meyer 1891: 98).

6.4. Homeric αἴξ aíks

I have circled around the important Greek form Hom. *αἴξ aíks*. It is left out of Ringe, Warnow and Taylor's dataset for the following reason:

> Greek is represented by Classical Attic rather than the earlier Homeric not only because the attestation of Attic is far more extensive, but also because Homeric Greek is known to be an artificial literary dialect. (Nakhleh et al. 2005: 393)

In Attic, the regular word for 'goat' is *χίμαιρα khímaira*. It is reasonable to choose the variety best equipped for filling out a word list of semantic slots. However, the fear of an anachronistic dataset is an inheritance from glottochronology and its derived disciplines, and it is most relevant when trying to date speciation events on the basis of lexical replacement (Nakhleh et al. 2005: 400). Adding cognates that no longer fit the semantic slot because of semantic shifts obviously violates the criteria, but it is important to realise that including older layers of the same language after its final split-off from any related languages only interferes with chronology, not with phylogeny, and therefore will not skew the results.

[4] Yet another alternative connects Alb. *dhi* with OHG *ziga* (Meyer 1891: 85; Demiraj 1997: 160; Topalli 2017: 430). OHG *ziga* is actually included in the Ringe wordlist, but not in the cognations, where OHG is only represented by *geiz*, a descendent of **gait*- (Ringe & Taylor 2007a: 49; 2012: 108). The reason for this coding choice is that the lack of cognates makes it not 'effectively polymorphic', and *ziga* can thus be left out without consequence (Ringe et al. 2002: 84). If Alb. *dhi* had been analysed as a descendent of the same etymon as OHG *ziga*, a polymorphic isogloss would split up the Germanic unity for this character.

This approach gives the well-attested Greek branch a minor disadvantage and makes it look more isolated or uninformative than it could have been. Attic has simply had more time to replace inherited cognates. Elsewhere, where the attestations of the included languages are more fragmentary, other tactics have been employed, and some gaps have been filled by forms inferable from derivations. Examples include the Old Prussian cardinal numerals – and the Tocharian B word for 'goat', which is represented by **ās* in the dataset, even though the simplex was not attested at the time of compilation. Instead, it was inferred from the derived adjective *aṣiye* 'pertaining to a goat', derived from a descendant of PT **asə*, as also TA *ās* 'goat'. In the meantime, TB *ās* 'goat, ewe' has securely been identified (Pinault 1998; Adams 2013: 62). More strikingly, the 'Very Old Latin' (dated ca. 600–300 BCE, (Weiss 2020: 24, 239)) reflex of the thematic gen. sg. ending **-osi̯o* has been coded for in the morphological data, despite Classical Latin dated ca. 100 BCE being the representative variant (Ringe et al. 2002: 81; Ringe & Taylor 2007b: 6–7).

Greek *αἴξ aíks* does occur in Homer and derivatives of it in Mycenaean as well as alphabetical Greek; though not in Attic, and the form *could* thus have been inferred for Proto-Greek, had the Ringe group chosen this approach. The concept of basic words is artificial to begin with, and even more so in poetic language. However, ruling Homeric Greek lexicon out because the language is artificial might be a bit unnuanced. It is an artificial variety in the sense that it mixes dialectal lexicon and grammar with synthetic forms patterned on archaic types or productive morphology; but we would not expect to find artificially invented lexical roots. This only shows that the analysis behind each and every cognation does have an impact on the outcome.

6.5. Satəm grouping or Balkan borrowing?

Greek *αἴξ aíks* and the Armenian nominative *ayc* clearly reflect a root-noun, but as Thorsø (2020: 255) remarks, perhaps the unusual compound vowel of forms like *αἰγίβοτος aigíbotos* 'grazed by goats' (Hom.+) and possibly Myc. *a*$_3$*-ki-pa-ta*, interpreted as *ai̯gipa(s)tās* 'goat heard',[5] is connected to the Armenian oblique stem *ayci-*, continuing an *ih*$_2$-derivative. If that is the case, Greek, Armenian and Albanian share not only the root variant **ai̯ĝ-* but also the complex of a root-noun and a collective derivative. Even Mycenaean attests the *i̯o*-stem adjective *a*$_3$*-za* in the phrase *di-pte-ra a*$_3$*-za* /dipʰtʰerā aigi̯ā/ 'goat hide' on PY Ub 1318 (Bernabé & Luján 2019: 573; Clackson 1994: 218–9 n. 10, 18–9; *pace* Jucquois & Devlamminck 1977: 22), which – if formed to the *ih*$_2$-stem would add further evidence for the complex.

However, since it is unclear which etymon the Indo-European protolanguage had filling the semantic slot for 'goat', it is impossible to distinguish the innovative subgroup from the branches simply preserving archaisms. As the previous discussion shows, the words for 'goat' have been or can be used to argue in favour of various subgroupings. Again, taking the coding of the Ringe group as a starting point, and adding additional tweaks to their analysis, we could arrive at arguments in favour of the following subgroupings:

a An Albano-Indo-Slavic group based on positive evidence from Baltic, Indic, and Albanian. This analysis rests on the separation of the etyma **aĝ-* and **ai̯ĝ-* and analysing Alb. *dhi* as a descendant of **aĝ* (possibly separating *dhi* from *edh*). The lexemes filling the

[5] Thorsø is probably right in glossing *a*$_3$*-ki-pa-ta* as a noun and not a personal name, though this is unclear from his references to Duhoux (2008: 295), who in passing mentions the broken and restored *[a*$_3$*-]ḳi-p̣a-ta* in a lacuna-filled context on PY Ae(1) 489, and Adrados (1985: 135), who separates the appellative of a man in PY Ae(1) 108 from the masc. pers. name in KN Fh 346. It further occurs on PY Ae(1) 264. Recent scholarship seems to have settled on the interpretation of the form as the occupational term of a goat herd (Rougemont 2019: 314; Bernabé & Luján 2020: 83).

semantic slots in Slavic and Iranian add no direct evidence in the strictest sense for or against this cognate grouping but the character is still compatible with such a phylum. Some derivatives (e.g. Slav. **jazьno*) might be easier explained if the etymon existed in the common ancestor.

b A satəm-group: If **aĝ* and **ai̯ĝ* are taken as the same etymon, and the variation is assigned to later influence from the root **h₂eĝ-* 'lead', Armenian *ayc* could also descend from the same form.

c A Graeco-Satəm group: If Hom. *αἴξ aíks* is taken into account, the character could support not just a satəm-group but a Graeco-Satəm grouping.

d A Balkan group opposing an Indo-Slavic group: if *αἴξ aíks* is taken into account, and Alb. *dhi* had been coded as **ai̯ĝ-*, the character would support a Balkan group consisting of Greek, Armenian, and Albanian as opposed to an Indo-Slavic group sharing **aĝ-*.

This last item (d) is the argument advanced by Thorsø – based on Kroonen (2012: 246). He further substantiates the point that while the etyma **aĝ-* and **ai̯ĝ-* are too far apart to be reconstructed as a single proto-form, they are so similar that we might suspect parallel borrowings from the same or similar sources. The morphological peculiarities are only shared in the Balkan group, which Thorsø (2020: 254) – and Matzinger (2006: 25) – see as an innovation, but Clackson (1994: 90) as an archaism. However, in a stratified family tree, archaisms can stem from any intermediate protolanguage.

6.6. *Ex Caucaso capra?*

The claim that the etymon or etyma are ultimately borrowings would of course be more convincing if a phonologically, morphologically, culturally, and geographically realistic donor language could be proposed. Given that we are dealing with a root etymology with no other meaning in Indo-European, the morphological criterion is less problematic.

A Caucasian source language has in fact been suggested: Proto-Nakh-Dagestani (PND, also known as North-East Caucasian) or – if it exists – Proto-North Caucasian (PNC; Kortlandt 2019: 385; Matasović 2012: 290). PND has a word for 'goat' reconstructed as **Hēǯ́V*, possibly going back to PNC **ħēǯ́ü* (Starostin 2007a: 292)[6] or **ʔēǯ́wē* (Nikolayev & Starostin 1994: 245). There is a superficial similarity between the Caucasian and Indo-European proto-forms, but the comparison rests on narrow grounds. There seems to be a correspondence between Adyghe (West Caucasian) *āča* 'billy goat' and the ND forms (Chech. *awst* 'goat kid', Chirag Agul *ʕač:a* 'goat') which is the argument in favour of reconstructing this lexeme to PNC instead of regarding it an Indo-European loanword in this branch (Starostin 2004: 214; Nikolayev & Starostin 1994: 245). However, the reconstruction is idiosyncratic, and many developments are unparalleled: The initial consonant is uncertain (but does, of course, resemble an Indo-European initial laryngeal before **a*), the **j* is highly speculative, and the cluster **ǯ́w* may equally well be **st*. Consequently, **HēstV-* looks a lot less like the preform **ai̯ĝ-* (based on the Indo-European evidence alone) than **Hēǯ́V* resembles an automatic transposition **h₂ei̯ĝ-*.

It is morphologically peculiar that this word, clearly ending in a vowel in the suggested donor language, should have been borrowed and incorporated as a root-noun. Phonologically, the oscillation between **a* and **ai̯* within the Indo-European languages has also been argued to result from an adaptation of a foreign phoneme (Matasović 2012: 90, n. 16. See also above). Another peculiarity is that the Caucasian "voiced hissing-hushing (= palatalised)

[6] Other notations: **Hēǯ́ü* (Starostin 1985: 76; 2004: 234), **Hēǯü* (Starostin 2004: 214), and **Hēǯ́u* (Starostin 2009: 80; 1988: 114; 2007b: 315).

affricate" (Nikolayev & Starostin 1994: 21) *ʒ́ should have been borrowed as a palato-velar stop *ĝ. This match fits a more conservative reconstruction of the PIE velars better than the progressive scenarios in which the palatal feature arose later (e.g. Sihler 1995: 153–5; cf. also Kortlandt 1978).

It is not impossible to come up with a scenario in which the speakers of the protolanguages involved were geographically close, but there seems to be no plausible archaeological reason for assuming a Caucasian borrowing since the domestication of the goat (*Capra hircus*) was wide-spread in Eurasia already after the last ice age (Mallory & Adams 1997: 230 with refs.). A solution according to which the different roots denoted different kinds, shapes, breeds, or colours of goats would be pure speculation.

In total, the judgement of the loanword origin is rather circular: PIE needs to have been spoken close to PNC in order for the borrowing event to take place, but the analysis of the form as a Caucasian loanword is mostly backed by the very same geographical suspicion. There may of course be additional arguments or a bulk of evidence to speak for such a solution, but on this lexical level, we cannot safely argue in favour of Caucasian connections. Similarly, we can indeed use the distribution of the forms of this isogloss to argue for various subgroupings of the Indo-European languages; but for the most part, the navigation between many complex and intertwined etymological analyses are to some extent dependent on the very subgroups we are trying to argue for to begin with.

7. Conclusion

In the precious sections, I have shown how the behind-the-scenes analysis of cognacy assignment and loanword detection in lexical data have vast consequences for later analysis. I have shown that loanword judgement directly affects the analysis, that preconceived ideas are deeply embedded in cognate coding, and that a certain bias in the analysis is unavoidable.

The word for 'axe' **peleḱu-* is probably borrowed into some stage of post-PIE, but despite numerous attempts at a connection with the Semitic verbal root **p-l-q* 'to split', neither Semitic nor Indo-European derivational patterns can explain the word formation. If it is treated as a single borrowing, it is compatible with an Indo-Greek clade (where the number of daughter branches is unclear as the lexeme could have been lost again). However, within Indo-Iranian, the material points in opposite directions: The sound changes in Nuristani and Indic point to Proto-Indo-Iranian age, and the Iranian material might rather point to independent later borrowings, which in turn relies on extra assumptions about the donor languages. If the lexemes were indeed treated as separate borrowings, the etymon would become completely uninformative for phylogenetic argumentation. The lack of attestations from most branches makes the danger of judging from positive evidence alone clear.

The word for 'goat' **(H)a(i̯)ĝ-* could be an archaism, a borrowing, or even several independent borrowings possibly from different languages – or a blending of inherited and borrowed forms. The character could speak in favour of 'Graeco-Satəm' clade (if the variation in the two root forms **aĝ-* and **ai̯ĝ-* is seen as internally 'Indo-Slavic'), or an Indo-Slavic branch (showing **aĝ-*) opposing a Balkan group (**ai̯ĝ-*) if the two root forms are kept apart. The analysis of the ambiguous Albanian forms *edh* and *dhi* play a key role in deciding between the two. If it is foreign influence, Nakh-Dagestani or North-Caucasian have been suggested as the source, but this rests on idiosyncratic reconstructions of very uncertain consonant clusters.

When comparing the two proposed lexemes, it is striking that the loan trajectories would have to mirror images of each other: The Caucasian word for goat would alledgedly be borrowed without its final vowel and become a root-noun, and its affricate or sibilant consonant (or cluster) would end up as a palatalised velar stop, whereas the uvular or back-velar stop of the pseudo-Semitic word for 'axe' would be adopted as a fronted velar – but most importantly, it would be borrowed as a *u*-stem either on the basis of a Semitic case ending – not a stem vowel – or by some rare spontaneous stem class reassignment.

Cognate character coding, etymology and loan word analysis are interdependent in nature. Raw lexical data contain no phylogenetic information, and coding cognates as data points cannot be undertaken completely without bias in the analysis. Lexical data in general and borrowings in particular need not be unusable for phylogenetic purposes, but it is difficult to see them standing alone, no matter the approach. This should not prevent us from trusting computational studies on cognacy databases, but it does call for careful coding and open discussions of the analyses chosen – and the consequences of their alternatives.

Correspondence
Simon Poulsen
Department of Nordic Studies and Linguistics
University of Copenhagen
Emil Holms Kanal 2
2300 København S
Denmark.
Email: simon.poulsen@hum.ku.dk

References

Abaev, V. I., 1973. *Istoriko-ètimologičeskij slovar' osetiskogo jazyka*, Volume 1. A–K. Moskva & Leningrad: Nauka.

Adams, Douglas Quentin, 2013. *A Dictionary of Tocharian B* (2nd ed.). (Leiden Studies in Indo-European 10). Amsterdam & New York (NY): Rodopi.

Adrados, Francisco Rodríguez, 1985, in Francisco Aura Jorro (ed.), *Diccionario micénico*, Volume 1. (Diccionario Griego-Español 1). Madrid: Consejo Superior de Investigaciones Científicas, Instituto de Filología.

Adrados, Francisco Rodríguez, 1991. 'Dictionaries of Indo-European and their problems', *Lexicographica* 7. 11–24.

ALEW[2] = Hock, Wolfgang, Elvira-Julia Bukevičiūtė, Christiane Schiller, Rainer Fecht, Anne Helene Feulner, Eugen Hill & Dagmar S. Wodtko, 2020. Altlitauisches etymologisches Wörterbuch. 2.0 https://alew.hu-berlin.de/ (26 January, 2023)

Bampiniōtēs, Geōrgios, 2019. *Etymologiko lexiko tes neas Ellenikēs glōssas: Istōria tōn lexeōn, me schólia kai énthetoys pínakes* (5th ed.). Athēna: Kentro Lexikologias.

Beekes, Robert Stephen Paul, 2010. *Etymological Dictionary of Greek*. (Leiden Indo-European Etymological Dictionary Series 10). Leiden & Boston (MA): Brill.

Bernabé, Alberto & Eugenio R. Luján, 2019. 'Testi relativi a pelle e manufatti in pelle', in Maurizio Del Freo & Massimo Perna (eds.), *Manuale di Epigrafia micenica: Introduzione allo studio dei testi in lineare B*, Volume 2. Padova: Libreriauniversitaria.it. 567–587.

Bernabé, Alberto & Eugenio R. Luján, 2020. *Introducción al griego micénico: Gramática, selección de textos y glosarios*. 2[a] edición, corregida y aumentada. (Monografías de Filología Griega 30). Zaragoza: Prensas Universitarias de Zaragoza.

Bernard, Chams Benoît A. S., 2023. *Like Dust on the Silk Road: An Investigation of the Earliest Iranian Loanwords and of Possible BMAC Borrowings in Tocharian*. Leiden: Universiteit Leiden PhD Dissertation (Proefschrift).

Bird, Norman, 1993. *The Roots and Non-roots of Indo-European: A Lexicostatistical Survey*. Wiesbaden: Harrassowitz.

Bjørn, Rasmus G., 2022. 'The lexicon of an old European Afro-Asiatic language: Evidence from early loanwords in proto-Indo-European', *Historische Sprachforschung* 135(1). 3–42.

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.

Butts, Aaron Michael, 2016. *Language Change in the Wake of Empire: Syriac in its Greco-Roman Context*. University Park (PA): Pennsylvania State University Press.

Campbell, Lyle, 2020. *Historical Linguistics: An Introduction* (4th ed.). Edinburgh: Edinburgh University Press. https://doi.org/10.1515/9781474463133

Canby, Marc E., Steven N. Evans, Donald Ringe & Tandy Warnow, 2024. 'Addressing polymorphism in linguistic phylogenetics', *Transactions of the Philological Society* 122(2). 191–222. https://doi.org/10.1111/1467-968X.12289

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

Cheung, Johnny, 2002. *Studies in the Historical Development of the Ossetic Vocalism*. (Beiträge zur Iranistik 21). Wiesbaden: Reichert.

Clackson, James, 1994. *The Linguistic Relationship between Armenian and Greek*. (Publications of the Philological Society 30). Oxford & Cambridge (MA): Blackwell.

Clackson, James, 2022. 'Methodology in linguistic subgrouping', in Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*. Cambridge: Cambridge University Press. 18–32. https://doi.org/10.1017/9781108758666.002

Demiraj, Bardhyl, 1997. *Albanische Etymologien: Untersuchungen zum albanischen Erbwortschatz*. (Leiden Studies in Indo-European 7). Amsterdam & Atlanta (GA): Rodopi. https://doi.org/10.1163/9789004653757

D'iakonov, Igor Mikhailovich, 1985. 'On the original home of the speakers of Indo-European', *Journal of Indo-European Studies* 13(1–2). 92–174.

Duhoux, Yves, 2008. 'Mycenaean anthology', in Yves Duhoux & Anna Morpurgo Davies (eds.), *A Companion to Linear B: Mycenaean Greek Texts and their World*, Volume 1. (Bibliothèque Des Cahiers de l'Institut de Linguistique de Louvain. Antiquité 120). Louvain-la-Neuve & Dudley (MA): Peeters. 243–394.

Euler, Wolfram, 1979. *Indoiranisch-griechische Gemeinsamkeiten der Nominalbildung und deren indogermanische Grundlagen*. (Innsbrucker Beiträge zur Sprachwissenschaft B30). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Falkner, Margit, 1952. 'Ist Idg. **pelek̂us* ein akkadisches Wort?', in Wilhelm Brandenstein (ed.), *Studien zur indogermanischen Grundsprache, 26*. (Arbeiten Aus dem Institut für Allgemeine und Vergleichende Sprachwissenschaft 4). Wien: Gerold.

Fortson, Benjamin Wynn, I. V., 2010. *Indo-European Language and Culture: An Introduction* (2nd ed.) (Blackwell Textbooks in Linguistics 19). Chichester & Malden (MA): Wiley-Blackwell.

Friis, Louise Skydsbjerg, 2024. *Studies in Tocharian Verbal Morphology Relevant to the Cladistic Position of Tocharian in Indo-European*. Leiden: Universiteit Leiden PhD Dissertation. https://hdl.handle.net/1887/4093377 (3 October, 2024)

Frisk, Hjalmar, 1966. *Griechisches etymologisches Wörterbuch*, Volume 2. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher 17). Heidelberg: Carl Winter.

Gamkrelidze, T'amaz & Vjačeslav V. Ivanov, 1995. *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture*. (Trends in Linguistics. Studies and Monographs 80). Berlin & New York (NY): Mouton de Gruyter. https://doi.org/10.1515/9783110815030

Goldstein, David M., 2022. 'There's no escaping phylogenetics', in Laura Grestenberger, Charles Reiss, Hannes A. Fellner, & Gabriel Z. Pantillon (eds.), *Ha! Linguistic Studies in Honor of Mark R. Hale*. Wiesbaden: Reichert. 71–91.

Gragg, Gene & Robert Hoberman, 2012. 'Semitic', in Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic Languages*. (Cambridge Language Surveys). Cambridge: Cambridge University Press. 145–235.

Gray, Russell D. & Quentin D. Atkinson, 2003. 'Language-tree divergence times support the Anatolian theory of Indo-European origin', *Nature* 426(6965). 435–439.

Greenhill, Simon J. & Russell D. Gray, 2012. 'Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation', *Diachronica* 29(4). 523–537.

Greenhill, Simon J., Paul Heggarty & Russell D. Gray, 2021. 'Bayesian phylolinguistics', in Richard D. Janda, Brian D. Joseph, & Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*. Hoboken (NJ): John Wiley & Sons. 226–253. https://doi.org/10.1002/9781118732168.ch11

Heggarty, Paul, 2021. 'Cognacy databases and phylogenetic research on Indo-European', *Annual Review of Linguistics* 7(1). 371–394.

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, Benedict King, Remco Bouckaert, Lechosław Jocz, Martin Joachim Kümmel, Thomas Jügel, Britta Irslinger, Roland Pooth, Henrik Liljegren, Richard F. Strand, Geoffrey Haig, Martin Macák, Ronald I. Kim, Erik Anonby, Tijmen Pronk, Oleg Belyaev, Tonya Kim Dewey-Findell, Matthew Boutilier, Cassandra Freiberg, Robert Tegethoff, Matilde Serangeli, Nikos Liosis, Krzysztof Stroński, Kim Schulte, Ganesh Kumar Gupta, Wolfgang Haak, Johannes Krause, Quentin D. Atkinson, Simon J. Greenhill, Denise Kühnert & Russell D. Gray, 2023a. 'Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages', *Science* 381(6656). eabg0818. https://doi.org/10.1126/science.abg0818

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, Benedict King, Remco Bouckaert, Lechosław Jocz, Martin Joachim Kümmel, Thomas Jügel, Britta Irslinger, Roland Pooth, Henrik Liljegren, Richard F. Strand, Geoffrey Haig, Martin Macák, Ronald I. Kim, Erik Anonby, Tijmen Pronk, Oleg Belyaev, Tonya Kim Dewey-Findell, Matthew Boutilier, Cassandra Freiberg, Robert Tegethoff, Matilde Serangeli, Nikos Liosis, Krzysztof Stroński, Kim Schulte, Ganesh Kumar Gupta, Wolfgang Haak, Johannes Krause, Quentin D. Atkinson, Simon J. Greenhill, Denise Kühnert & Russell D. Gray, 2023b. 'Supplementary materials for "Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages."', *Science* 381(6656). eabg0818. https://www.science.org/action/downloadSupplement?doi=10.1126%2Fscience.abg0818&file=science.abg0818_sm.pdf (27 July, 2023)

Hock, Hans Henrich, 1988. *Principles of Historical Linguistics* (2nd ed.). Berlin, New York (NY) & Amsterdam: Mouton de Gruyter. https://doi.org/10.1515/9783110871975

Hoenigswald, Henry M., 1966. 'Criteria for the subgrouping of languages', in Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European Dialects: Proceedings of the Conference on Indo-European Linguistics, Held at the University of California, los Angeles, April 25–27, 1963*. Berkeley (CA) & Los Angeles (CA): University of California Press. 1–12. https://doi.org/10.1525/9780520312418-002

Hoffmann, Karl, 1967. 'Drei indogermanische Tiernamen in einem Avesta-Fragment', *Münchener Studien zur Sprachwissenschaft* 22. 29–38.

Höfler, Stefan, 2024. 'Linnaean linguistics: 'Bear', 'horse', 'wolf' and the Indo-European phylogeny from a zoographical perspective', in Jenny Larsson, Thomas Olander & Anders Richard Jørgensen (eds.), *European Interfaces: Integrating Linguistics, Mythology and Archaeology. (Stockholm Studies in Indo-European Language and Culture 1)*. Stockholm: Stockholm University Press. 57–90. https://doi.org/10.16993/bcn.d

Jacques, Guillaume & Johann-Mattis List, 2019. 'Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them)', *Journal of Historical Linguistics* 9(1). 128–167.

Jucquois, Guy & Bernard Devlamminck, 1977. *Compléments aux dictionnaires étymologiques du grec ancien*, Volume 1. A-K. (Bibliothéque Des Cahiers de L'Institut de Linguistique de Louvain 8). Louvain: Peeters.

Kassian, Alexei, George Starostin, Anna Dybo & Vasiliy Chernov, 2010. The Swadesh wordlist: An attempt at semantic specification. *Journal of Language Relationship • Вопросы языкового родства* 4. 46–89.

Kassian, Alexei S., Mikhail Zhivlov, George Starostin, Artem A. Trofimov, Petr A. Kocharov, Anna Kuritsyna & Mikhail N. Saenko, 2021. 'Rapid radiation of the inner Indo-European languages: An advanced approach to Indo-European lexicostatistics', *Linguistics* 59(4). 949–979.

Kaufman, Stephen A., 1974. *The Akkadian Influences on Aramaic*. (Assyriological Studies 19). Chicago (IL): University of Chicago Press.

Kogan, Leonid, Manfred Krebernik, Oleg Linkohr, Rim Nurullin, Olga I. Sęk. (eds.), 2020. *Etymological Dictionary of Akkadian. Vol. 1. Roots Beginning with P and B*. Berlin & Boston (MA): De Gruyter. https://doi.org/10.1515/9781614512394

Kortlandt, Frederik Herman Henri, 1978. 'I.-E. Palatovelars before resonants in Balto-Slavic', in Jacek Fisiak (ed.), *Recent Developments in Historical Phonology*. (Trends in Linguistics. Studies and Monographs 4). The Hague, Paris and New York (NY): Mouton de Gruyter. https://doi.org/10.1515/9783110810929.237

Kortlandt, Frederik Herman Henri, 2016. 'Balto-Slavic and Indo-Iranian', *Baltistica* 51(2). 355–364.

Kortlandt, Frederik Herman Henri, 2019. 'Indo-European between Uralic and Caucasian', *Journal of Indo-European Studies* 47(3–4). 383–388.

Kretschmer, Paul, 1896. *Einleitung in die Geschichte der Griechischen Sprache*. Göttingen: Vandenhoeck & Ruprecht.

Kroonen, Guus, 2012. 'Non-Indo-European root nouns in Germanic: Evidence in support of the agricultural substrate hypothesis', in Riho Grünthal & Petri Kallio (eds.), *A Linguistic Map of Prehistoric Northern Europe*. (Suomalais-Ugrilaisen Seuran Toimituksia / Mémoires de La Société Finno-Ougrienne). Helsinki: Suomalais-Ugrilainen Seura. 239–260.

Lubotsky, Alexander M., 1988. *The System of Nominal Accentuation in Sanskrit and Proto-Indo-European*. (Memoirs of the Kern Institute). Leiden, New York (NY), København & Köln: Brill. https://doi.org/10.1163/9789004643987

Makkay, János, 1998. 'Greek ἀξίνε and πέλεκυ̃ς as semitic loan-words in Greek and the corresponding axe type', in Takahito Mikasa (ed.), *Essays on Ancient Anatolia in the Second Millennium B.C.* Wiesbaden: Harrassowitz. 183–197. (Bulletin of the Middle Eastern Culture Center in Japan 10).

Mallory, James Patrick & Douglas Quentin Adams (eds.), 1997. *Encyclopedia of Indo-European Culture*. London & Chicago (IL): Fitzroy Dearborn.

Malzahn, Melanie, 2016. 'The second one to branch off? The Tocharian lexicon revisited', in Bjarne Simmelkjaer Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander, & Birgit Anette Olsen (eds.), *Etymology and the European Lexicon: Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*. Wiesbaden: Reichert. 281–292.

Mann, Stuart E., 1984. *An Indo-European Comparative Dictionary*. Hamburg: Buske.

Martirosyan, Hrach, 2013. 'The place of Armenian in the Indo-European language family: The relationship with Greek and Indo-Iranian', *Journal of Language Relationship • Вопросы языкового родства* 10(1). 85–138.

Matasović, Ranko, 2012. 'Areal typology of proto-Indo-European: The case for Caucasian connections', *Transactions of the Philological Society* 110(2). 283–310.

Matzinger, Joachim, 2006. *Der altalbanische Text Mbsuame e Krështerë (Dottrina cristiana) des Lekë Matrënga von 1592: Eine Einführung in die albanische Sprachwissenschaft*. (Jenaer Indogermanistische Textbearbeitung 3). Dettelbach: J.H. Röll.

Mayrhofer, Manfred, 1986. *Etymologisches Wörterbuch des Altindoarischen*, Volume 1. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Carl Winter.

Mayrhofer, Manfred, 1996. *Etymologisches Wörterbuch des Altindoarischen*, Volume 2. (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Carl Winter.

Meyer, Gustav, 1891. *Etymologisches Wörterbuch der albanesischen Sprache*. Straßburg: Trübner. https://doi.org/10.1515/9783111562261

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.

Nikolayev, S. L. & Sergej Anatol'evič Starostin, 1994. *A North Caucasian Etymological Dictionary*. Moscow: Asterisk.

Olander, Thomas, 2018. 'Connecting the dots: The Indo-European family tree as a heuristic device', in David M. Goldstein, Stephanie W. Jamison, & Brent Vine (eds.), *Proceedings of the 29th Annual UCLA Indo-European Conference*. (Journal of Indo-European Studies Monograph Series) Bremen: Hempen. 181–202.
Olander, Thomas, 2022. 'To **b* or not to **b*: Proto-Indo-European **b* in a phylogenetic perspective', *Historische Sprachforschung* 133(1). 182–208.
O'Leary, De Lacy, 1923. *Comparative Grammar of the Semitic Languages*. (Trübner's Oriental Series). London & New York (NY): Trübner & Dutton.
Olsen, Birgit Anette & Rasmus Thorsø, 2022. 'Armenian', in Thomas Olander (ed.), *The Indo-European Language Family: A phylogenetic perspective*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108758666.012
Orel, Vladimir Ėmmanuilovič, 1998. *Albanian Etymological Dictionary*. Leiden: Brill.
Palmér, Axel I. 2024. *Indo-Slavic lexical isoglosses and the prehistoric dispersal of Indo-Iranian*. Leiden: Universiteit Leiden PhD Dissertation (Proefschrift). https://scholarlypublications.universiteitleiden.nl/handle/1887/3765823 (9 November, 2024).
Pellard, Thomas, Robin J. Ryder & Guillaume Jacques, forthcoming. 'The family tree model', in Ledgeway (ed.), *The Wiley Blackwell Companion to Diachronic Linguistics*. Wiley-Blackwell. https://www.academia.edu/101656989/The_Family_Tree_model.
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
Peyrot, Michaël, 2022. 'Tocharian', in Thomas Olander (ed.), *The Indo-European language family: A phylogentic perspective*. Cambridge: Cambridge University Press.
Pinault, Georges-Jean, 1998. 'Economic and administrative documents in Tocharian B from the Berezovsky and Petrovsky collections', *Manuscripta Orientalia* 4(4). 3–20.
Porzig, Walter, 1954. *Die Gliederung des indogermanischen Sprachgebiets*. (Indogermanische Bibliothek. Reihe 3. Untersuchungen). Heidelberg: Carl Winter.
Ringe, Donald A., 2022. 'What we can (and Can't) learn from computational cladistics', in Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*. Cambridge: Cambridge University Press. 52–62. https://doi.org/10.1017/9781108758666.004
Ringe, Donald A. & Joseph F. Eska, 2013. *Historical Linguistics: Toward a Twenty-First Century Reintegration*. Cambridge & New York (NY): Cambridge University Press. https://doi.org/10.1017/CBO9780511980183
Ringe, Donald A. & Ann A. Taylor, 2007a. The Indo-European wordlists. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/˜nakhleh/CPHL/ie-wordlist-07.pdf (1 June, 2022).
Ringe, Donald A. & Ann A. Taylor, 2007b. Morphological characters. Supplementary material of Ringe, Warnow & Taylor 2002. http://www.cs.rice.edu/˜nakhleh/CPHL/code-m-07.pdf (1 June, 2022).
Ringe, Donald A. & Ann A. Taylor, 2012. Cognations of lexical characters. Supplementary material of Ringe, Warnow & Taylor 2002. http://tandy.cs.illinois.edu/cognations-2k-revised.pdf (1 June, 2022).
Ringe, Donald A., Tandy Warnow & Ann A. Taylor, 2002. 'Indo-European and computational linguistics', *Transactions of the Philological Society* 100(1). 59–129.
Rougemont, Françoise, 2019. 'Animali e allevamento', in Maurizio Del Freo & Massimo Perna (eds.), *Manuale di Epigrafia micenica: Introduzione allo studio dei testi in lineare B*, Volume 2. Padova: Libreriauniversitaria.it. 305–347.
Scarborough, Matthew Joseph Charles, 2020. 'Cognacy and computational cladistics: Issues in determining lexical cognacy for Indo-European cladistic research', in Thomas Olander & Matilde Serangeli (eds.), *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European, 179–208. (Brill's Studies in Indo-European Languages and Linguistics)*. Leiden & Boston (MA): Brill. https://doi.org/10.1163/9789004416192_011
Schmidt, Klaus T., 1992. 'Archaismen des Tocharischen und ihre Bedeutung für Fragen der Rekonstruktion der Ausgliederung', in Robert Stephen Paul Beekes, Alexander Lubotsky, & Joseph Johannes Sicco Weitenberg (eds.), *Rekonstruktion und relative Chronologie: Akten der VIII. Fachtagung der Indogermanischen Gesellschaft, Leiden, 31. August - 4. September 1987*. (Innsbrucker Beiträge zur Sprachwissenschaft 65). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck. 101–114.
Starostin, Sergej Anatol'evič, 2004. 'The cultural vocabulary in the common north caucasian lexical stock', in Vitaly Naumkin (ed.), *Russian Oriental Studies: Current Research on Past and Present Asian and African Societies*. Boston (MA): Brill. 211–245. Translation of Starostin 1985.
Sergej Anatol'evič, Starostin, 2009. 'Indo-European-north Caucasian isoglosses', *Mother Tongue: Journal of the Association for the Study of Language in Prehistory* 14. 77–135. Translation by Ronald W. Thornton of Starostin 1988.
Sihler, Andrew Littleton, 1995. *New Comparative Grammar of Greek and Latin*. Oxford & New York (NY): Oxford University Press. https://doi.org/10.1093/oso/9780195083453.001.0001
Sokoloff, Michael & Carl Brockelmann, 2009. 'A Syriac lexicon: A translation from the Latin, correction, expansion, and update of C', in *Brockelmann's Lexicon Syriacum*. Winona Lake (IN) & Piscataway (NJ): Eisenbrauns & Gorgias Press.
Starostin, Sergej Anatol'evič, 1985. 'Kul'turnaja leksika v obščesevernokavkazskom slovarnom fonde', in E. K. Borisova (ed.), *Drevnjaja Anatolija*. Moskva: Nauka. 74–94.
Starostin, Sergej Anatol'evič, 1988. 'Indoevropejsko-severnokavkazskie izoglossy', in *Drevnij Vostok: Ėtnokul'turnye svjazi*. Moskva: Nauka. 112–163.

Starostin, Sergej Anatol'evič, 2007a. 'Kul'turnaja leksika v obščesevernokavkazskom slovarnom fonde', in Sergej Anatol'evič Starostin (ed.), *Trudy po jazykoznaniju*. Moskva: Jazyki Slavjanskix Kul'tur. 289–311. Reprint of Starostin 1985.

Starostin, Sergej Anatol'evič, 2007b. 'Indoevropejsko-severnokavkazskie izoglossy', in Sergej Anatol'evič Starostin (ed.), *Trudy po jazykoznaniju*. Moskva: Jazyki Slavjanskix Kul'tur. 312–358. Reprint of Starostin 1988.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor, 2010. 'Borrowability and the notion of basic vocabulary', *Diachronica* 27(2). 226–246.

Thorsø, Rasmus, 2020. 'Two Balkan Indo-European loanwords', in Matilde Serangeli & Thomas Olander (eds.), *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*. Leiden & Boston (MA): Brill. 251–262. https://doi.org/10.1163/9789004416192_014

Tischler, Johann, 1973. *Glottochronologie und Lexikostatistik*. (Innsbrucker Beiträge zur Sprachwissenschaft 11). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Topalli, Kolec, 2017. *Fjalor etimologjik i gjuhës shqipe*. Tiranë: Botimet Albanologjike.

Tremblay, Xavier, 2005. 'Iranian historical linguistics in the twentieth century', *Friends and Alumni of Indo-European Studies Bulletin* 11(1). 1–23.

Verhasselt, Gertjan, 2011. 'The pre-Greek linguistic substratum: A critical assessment of recent theories', *Les Études Classiques* 79. 257–283.

von Soden, Wolfram, 1972. *Akkadisches Handworterbuch*, Volume 2: M-S. Wiesbaden: Harrassowitz.

Warnow, Tandy, Don Ringe & Ann A. Taylor, 1996. 'Reconstructing the evolutionary history of natural languages', in Éva Tardos (ed.), *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms: Atlanta, Georgia, January 28–30, 1996*. New York (NY) & Philadelphia (PA): Society for Industrial and Applied Mathematics. 314–321.

Watson, Wilfred G. E., 2013. 'Indo-European and Semitic: Two-way traffic', in Juan Pedro Monferrer-Sala & Wilfred G. E. Watson (eds.), *Archaism and Innovation in the Semitic Languages: Selected Papers*. (Semitica Antiqva 1). Córdoba: Oriens Academic. 163–193.

Weiss, Michael L., 2020. *Outline of the Historical and Comparative Grammar of Latin* (2nd ed.). Ann Arbor (MI) & New York (NY): Beech Stave Press.

Weninger, Stefan (ed.), 2011. *The Semitic Languages: An International Handbook*. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 36). Berlin: de Gruyter. https://doi.org/10.1515/9783110251586

Winter, Werner, 1997. 'Lexical archaisms of the Tokharian languages', in Hans Henrich Hock (ed.), *Historical, Indo-European, and Lexicographical Studies: A Festschrift for Ladislav Zgusta on the Occasion of his 70th Birthday*. (Trends in Linguistics. Studies and Monographs 90). Berlin & New York (NY): Mouton de Gruyter. 183–193. https://doi.org/10.1515/9783110809718.183

Wüst, Walther, 1956. *Idg.* *pelek̑u- *"Axt, Beil": Eine paläo-linguistische Studie*. (Annales Academiae Scientiarum Fennicae. Series B 93.1). Helsinki: Suomalaisen Tiedeakatemian Toimituksia.

Addenda et corrigenda to S. Poulsen (2025) “Loanwords and Linguistic Phylogenetics: **pelek̂u-* ‘axe’ and **(H)a(i̯)ĝ-* ‘goat’”, *Transactions of the Philological Society* 123(1). 116-136.

(See also [https://zenodo.org/records/15075920 [zenodo.org]](https://zenodo.org/records/15075920))

1. Introduction

In the days following the initial publication of Poulsen (2025) I received valuable comments and extra information I wish I had known or realised earlier. These new insights on the difficulty of reconstructing PIr. **aǰa-* ‘goat’ and the reconstruction of “hissing-hushing” affricates in Caucasian languages call for some adjustments to be made to the conclusions on the analyses of the word for ‘goat’ in the original article. The overall conclusions of the paper remain the same.

2. Iranian **aǰa-* as a source of Tocharian **asə*

In an email on the date of the publication (January 27, 2025), Chams Benoît Bernard made me aware of Bernard 2024 (review of Carling & Pinault (2023)), in which he comments on the proposed Iranian origin of Proto-Tocharian **asə* ‘goat’, allegedly borrowed from Iranian **aǰa-*. Bernard (2024: 293) notes that the attestations of Iranian *az-* ‘goat’ are all “difficult to find, rare, or *hapax legomena*”, and that they are not attested in any Middle Iranian language which was in contact with Tocharian. Had the simplex **aǰa-* been borrowed into Tocharian, it would probably have been rendered as ***etse* (if from “Old Steppe Iranian” or ***āso* or the like if through Khotanese).

The attestations of **aǰa-* are certainly rare or difficult. Nevertheless, Bailey (1979: 6) has suggested that it survives in the Khotanese compound *aysdāṁ* ‘a commodity’ – if he is correct in tracing this back to PIr. **Haǰa-dānā-* < PIIr. **Haǰa-dʰaHnaH-* ‘goat + grain’ (Palmér 2024: 42). This is unlikely. Although PIr. **(H)aǰ-* would yield Khot. *az-* <ays>, the etymology is doubtful: There is nothing forcing a semantic connection of Khot. *aysdāṁ* and the unclear Tumshuqese *eźdana kalasta* ‘(leather?) vessel or bag for food (grain?)’ with the potentially non-existing PIr. **aǰ-* (Rastorgueva & Ėdel’man 2000: 293).

Although the Iranian reflexes of *(*H*)*aj́*- 'goat' are difficult, there is certainly basis for reconstructing **(H)ij́*- 'skin-, leather' for Proto-Iranian on the basis of YAv. *īzaena*-, Khot. *häysa*- 'skin, hide', Balochi *hīz* 'leather churn' and Yidgha *īze*, *yiǰya* 'goatskin-bag used for carrying sour milk' (Bernard 2023: 63). I remain sceptical towards the connection between the stem **(H)ij́*- denoting 'leather', 'skin', 'hide' and derived products and the zoonym **(H)ai̯ĝ*- found in Greek and Armenian because *(*H*)*ai̯ĝ*- is otherwise never attested in the zero grade outside this Iranian etymon.

That said, it is difficult to avoid the Iranian cognates of Ved. *ajína*- 'fur of a black antelope or tiger' (AV+), although there are few or no traces of the simplex **Haj́a*- 'goat'. YAv. *azina-uuaṇt*- 'wearing a hide' and Wakhi *yazn* 'inflated skin, mussuck (water bag made of leather)' certainly point to the existence of this derivative denoting 'hide' in Proto-Iranian (Palmér 2024: 44, esp. n. 49). These forms are relevant to the discussion of the hypothesis that the ancestor of Lith. *ožìnis* was borrowed from Iranian (*ALEW*[2]: 830). This loan can, as stated (Poulsen 2025: 12), only be considered valid if the etymon was preserved in the potential donor language. It seems likely that this derivative was preserved in Proto-Iranian. It remains peculiar that an Iranian loan would show Winter's law in Balto-Slavic.

In conclusion, while the simplex **ája*- 'goat' cannot be reconstructed for Proto-Iranian on the basis of certain attestations, **ajína*- 'leather, hide' which seems to be a derivative of it can.

3. Caucasian corrections

In an email (January 30, 2025), Samopriya Basu pointed me to valuable publications and provided me with an excellent commentary on the Caucasian reconstructions. The Caucasian reconstructions could certainly have benefitted from more context. As briefly noted, Nikolayev & Starostin (1994) advocate for the existence of North-Caucasian, the proposed "macro-family" entailing Northeast Caucasian (Nakh-Daghestani) and Northwest Caucasian (West Caucasian or Abkhaz-Adyghe). It remains highly disputed. However, not only the theory, but the entire dictionary by Nikolayev & Starostin have been heavily criticised (e.g. Schulze 1997; Nichols 2003; but see also Bengtson & Leschber 2022). In Figure 1 below, I have given the main branches of the proposed "macro-family" for readers who – like myself – are not specialists in Caucasian linguistic history.

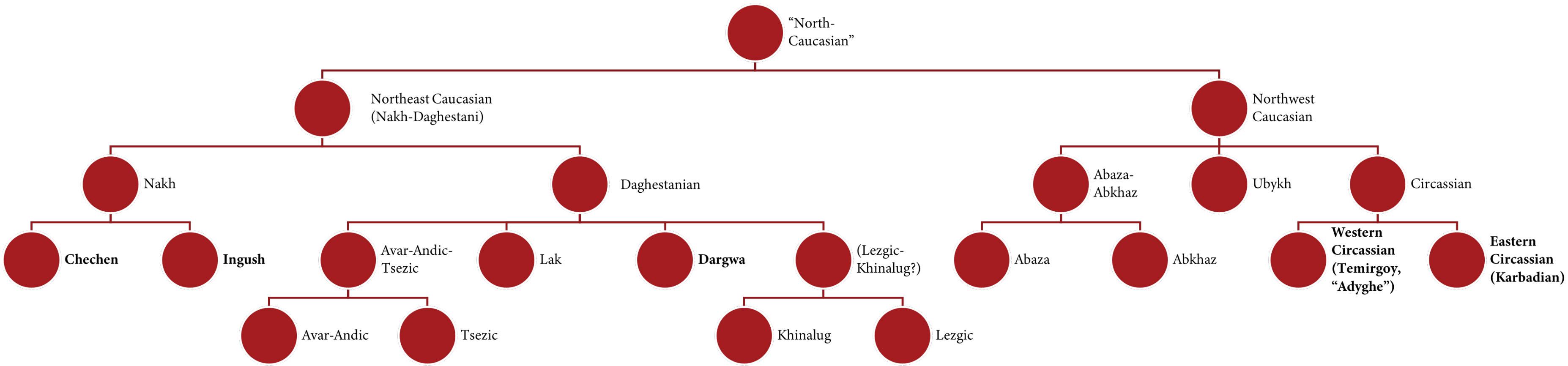


Figure 1. Family tree connecting the Northeast and Northwest Caucasian branches with their major descendants (Nichols 2003: 207; Ganenkov & Maisak 2021: 86; Arkadiev & Lander 2021: 368). The branches and languages from which there is evidence for the discussed cognate are marked bold.

3.1. North and Northeast Caucasian

The reconstruction **ʡējʒ́wē* provided for Proto-North-Caucasian by Nikolayev & Starostin (1994: 245) is problematic beyond the fact that the protolanguage possibly never existed at all, because it rests on problematic assumptions. PNC **ʡējʒ́wē* is defended by the following comment, but each point should be criticized:

> Medial *-*j*- is postulated to account for the development *-*ʒ́*- > *-*st*- in PN, while *-*w*- is reflected as *-*b*- in PN and as labialisation in PWC. (Nikolayev & Starostin 1994: 245)

First, no Nakh-Daghestani languages "preserve" the "hissing-hushing (= palatalized)" affricates, making the PND form **Hējʒ́V* problematic from an internally Nakh-Daghestani point of view. Basu further informs me that the peculiar "hissing-hushing" sibilants and affricates are, in fact, not palatalised in the West Caucasian languages where they belong. They are rather produced with a distinct tongue-shape whereby the apex rests against the alveoles of the lower teeth (like [z̠]), and the sides of the tongue are pressed against the upper molars (like [ʒ]), giving a lamino-postalveolar fricative with no sublingual space ([ẑ]) (Ladefoged & Maddieson 1996: 191; Catford 1977: 290). This makes palatalisation impossible. They may, however, be slightly pharyngealised (Beguš 2021: 704).

Second, PNC **ʒ́* does not even yield "hissing-hushing" affricates anywhere in the table of sound correspondences provided by the authors, and PNC (as well as PND) **ʒ́w* yield PN **ʒ*, not **st* as in this cognate (Nikolayev & Starostin 1994: 50). Third, it is left unsaid why the medial **j* is relevant for the proposed change of PND **ʒ́w* > PN **st*. Fourth, the labial **w* is reconstructed on the basis of the **b* in Proto-Nakh **ˀābst* (Chechen *awst*, Ingush *ɔsti*), but as we have just established, PND **ʒ́w* is said to yield PN **ʒ* in one account, **st* in another, but somehow, it also yields **bst* in this reconstruction. Notably, the intermediate form, PND **Hējʒ́V*, is reconstructed without the labial, which means it would have to reappear in Proto-Nakh by Nikolayev & Starostin's own account.

To make matters worse, Schulze (2014: 260) argues that the entire *Wortsippe* of Dargwa *ˁeža* (Itsari *eč̣ːa*, Kubachi *ič̣ːa*, Kaytak and Muiri *ˁič̣ːa*) is borrowed from a Turkic form **ička*. Azeri and Kipchak have *keči*, but Kumyk *ečki* and Nogai *eški* show the reverse order of the medial consonants which would be the order in the proposed donor language. However, further cognates than the ones given by Schulze in the Dargwa branch (e.g. Akusha *ʕeža, ʕežn*-; Urakhi *ʕid͡za*; Chirag *ʕàˁč̣ːa*) allow for a reconstruction of a Proto-Dargwa form **ʕed͡ʒȍm(nə)* (Mudrak 2016: 234–5),[1] which means that the Turkic loan would have to be suspiciously early and before

[1] **'eǯȍm(nə)* ~ **'ĕǯȍm(nə)* in the original spelling. According to Basu, the medial affricate is probably geminate **d͡zː*, but neither palatalised, labialised nor "hissing-hushing".

the break-up of the Dargwa branch. If it is a loan, the Daghestani comparanda are irrelevant for the reconstruction of Proto-Nakh-Daghestani.

It should further be noted that the division of the Nakh-Daghestani family into the two major branches, Nakh and Daghestanian, which is usually assumed (Nichols 2003: 207), is debated (Ganenkov & Maisak 2021: 86; Schulze 2017: 108).

3.2. Northwest Caucasian

In the comment above, Nikolayev & Starostin further claim that the **w* reconstructed for "PNC" is preserved as labialisation in Proto-West-Caucasian **ač̣́ːʷá*. However, the cognates quoted for this etymon do not show labialization: "Adyghe" *āča* and Kabardian *āža*. Basu informs me that "Adyghe" is more correctly referred to as the Temirgoy (literary West Circassian), and that the form *āča* should be rendered *ачъэ* /äˑt͡ʂʰɜ/. Similarly, Kabardian (literary East Circassian) *āža* should be *ажэ* /äˑʐ̻ɜ/. The Proto-Circassian form reconstructed from these cognates would look something like */äˑt͡ʂɜ/ (or **āčːa* in more traditional notation) – with no trace of labialization or palatalization, nor of "hissing-hushing" affricates.

If we really wanted to see a connection with Indo-European, this form *could* resemble Proto-Indo-Iranian **Hajá-* slightly, but it would probably be significantly younger than Proto-Indo-Iranian! The lack of comparanda in Ubykh and Abkhaz-Abaza, the two remaining branches making up Northwest Caucasian (Arkadiev & Lander 2021: 368) – make it even more difficult to project the form further back in time (Šagirov 1977: 58).

In conclusion, a lot more work has to be done on the reconstruction of the internal nodes in Northeast and Northwest Caucasian before any credible connection with Indo-European can be made.

4. Conclusion

The points presented above bring some emendations to the material and conclusions of Poulsen (2025):

- PT **asə* was probably not borrowed from Iranian **aja-* – but the preform still entered the languages before the breakup of Proto-Tocharian.
- It is doubtful at best if PIr. **aja-* 'goat' ever existed as there are only spurious traces of it in hapax legomena and philologically difficult attestations. However, the derivative **Hajina-* '(of) hide' must be reconstructed for Proto-Iranian. It is still questionable if this form was borrowed into Proto-Balto-Slavic.
- Even if the existence of North-Caucasian should be proven, the reconstruction **ʔēǯ́wē* is not compatible with the intermediate reconstructions based on the material.

- It is problematic to reconstruct a common form for Proto-Nakh-Daghestani: Proto-Nakh **ābst-* is hardly compatible with Proto-Dargwa **ʕed͡ʒzö̀m(nə)*, which in turn might be a loan from a Turkic language.
- A Proto-West-Caucasian form is uncertain because of the lack of comparanda from Abkhaz-Abaza, but Proto-Circassian points to */äˑt͡ʂ̰ɜ/ (**āčːa*). Although this form superficially looks like Proto-Indo-Iranian **Hajá-*, it is thousands of years younger.

5. Bibliography

ALEW[2] = Hock, Wolfgang, Elvira-Julia Bukevičiūtė, Christiane Schiller, et al. (eds.). 2020. *Altlitauisches etymologisches Wörterbuch*. 2.0 [Online]. https://alew.hu-berlin.de/.

Arkadiev, Peter & Yury Lander. 2021. The Northwest Caucasian Languages. In Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*, 368–446. (Oxford Handbooks). Oxford University Press. doi:10.1093/oxfordhb/9780190690694.013.3. https://doi.org/10.1093/oxfordhb/9780190690694.013.3 (31 January, 2025).

Bailey, Harold Walter. 1979. *Dictionary of Khotan Saka*. Cambridge & New York (NY): Cambridge University Press.

Beguš, Gašper. 2021. Segmental Phonetics and Phonology in Caucasian languages. In Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*, 688–728. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190690694.013.18 (24 March, 2025).

Bengtson, John D. & Corinna Leschber. 2022. On Criticism of S. L. Nikolayev/S. A. Starostin, A North Caucasian Etymological Dictionary. *Iran and the Caucasus* 26(4). 392–406. doi:10.1163/1573384X-20220405.

Bernard, Chams Benoît. 2023. Like dust on the Silk Road: An investigation of the earliest Iranian loanwords and of possible BMAC borrowings in Tocharian. Leiden: Universiteit Leiden PhD-dissertation (Proefschrift).

Bernard, Chams Benoît. 2024. Review of Carling & Pinault (2024). *Wékʷos: Revue d'études indo-européennes* 7. 285–302.

Carling, Gerd & Georges-Jean Pinault. 2023. *Dictionary and thesaurus of Tocharian A*. Wiesbaden: Harrassowitz.

Catford, John Cunnison. 1977. Mountain of Tongues: The Languages of the Caucasus. *Annual Review of Anthropology*. 283–314.

Ganenkov, Dmitry & Timur Maisak. 2021. Nakh-Dagestanian Languages. In Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*, 86–145. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190690694.013.4 (31 January, 2025).

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Mudrak, Oleg Alekseevič. 2016. *Darginskie osnovy*. Vol. 1. Ėtimologičeskij slovar'. Moskva: Jazyki Narodov Mira.

Nichols, Johanna. 2003. The Nakh-Daghestanian Consonant Correspondences. In Dee Ann Holisky & Kevin Tuite (eds.), *Current Trends in Caucasian, East European and Inner

Asian Linguistics: Papers in honor of Howard I. Aronson, 207–264. (Current Issues in Linguistic Theory 246). John Benjamins. doi:10.1075/cilt.246.14nic. https://benjamins.com/catalog/cilt.246.14nic (30 January, 2025).

Nikolayev, S. L. & Sergej Anatol'evič Starostin. 1994. *A North Caucasian etymological dictionary*. Moscow: Asterisk.

Palmér, Axel I. 2024. Indo-Slavic lexical isoglosses and the prehistoric dispersal of Indo-Iranian. Leiden: Universiteit Leiden PhD dissertation (Proefschrift). https://scholarlypublications.universiteitleiden.nl/handle/1887/3765823 (3 January, 2025).

Poulsen, Simon. 2025. Loanwords and Linguistic Phylogenetics: **pelek̑u*- 'axe' and **(H)a(i̯)ĝ*- 'goat.' *Transactions of the Philological Society* 123(1). 116–136. doi:10.1111/1467-968X.12308.

Rastorgueva, Vera Sergeevna & Džoj Iosifovna Ėdel'man. 2000. *Ėtimologičeskij slovar' iranskix jazykov*. Vol. 1. a–ā. Moskva: Vostočnaja literatura.

Šagirov, Amin Kabcuevič. 1977. *Ėtimologičeskij slovar' adygskix (čerkesskix) jazykov*. Vol. 1. A-N. Moskva: Nauka.

Schulze, Wolfgang. 1997. Review of Nikolayev & Starostin (1994). *Diachronica* 14(1). John Benjamins. 149–161. doi:10.1075/dia.14.1.11sch.

Schulze, Wolfgang. 2014. Generic Terms for Domestic Animals in East Caucasian. *Iran and the Caucasus* 18(3). Brill. 213–274. doi:10.1163/1573384X-20140303.

Schulze, Wolfgang. 2017. The comparative method in Caucasian linguistics. In Jared Klein, Matthias Fritz & Brian Joseph (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 105–114. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 41). Berlin & Boston (MA): De Gruyter. doi:10.1515/9783110261288-011.

Article 3: Korn & Poulsen: A Tree or Not? An East Iranian Experiment

Preamble of “A Tree or Not: An East Iranian Experiment”

1. The place of East Iranian in a thesis on “Indo-Greek”

It would not be unreasonable to question the relevance of a paper on the internal relations of the East Iranian in a thesis investigating the relationship between Indo-Iranian and Greek. One way to argue would be to insist that a higher-order group can never be stronger than the nodes on which it rests, and that all intermediate proto-languages should be reconstructed bottom-up. Thus, the validity of an East Iranian clade has implications for the reconstruction of Proto-Iranian which again is important for the reconstruction of Proto-Indo-Iranian – and finally for any potential “Indo-Greek” clade and proto-language. In order for this argument to hold any water, it would require the same effort for all other potential descendants of the latest shared ancestor of Indo-Iranian and Greek.

There would be considerable implications of such a line of argumentation. Problematically, even the existence of complete unity and the reconstructability of Proto-Greek (Garrett 2006) and Proto-Iranian (Tremblay 2005a)[1] have been questioned, and should thus be addressed first. Similarly, since the questions of how Macedonian (Brixhe 2018: 1864–6) and Phrygian (Obrador-Cursach 2019) are related to Greek, as well as the questions of the relationship between the Dardic and Indic languages (Ėdel’man 1999; Hegedűs 2005; Kogan 2005: 201) and the placement of Nuristani within Indo-Iranian (Strand 2023; Kümmel 2022: 253–6; Smith 2017: 443–4; Blažek & Hegedűs 2012; Degener 2002; Budruss 1977; 1978) remain unsettled, this argument would lead to substantial work beyond what is possible in this thesis in order to be credible. While it is necessary to have a clear and structured view of the phylogenies and subdivisions of these branches when reconstructing linguistic features of their proto-language or latest common ancestor, it would indeed be impossible for this thesis to cover such detail of all Indic, Dardic, Nuristani, Iranian, Hellenic, Phrygian, Macedonian languages and dialects. Moreover, it has been made clear ad nauseam in this thesis that we cannot base phylogenetic claims on positive evidence alone. Accordingly, in a study of higher-

[1] Or its difference from Proto-Indo-Iranian (Kümmel, Abstract for *Fachtagung der Indogermanischen Gesellschaft*, September 2024).

order subgrouping, this would even have to extend beyond just the two branches in question but also be relevant for all potential sister-branches. Usual candidates are Armenian, Albanian and Balto-Slavic, but across the literature, it is possible to find claims of all other Indo-European branches also potentially descending from the latest shared ancestor of Indo-Iranian and (cf. Article 1, sect. 2). Undertaking this task for "Indo-Greek" would essentially mean reconfiguring the Indo-European family tree – which is on the to-do list, but sadly not achievable in this thesis.

A much more honest and realistic argument is the sheer fact that East Iranian offers a very good case on which to apply the methodology of linguistic phylogenetics, more specifically of computational linguistic phylogenetics, and in practise to test a Maximum Parsimony analysis. The topic of this thesis is not just the relation between the Indo-Iranian and Greek clades, it is also very deeply focused on the methodology of linguistic phylogenetics and specifically higher-order subgrouping. The "Indo-Greek" clade, whatever its nature and composition of branches was, is mostly researched in a dialect-geographical framework (cf. Article 1, sect. 1), which means that a lot of the arguments in favour of a special relationship revolve around evidence that is, for a phylogeneticist, rather hard to engage with, prove or falsify. Isoglosses in dialectology tend to be more fluid, as they also cover shared tendencies and similar responses to the same linguistic pressure. What is described as innovations need not translate directly to the phylogenetic definition of a shared innovation.

For instance, the fact that Indo-Iranian and Greek innovated a pluperfect and a perfect middle with roughly the same morphemes might very well constitute a shared innovation in some sense – but neither can be a candidate of a significant shared innovation in phylogenetics, since they must be reconstructible in form and function for a common prestage. As we cannot reconstruct shared "Proto-Indo-Greek" formations of the actually occurring forms of the pluperfect or perfect middle in Indo-Iranian and Greek, it is almost impossible for the innovations to have occurred only once when the speakers of the languages still formed a single speech community. I do not mean to imply that there is no methodology in dialectology, but I will stress that conclusions projecting such shared or similar "Indo-Greek" features back to a hypothetical prolonged period of contact while the dialects were diversifying (i.e. the conclusions of Euler (1979) and Birwé (1956)) cannot be regarded as final. Such hypotheses would stand stronger if they were backed by circumstantial evidence from aDNA or archaeology – and indeed if they had been tested properly with phylogenetic methods.

This is where East Iranian proves its worth as an experiment within reach. For two reasons, it is an ideal linguistic area on which to apply phylogenetic methodology in search of novel results: First, the languages no doubt formed a dialect continuum at some point in their development – second, the directionality of changes and the ancestral system is known. These two points above obviously ignore the crucial fact that East Iranian is valuable in itself. It is an area that has not received as much scholarly attention as many other Indo-European branches. The relatively frequent discovery of new documents attesting various stages of the middle Iranian languages as well as field work on the modern varieties has enlarged the data volume and continuously paves the way for novel analyses in the etymology of these languages, and these indeed deserve serious attention in phylogenetics.

2. Why phylogenetics rather than dialectology

The East Iranian languages have, like the relationship between Indo-Iranian and Greek, often been viewed as a dialect-continuum (Sims-Williams 1996: 651; Korn 2016; 2019). As discussed in Chapter 2, many of the claims advanced in prehistoric dialectology are difficult to falsify, and there is no inherent incompatibility between languages having developed as bifurcating splits or in a wave-like fashion. Therefore, it is preferable for the scholarly debate too keep some close ties to falsifiable – albeit simplistic – hypotheses since these drive research forward just as much – if not more – than wild ideas and guestimates.

Unfortunately, it is a lot harder to disprove areal relations in prehistory than to propose them. It is often argued, for instance, that the ancestors of Khotanese and Wakhi belonged to the same dialect group in prehistory (Skjærvø 1989: 375) – or Wakhi is simply listed as a member of a the Sakan dialect group (Skjærvø 2017: 476; Huyse 2017: 601; Tremblay 2005a: 678; 2005b: 423). The basis for this is one single phonemic isogloss, namely the retention of relics of the Proto-Iranian palatal character of **ć* in **aću̯a-* 'horse' > Khot. *aśśa-* 'id.', Wakhi *yaš* 'id.' as opposed to the other Iranian dialects in which this trait is lost: **aću̯a-* > **acu̯a-* > Av. *aspa-*, OP *asa-* and Oss. *æfsæ* 'mare' (*-*ā*) (Peyrot 2018: 271–2). Now, this isogloss occurs only in one Wakhi word, and it is not backed by any other shared phonological, morphological or lexical isoglosses (Cheung 2015: 55–6, n. 32). Additionally, the isogloss is best described as a common retention (Emmerick 1989: 216).

Consequently, claiming that "Wakhi is an example of the repeated invasions of Saka since antiquity" is indeed a conclusion that is very far from what can be safely extrapolated from the data (*pace* Windfuhr 2009: 15). Most importantly, areal relations do not exclude the possibility of an underlying genetic relationship subsequently muddled by areal innovations.

The fact that the languages innovated horizontally together, i.e. developed common tendencies, reacted to the same inherited pressures or were in contact with the same languages in the same geographical area, does not mean that they cannot simultaneously have developed in a tree-like fashion. This tree could potentially be uncoverable because of later or more prominent innovations, or it could indeed be hiding right at our feet, only muddled by the preconceived idea that we are dealing with a linguistic areal and therefore need not even try to distinguish between isoglosses as systemic tendencies and potentially shared innovations that go back to the same reconstructible preform.

Little reconstruction of the potential "higher-order" branches between Proto-(East-)Iranian and the attested languages has been done. This is not to claim that researchers who describe linguistic innovations in terms of areal features ignore reconstruction, but remaining agnostic of the relative age of innovations can indeed be a draw-back of a solely wave-model approach to relationship. Indeed, a phylogenetic approach might help qualify the reconstruction of intermediate steps in the development of the East Iranian branch. In other words, unresolved family trees should spike caution and require some action before any reconstruction is taken at face value (Olander 2018; Goldstein 2022; Jacques & List 2019); and such an action can be phylogenetic modelling.

Thus, the East Iranian "dialect continuum" is an ideal candidate for a phylogenetic investigation; its internal diversification remains untested with phylogenetic methodology, and it simultaneously serves a good test case to examine the influence of known branch-internal contact on phylogenetic results. It is a further advantage that the ancestral state is somewhat known: The East Iranian languages can largely be derived from the Proto-Iranian reconstructible from the Old Iranian and Middle West Iranian languages which. Therefore, we have a reasonable proxy for the ancestral state and can rely on the directionality of innovations.

3. Why computational methodology rather than traditional methods

Testing the hypothesis that the East Iranian languages do not only form a dialect continuum but that there is indeed a recoverable phylogenetic tree underneath their subsequent horizontal influence on each other is where we see the greatest potential for novel achievements. However, the exploration of the middle and modern Iranian languages cannot be separated from immense philological work, and thus the analyses tend to focus on the isoglosses in isolation. Comparing languages on the basis of what springs to mind is a dangerous sport when it comes to languages as poorly attested as the Middle Iranian ones and as late or innovative as many of the Modern East Iranian ones. What coincidentally survives is, of course, all the evidence we can base our grouping on, but such material calls for a systematic rethinking of what surely is and what merely could be shared innovations. Not preserving the evidence in favour of having been part of a dialect continuum is not the same as preserving counterevidence. Computational methodologies offer great possibilities for over-arching analyses encompassing as much evidence as possible at once; and though it might seem banal, the data preparation will lead to analyses forcing a sharp distinction between clear positive evidence for subgroupings, and arguments that are merely compatible with the favoured solution. Importantly, the computational analyses make positive counterevidence very clear.

Further, etymological analyses will indeed benefit from a strict methodological approach to reconstruction. Since we do believe that linguistic prehistory can to some extent be recovered and reconstructed protolanguages possess some elements of linguistic reality, they also need to be reconstructed through a phylogenetic lens.

Although linguistic reconstruction is more complex than the "numbers' game" that is a computational phylogenetic analysis, viewing the traditional assumptions through a quantitative lens is a useful tool. In some cases, the basis for and perhaps even the motivation for an innovation was already present in the proto-language (Garrett 1999), but in other cases the innovation cannot necessarily be projected all the way back to the parent language. The former is especially clear in cases of functional and formal mergers – and perhaps for phonetic changes. In the case of the East Iranian languages, this goes for the initial lenition of stops and for the reduction of the case system. The latter becomes especially prominent when a formant or category does not have reflexes in all the languages examined. In the language of computational analysis, the transition must have happened between the root state and the leaves – in traditional terms, the innovation occurred at some point after the dissolution of the proto-language. In the case of the East Iranian languages, this clearly goes for the nominal plurals in *-*t*-. They go back to the old collective suffix *-*tā*, still function morphosyntactically as such in the Sogdian Ancient letters where they are semantically plural (Sims-Williams 2021), but they have become fully fledged regular plurals in later Sogdian, Yaghnobi and Ossetic (Kim 2025). If these languages constitute a subgroup, we can easily reconstruct the innovation once for their common ancestor. But when Sogdian also shares innovations of the same sort with the Pamir languages, especially Shughni and Wakhi, with Chorasmian and Bactrian and with Khotanese, a crucial question should be raised: What is the consequence of stating that languages share these innovations. If they are truly genetic innovations, this criss-cross of isoglosses must mean that the innovations arose early on and were lost multiple times independently. A computational approach makes it easy and transparent to examine such scenarios to optimise the traditional reconstruction of linguistic prehistory.

As linguistic phylogenetics most definitely rests on these very same qualitative arguments, there is no obvious opposition between the philological approach and computational phylogenetics. Whether the data of choice is lexical cognates or grammatical features, the analysis of archaism and innovation is in its essence qualitative and can only be carried out by skilled linguists.

4. Why Maximum Parsimony rather than other computational methods

As elaborated in the paper the reason for choosing Maximum Parsimony over other computational methods is that is completely transparent and is easily compatible with and complements traditional methods. We are not aiming to rediscover the spoked wheel; we are trying to disentangle a complex linguistic area and to get a better grasp of the linguistic prehistory. What we need is a tool that will mimic what linguistics will do anyway and to allow us to face the consequences of competing analyses.

Maximum Parsimony searches for the easiest way to get from a root to the leaves with the cumulative costs of transitions as the optimisation criterion. This is ideal because it allows us to use the type of data we – and the rest of the linguistic community – deem most appropriate for research in linguistic phylogenetics, and it plays emphasis on the linguistic innovation, not just difference on the surface or similarity. This optimisation criterion is remarkably similar what would be done by hand by any historical linguist: It is – all other things equal – more economic to assume that an innovation happened only once in the prehistory of two languages that share the same innovation that to posit that these two languages underwent identical innovations in parallel. The method becomes appropriate when the traditional methods come to a halt: In the case of the East Iranian languages, it is impossible to draw a tree that is compatible with the inconsistently overlapping isoglosses. For instance, most of the isoglosses shared between any languages of the group are shared with Sogdian. This is compatible with multiple scenarios. Perhaps all such traits were inherited from an earlier stage and coincidentally lost everywhere but in Sogdian and whomever shares the isogloss. Perhaps they were spread horizontally after the dissolution of the common ancestor and are thus not compatible with a tree model. Or perhaps it is possible to combine these scenarios: An innovation can be genetic in one part of the family and horizontal in other parts. This is what we wish to explore by analysing all isoglosses traditionally held to be significant for the subgrouping of the East Iranian languages at once.

4.1.Why not another computational method

Other computational methods are less adequate. Distance-based methods, known from the early glottochronological studies have long been recognised as being too simplistic for linguistic analysis (Pellard, Ryder & Jacques forthc.: 16). The biggest problem is that they do not distinguish between archaisms and innovations. For East Iranian, there is no need for a study on surface similarities.

Maximum Compatibility, the method applied by the influential studies of Ringe et al. (2002), is less adequate for the East Iranian languages, because it finds the best result on the trees with which most characters are compatible. This is ideal if one has a large number of characters available and if most of these characters are expected to be somewhat aligned (Ringe, Warnow & Taylor 2002: 76, n. 11). But the reality of the East Iranian languages where the isoglosses are not confined to clear patterns make it problematic. There is no doubt that a Maximum Compatibility search would find almost all of the East Iranian isoglosses incompatible with the best tree(s), and the final result would in turn only be based on a few isoglosses that, maybe coincidentally, maybe because of shared linguistic history, are in agreement. For such a search to be viable for East Iranian, it would require a much larger dataset than what we are able to compile at present. We would not be able to use this method to advance the discussion much; the core of the issue is that the isoglosses are not in agreement. Different researchers emphasise different isoglosses which has led to the diverging proposals.

4.2.Why not a Bayesian analysis

Currently, Bayesian analysis is in fashion as the most statistically reliable phylogenetic method (Pellard, Ryder & Jacques forthc.: 17; 2022: 59). While this may very well be the case, it has several drawbacks for linguistic phylogenetic analysis of closely related languages.

First, it requires large quantities of data (hundreds of characters) to function. This is difficult to meet with phonological data. Phonemic systems are, after all, of variable but ultimately limited size. Second, all data must be of very similar if not equal significance. This is similar to the criterion of "identifiability" that also adherents of traditional methodology advocate for: "the linguistic element adduced as a shared innovation in the lower node should be clearly identifiable in the higher as well as in the lower node" (Peyrot 2022: 90).[2] The Bayesian approach practically excludes morphological isoglosses as they need not be preserved or ever have existed in the analysed languages (Greenhill & Gray 2012; Greenhill, Heggarty & Gray 2021). Thus, the preferred data type is lexical cognacy: All languages will have lexical material – and all related languages should, although this is hotly debated, be able to fill the same universal semantic slots with cognates of the same etyma.

However, lexical data – not matter how strictly defined or how much extra linguistic information goes into the coding of cognates (in the sense that cognates can only be determined on the basis of established sound laws), is not particularly adequate for determining linguistic relationships in general (Clackson 1994: 23–25; 2022), and not for languages usually described as a dialect continuum in particular. If we consider for a moment how new cognates arise (cf. Poulsen 2025: 121–2) and apply this line of thought to the East Iranian languages, detectable differences in cognates become even less compelling. It is beyond all doubt that the East Iranian languages descend from the same common ancestor, whether this is "Proto-East-Iranian" or

[2] From a formal point of view, not all characters in the East Iranian dataset are "identifiable": Not all languages show signs of the plurals in *-*t*-, and it quickly becomes absurd to evaluate the lack of evidence quantitatively. However, our characters are "identifiable" form a functional point of view: All languages either share the category (e.g. "plural"), or it is above doubt that they preserved the etymological material on which other innovations rest (e.g. the PIr. demonstrative stems or adpositions) at some point in their prehistory.

simply "Proto-Iranian". These languages share lexical isoglosses that set them apart from the other branches of Iranian (Sims-Williams 1996: 651; 1989: 169). The most prominent of these are used as examples in the following and will be treated in more detail in the Excursus below.

Differences in cognates can then arise by various lexical innovations – borrowings being the most prominent, and sadly, most irrelevant for establishing genetic relations. Onomatopoeia, spontaneous *ex nihilo* formations and various neologisms rarely enter the basic vocabulary and never in bulk. That said, there are unique lexical roots only shared among the East Iranian languages with unclear etymologies, e.g. **kapā-* 'fish', **pati-gāz-* 'receive, accept', possibly **sāna-* 'enemy' and **drau̯a-* 'hair'.

This leaves two likely ways for closely related languages to exhibit different etyma in the same semantic slot. First are semantic shifts of already existing material which are rarely significant. Most such semantic shifts would be between near-synonyms, and it would be impossible to distinguish between inherited and independent shifts for closely related languages (cf. 'hair' below). Additionally, borrowings and calques between neighbouring or still diverging languages would make it impossible to distinguish contact from true genetic innovation. It is, of course, possible for a semantic shift to be severely typologically unlikely and therefore highly unlikely to be independently repeatable (e.g. **mai̯θa-* 'vacillating' > 'day'?, **-suxta-* 'burned' > 'pure'?). Shifts like these could, perhaps, constitute an argument for linguistic subgrouping – but it would be most likely be excessively rare. Such an argument, no matter how strong, would drown in the bulk of irrelevant lexical evolutions. The final type lies in derivation. Once again, given that the languages are closely related and – at least in the beginning of their prehistory after the dissolution of the ancestral language, but for most languages also well into the period of their attestation – inherited the same patterns of word formation, an innovation within the system of productive morphology is not very convincing (cf. **abi-Har-* 'to find').

Third, just like distance-based methods, Bayesian methods do not distinguish shared archaisms from shared innovations, but infer the original state from the analysis of the distribution of the states attested at the level of the leaves (Pereltsvaig & Lewis 2015: 64ff.). Although proponents of the method deem this a quality, not a flaw of the method (Greenhill & Gray 2012: 525–6), we remain sceptical and prefer not to leave the fate of the results in the hands of the model. For the East Iranian languages, knowing the direction of the innovation is the strongest suit, and it would frankly be a shame not to include in the analysis.

Fourth, the results of Bayesian analysis are very opaque and very far from the linguistic input. The goal of a Bayesian analysis is not to return a single tree, but rather to calculate the likelihood of all possible trees. In lieu of a single tree, all trees of the analysis can be visualised on top of each other (as a *DensiTree*) – while this tool is useful and makes the results more intuitive, the analysis does not come closer to answering the questions of subgrouping with more than competing scenarios and numerical likelihoods. We don't mean to downplay the statistical strength of the methods, but at present, we deem it more fruitful to examine the grammatical isoglosses themselves.

4.3. Maximum Parsimony ideally addresses the task at hand

Instead, Maximum parsimony allows us to build a dataset around the most salient linguistic materials.

Since the languages all ultimately go back to Proto-Iranian, we can make use of Proto-Iranian reconstructions as a proxy for a putative "Proto-East Iranian". This is in stark contrast to the more complex question of higher-order subgrouping of the Indo-European languages where the existence of grammatical categories and the reconstruction of entire morphological systems – especially in the verbal domain – are up for debate or reinterpretation.

Basing our root state on Proto-Iranian as a Proxy for "Proto-East Iranian" is in fact not far off, as it is what tends to happen when experts try to reconstruct the latter (Sims-Williams 1989: 165). It would be a shame not to utilise this known directionality when it is in fact one of the strongest assets of the East Iranian languages for phylogenetic purposes; and Maximum Parsimony allows us to base the analysis on linguistic innovations and gives us tools to assign different weights and costs to different innovations, much like judging significance in traditional methodology.

The issue of chronology and reconstruction as well as the issue of contact-induced lexical changes are the reasons for choosing maximum parsimony over Bayesian analyses. The data would probably not be very reliable for a trustworthy Bayesian analysis of the topology in the first place, and thus – even if we accepted its reliability for a brief moment – the absolute chronology inferred by a Bayesian analysis would not be even close to trustworthy. As linguists, we value the relative chronology of changes much more than absolute chronology. We are much more interested in the relationships, development, branching and reconstruction of intermediate protolanguages (recovering the linguistic prehistory through linguistics) than we are with making postulates about migrations on the basis of grammar.

Reconstruction and etymological analysis are intimately tied to phylogenetics; especially when the languages of the family or branch are not contemporary or equally well-attested. This does not imply that older languages are always more conservative, or conversely, that modern languages are excessively innovative. However, we need to ask the question: Did an ancestor of this language ever possess the feature we observe, and can we thus reconstruct an unbroken succession of it? Or may a given feature be better accounted for by another explanation? This analysis will help focus and improve future research.

Maximum parsimony cannot prove that isoglosses that do fall neatly onto a tree structure developed this way. The method is only a tool for analysing the consequences of many linguistic argumentations simultaneously and to help us as linguistics determine which analysis of a given isogloss or conglomerate of these is more economic or likely. Not everything that could in principle be reconstructed for a common prestage came into being as a single innovation. Independent and parallel innovations are very frequent, but one way to sift through them is to view them in light of a phylogenetic framework. If even the insignificant potential innovations run along the same lines, it gives us an opportunity to sharpen our hypotheses with the more economical options.

5. Excursus: East Iranian Lexical Isoglosses

For East Iranian, there are numerous lexical isoglosses setting East Iranian apart from West Iranian and Avestan, but usually the same few examples are given. Unless otherwise stated, the list below relies heavily on (Sims-Williams 1996: 651; 1989: 169).[3] These isoglosses fall into different and relevant categories:

A common archaism against West Iranian innovation is *_gari-_ 'mountain' (PIIr. *_gr̥H-i-_ < PIE *_gʷr̥Hí-_, cf. Ved. _girí-_) which is preserved in Khot. _ggara-_ and Sogd.

[3] For 'fish', 'day' and 'hair', I have also relied on the references given by Scarborough (_IE-CoR_: v. cognate sets 38, 703, 7233, 1202).

γr-. This etymon is replaced by **kaufa-* (YAv. *kaofa* 'mountain ridge') > **kōf* in West Middle Iranian.

There are common innovations against West Iranian. Among these are a bunch pertaining to the lexical root. The most famous innovation among these is **kapā* 'fish' (Khot. *kava-,* Sogd. *kp-,* Chor. *kb*, Scythian hydronym *Παντικαπης*, *Παντικαπαιον* litt. "fish-path" < **panθi-kapā-*, Wakhi *kūp*, Yidgha *kap*, Pashto *kab*, Oss *kæf*) which replaced PIr. **mā̆si̯a-* (MP *m'hyg,* Parth. *m'sy'g* < **māsi̯ā-ka-*; cf. Ved. *mátsya-*) (Bielmeier 1989: 243). The ultimate etymology is unclear; it is possibly to be connected with the root in Ved. *kapilá-* (usually cited as 'grey', but also '"monkey-coloured", reddish, brown, tawny') (Ėdel'man 2011: 241; Monier-Williams 1899: s.v.).

An isolated East Iranian root is **gāz-* 'to accept, receive', which is furthermore often found in combination with **pati-*: Khot. *pajāys-* 'enjoy, accept', Sogd. *pcγ'z* 'receive', Chor. *pcy'z-* 'accept' and Wakhi *рысˇыз* 'ask, wish, give back' (Cheung 2007: 118–9).

Similarly, **sāna-* 'enemy' is only found in EIr. (Khot. *sanä,* Sogd. *s'n,* Oss. *son*), but the etymology is unclear. Notably, most suggestions revolve around semantic shifts of Iranian roots, e.g. the connections with Av. *sā-* 'forbid, fight against', *paiti-sā-* 'resist', *sāsta* 'gruesome', *fra-sāna-* 'defeat' or *sād-ra* 'torment' (Abaev 1979: 135).

Also **drau̯a-* 'hair' (Khot. *drau-*, Oss. *ærdo*, Sogd. *žw-*, Yaghn. *dirau*, Orm. *drī*) is limited to East Iranian, according to Sims-Williams (1989; 1996). In the *IE-CoR*, these East Iranian languages share this cognation with Nuristani languages: Waigali (Kalasha-ala) *dru* is traced to PIr. **drau̯-*, and Bailey (Bailey 1979: 170) cites Ashkun *dro*. If these connections between Nuristani and Iranian are correct, and **drau̯a-* is to be reconstructed all the way back to PIIr., not just PEIr., it seems that the innovation is rather the generalisation of one near-synonym over the other in East Iranian. However, given the geographic position of the Nuristani occurrences, one could suspect an "undetected borrowing" from East Iranian. While it is striking at surface level that the East Iranian languages agree on sharing an etymon not found elsewhere in the same meaning, PIr. must have had quite the assortment of near-synonyms since there are many more seemingly old etyma in this meaning in Iranian. Some of these are *not* limited to just one branch, e.g. **u̯ars-* (PIE (or Indo-Slavic?)) **u̯elk̂-*, cf. Slav. **vôlsъ* 'hair', Ved. *válśa-* 'twig') in YAv. *varəsa-*, Sogd. *wrs*, Pashto *wekhta*.

There are also innovations of the semantics of inherited words, such as **mai̯θa-* which means 'day' in EIr. – Sogd., Chor *myθ*, Yaghnobi *meθ*, Shughni *meθ,* Sarikoli *maθ*, Yazghulami *miθ*, Yidgha-Munji *mīx̌*, but Av. *maēθa-* means 'unstable, vacillating'. The proposed connection with Lith. *mẽtas* 'time' is impossible. Iranian **maiθ-* < **mVi̯tH-* is incompatible with **meh₁-* or **med-* 'meassure' (Ėdel'man 2015: 160–2; Morgenstierne 1974: 45–6; Junker 1914: 13; *ALEW*[2]: s.v. *mẽtas₁*).

Finally, there are clear innovations consisting of semantic innovations in combination with derivational morphology. Khot. *vasuta-* 'pure' and Sogd. *'wswγty* 'id.' continue **ava-suxta-* (Av *suxta-*, the perfect participle of *saok/saoc-* 'burn, emit flames' (Cheung 2007: 338–9). Sogd. *βyr,* Chor. *βyr-*, Manic. Bactr. *'βyr-*, Bactr. *αβιρ-* and Yazghulami *vir* all mean 'to find, obtain' and go back to **abi-Har-*. The root **Har-* 'to go to(wards), reach' (< PIE **h₁er-*) is widespread across Iranian, but only East Iranian continue the semantically narrowed extension with the preverb **abi-* (Cheung 2007: 160–3).

6. Bibliography

Abaev, Vasilij Ivanovič. 1979. *Istoriko-ėtimologičeskij slovar' osetiskogo jazyka*. Vol. 3. S-T'. Moskva & Leningrad: Nauka.

ALEW[2] = Hock, Wolfgang, Elvira-Julia Bukevičiūtė, Christiane Schiller, et al. (eds.). 2020. *Altlitauisches etymologisches Wörterbuch*. 2.0 [Online]. https://alew.hu-berlin.de/.

Bailey, Harold Walter. 1979. *Dictionary of Khotan Saka*. Cambridge & New York (NY): Cambridge University Press.

Bielmeier, Roland. 1989. Sarmatisch, Alanisch, Jassisch und Altossetisch. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 236–245. Wiesbaden: Reichert.

Birwé, Robert. 1956. *Griechisch-Arische Sprachbeziehungen im Verbalsystem*. (Beiträge zur Sprach- und Kulturgeschichte des Orients 8). Walldorf Hessen: Verlag für Orientkunde Dr. H. Vorndran.

Blažek, Václav & Irén Hegedűs. 2012. On the position of Nuristani within Indo-European. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphophonemics*, 40–66. (LINCOM Studies in Indo-European Linguistics 41). München: Lincom Europa.

Brixhe, Claude. 2018. Macedonian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 3, 1862–1867. (Handbücher Zur Sprach- Und Kommunikationswissenschaft 41). Berlin & Boston (MA): De Gruyter. doi:10.1515/9783110542431-028. https://www.degruyter.com/document/doi/10.1515/9783110542431-028/html (28 June, 2024).

Budruss, Georg. 1977. Nochmals zur Stellung der Nūristān-Sparchen des Afghanischen Hindukush. *Nachtrag zu MSS 37* 38. 19–38.

Budruss, Georg. 1978. Nachtrag zu MSS 36, 1977, S. 19ff. *Münchener Studien zur Sprachwissenschaft* 38. 37–38.

Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb*. (Leiden Indo-European Etymological Dictionary Series 2). Leiden: Brill.

Cheung, Johnny. 2015. The linguistic and geographic position of Pashto within the east iranian language group: 1. A phonological and lexical analysis. *Iranian Languages and Literatures of Central Asia. From the Eighteenth Century to the Present*, 29–65. (Cahiers de Studia Iranica 57). Leuven: Peeters.

Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. (Publications of the Philological Society 30). Oxford & Cambridge (MA): Blackwell.

Clackson, James. 2022. Methodology in Linguistic Subgrouping. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 18–32. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.002 (23 September, 2022).

Degener, Almuth. 2002. The Nuristani languages. In Nicholas Sims-Williams (ed.), *Indo-Iranian languages and peoples*, 103–118. (Proceedings of the British Academy 116). Oxford: Oxford University Press.

Ėdel'man, Džoj Iosifovna. 2011. *Ėtimologičeskij slovar' iranskix jazykov*. Vol. 4. i-k. Moskva: Vostočnaja literatura.

Ėdel'man, Džoj Iosifovna. 2015. *Ėtimologičeskij slovar' iranskix jazykov*. Vol. 5. l-n. Moskva: Vostočnaja literatura.

Ėdel'man, Džoj Iosisovna. 1999. *Dardskie i nuristanskie jazyki*. (Jazyki Mira 8). Moskva: Indrik.

Emmerick, Ronald E. 1989. Khotanese and Tumshuqese. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 204–229. Wiesbaden: Reichert.

Euler, Wolfram. 1979. *Indoiranisch-griechische Gemeinsamkeiten der Nominalbildung und deren indogermanische Grundlagen*. (Innsbrucker Beiträge Zur Sprachwissenschaft B30). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Garrett, Andrew. 1999. A new model of Indo-European subgrouping and dispersal. In Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society, February 12–15*, 146–156. Berkeley: Berkeley Linguistics Society.

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 for Archaeological Research, University of Cambridge.

Goldstein, David M. 2022. There's no escaping phylogenetics. In Laura Grestenberger, Charles Reiss, Hannes A. Fellner & Gabriel Z. Pantillon (eds.), *Ha! Linguistic studies in honor of Mark R. Hale*, 71–91. Wiesbaden: Reichert.

Greenhill, Simon J. & Russell D. Gray. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica* 29(4). 523–537. doi:10.1075/dia.29.4.05gre.

Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2021. Bayesian Phylolinguistics. In Richard D. Janda, Joseph, Brian D. & Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, 226–253. John Wiley & Sons. doi:10.1002/9781118732168.ch11 (28 September, 2021).

Hegedűs, Irén. 2005. Review of Ėdel'man (1999). *Indogermanische Forschungen* 110. 319–325.

Huyse, Philip. 2017. The dialectology of Iranian. *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 599–608. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). Berlin & Boston (MA): De Gruyter. doi:10.1515/9783110261288-037.

IE-CoR = Heggarty, Paul, Cormac Anderson & Matthew Scarborough (eds.). 2023. *IE-CoR*. Zenodo. doi:10.5281/zenodo.8089434. https://doi.org/10.5281/zenodo.8089434 (18 March, 2025).

Jacques, Guillaume & Johann-Mattis List. 2019. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1). 128–167. doi:10.1075/jhl.17008.mat.

Junker, Heinrich F. J. 1914. *Drei Erzählungen auf Yaynå̄bī.* (Sitzungsberichte Der Heidelberger Akademie Der Wissenschaften, Philosophisch-Historische Klasse 14). Heidelberg.

Kim, Ronald I. 2025. Case Marking from Old Iranian to Ossetic: A Comparative and Typological Approach. *Transactions of the Philological Society*. doi:10.1111/1467-968X.12316. https://onlinelibrary.wiley.com/doi/10.1111/1467-968X.12316 (17 March, 2025). Online Version of Record.

Kogan, Anton I. 2005. *Dardskie jazyki: Cenetičeskaja xarakteristika*. Moskva: Vostočnaja Literatura.

Korn, Agnes. 2016. A partial tree of Central Iranian: A new look at Iranian subphyla. *Indogermanische Forschungen* 121(1). 401–434. doi:10.1515/if-2016-0021.

Korn, Agnes. 2019. Isoglosses and subdivisions of Iranian. *Journal of Historical Linguistics* 9(2). 239–281. doi:10.1075/jhl.17010.kor.

Kümmel, Martin Joachim. 2022. Indo-Iranian. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 246–268. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.014. https://www.cambridge.org/core/books/indoeuropean-language-family/indoiranian/C01D66B2DA24680E0808EFEC10E389C9 (23 September, 2022).

Monier-Williams, Monier. 1899. *A Sanskrit-English dictionary: Etymologically and philologically arranged with special reference to cognate Indo-European languages*. 2nd ed. Oxford: Clarendon Press.

Morgenstierne, Georg. 1974. *Etymological vocabulary of the Shughni group*. (Beiträge zur Iranistik 6). Wiesbaden: Reichert.

Obrador-Cursach, Bartomeu. 2019. On the place of Phrygian among the Indo-European languages. *Journal of Language Relationship • Вопросы языкового родства* 17(3). 233–245.

Olander, Thomas. 2018. Connecting the dots: The Indo-European family tree as a heuristic device. In David M. Goldstein, Stephanie W. Jamison & Brent Vine (eds.), *Proceedings of the 29th annual UCLA Indo-European conference*, 181–202. (Journal of Indo-European Studies Monograph Series). Bremen: Hempen.

Pellard, Thomas, Robin J. Ryder & Guillaume Jacques. forthc. The Family Tree model. In Ledgeway (ed.), *The Wiley Blackwell companion to diachronic linguistics*. Wiley-Blackwell. https://www.academia.edu/101656989/The_Family_Tree_model (22 May, 2023). Preprint.

Pereltsvaig, Asya & Martin W. Lewis. 2015. *The Indo-European controversy: Facts and fallacies in historical linguistics*. Cambridge: Cambridge University Press.

Peyrot, Michaël. 2018. Tocharian B *etswe* "mule" and Eastern East Iranian. In Lucien Christiaan van Beek, Guus Jan Kroonen, Tijmen Pronk, et al. (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*, 270–283. Ann Arbor (MI) & New York (NY): Beech Stave Press.

Peyrot, Michaël. 2022. Tocharian. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 83–101. Cambridge, UK, & New York: Cambridge University Press.

Poulsen, Simon. 2025. Loanwords and Linguistic Phylogenetics: **peleḱu-* 'axe' and *(*H*)*a*(*i̯*)*ĝ-* 'goat.' *Transactions of the Philological Society* 123(1). 116–136. doi:10.1111/1467-968X.12308.

Ringe, Donald A. 2022. What We Can (and Can't) Learn from Computational Cladistics. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 52–62. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.004 (27 September, 2022).

Ringe, Donald A., Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1). 59–129. doi:10.1111/1467-968X.00091.

Sims-Williams, Nicholas. 1989. Eastern Middle Iranian. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 165–172. Wiesbaden: Reichert.

Sims-Williams, Nicholas. 1996. Eastern Iranian languages. *Encyclopædia Iranica* VII. 649–652. doi:10.1163/2330-4804_EIRO_COM_8633.

Sims-Williams, Nicholas. 2021. Plural and Collective in the Sogdian Ancient Letters. *Transactions of the Philological Society* 119(2). 174–181. doi:10.1111/1467-968X.12211.

Skjærvø, Prods Oktor. 1989. Modern East lranian languages. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 370–383. Wiesbaden: Reichert.

Skjærvø, Prods Oktor. 2017. The documentation of Iranian. *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 471–481. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). Berlin & Boston (MA): De Gruyter. doi:10.1515/9783110261288-032.

Smith, Caley. 2017. The dialectology of Indic. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, 417–447. De Gruyter. doi:10.1515/9783110261288-030.

https://www.degruyter.com/document/doi/10.1515/9783110261288-030/html (18 February, 2024).

Strand, Richard F. 2023. Kâmbojâs and Sakas in the Holly-Oak Mountains: On the origins of the Nûristânîs. In Alberto M. Cacopardo & Augusto S. Cacopardo (eds.), *Roots of Peristan: the pre-islamic cultures of the Hindukush/Karakorum: proceedings of the International Interdisciplinary Conference ISMEO, Rome, Palazzo Baleani, 5-7 October, 2022*, 781–808. (Serie Orientale Roma. Nuova Serie 37). Roma: Ismeo.

Tremblay, Xavier. 2005a. Bildeten die iranischen Sprachen ursprünglich eine genetische Familie oder einen Sprachbund innerhalb des indo-iranischen Zweiges? Beiträge zur vergleichenden Grammatik der iranischen Sprachen V. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel: Akten der XI. Fachtagung der Indogermanischen Gesellschaft, 17.–23. September 2000, Halle an der Saale*, 673–688. Wiesbaden: Reichert.

Tremblay, Xavier. 2005b. Irano-Tocharica et Tocharo-Iranica. *Bulletin of the School of Oriental and African Studies* 68(3). 421–449. doi:10.1017/S0041977X05000248.

Windfuhr, Gernot. 2009. Dialectology and topics. In Gernot Windfuhr (ed.), *The Iranian languages*, 5–42. Abingdon, Oxon, & New York: Routledge.

A tree or not? An East Iranian experiment

Agnes Korn (CNRS, Paris) & Simon Poulsen (University of Copenhagen)

Abstract

While the languages of the Iranian branch have traditionally been subdivided into West and East Iranian (EIr.), recent research has suggested that EIr. is not a genetic entity, and no EIr. protolanguage can be reconstructed. Moreover, suggestions for the subgrouping of EIr. have diverged widely. Depending on the favoured solution, some isoglosses have been considered important, and the remaining ones irrelevant, revealing a certain amount of arbitrariness.

In this paper, we suggest tackling this arbitrariness by presenting an attempt at a computational phylogenetic analysis of the subgroups within EIr. We interpret the traditional phonological and morphological isoglosses of EIr. as potentially shared innovations and use them in a manually coded dataset of 17 EIr. languages (ancient and modern). We weigh the isoglosses and assign typologically informed differentiated costs to the changes between all outcomes attested. We then apply the Graphic User Interface *LinguiPhyR* to perform a Maximum Parsimony analysis. The software generates the most economic ("parsimonious") phylogenetic tree, i.e. the tree that requires the lowest cost to get from Proto-Iranian (which we use as the ancestral state) to the attested languages. We evaluate the trees using the app's post-processing tools to assess the inferred developments, the relative chronology of each isogloss, and the influence of each isogloss on the intermediate nodes.

Our results are inconclusive. Although the app finds some lower-order groups (e.g. Wakhi as a sister of Yidgha-Munji and Ishkashimi as belonging to the Shughni-Yazghulami group, multiple iterations yield differing results, which underlines the arbitrariness problem outlined above. Our best trees do not entirely agree with any of the models proposed in the literature; e.g. Khotanese is always returned as an outlier – either as the first or second language to branch off; Sogdian does not come out as a close relative of Yaghnobi, Ormuri and Pashto are grouped together rather than Ormuri and Parachi.

Some isoglosses are consistently incompatible with the best trees. In some cases, the software must infer large-scale loss or parallel innovations to fit them onto a genetic tree. Although inconclusive, the results leave us better equipped to evaluate the EIr. isoglosses – some may be compatible with an underlying tree, and others (notably the innovated nominal plural endings from the collective in *-*tā*) should probably be regarded as areal features of the "Sprachbund" that EIr. (as well as some of its subgroups) has been suggested to represent.

1. Introduction

1.1. Background

The purpose of this paper is to test the Graphic User Interface *LinguiPhyR* (Canby 2024), called "the App" in what follows, on a group of languages that according to current scholarship does not form a family tree, but could instead be regarded as a linguistic area (*Sprachbund*). The App does not in itself provide a novel analysis method, but it makes Maximum Parsimony (see Section **1.4.** below) much more accessible for non-computationalists. The objective is two-fold: on the one hand, we will evaluate the methodology of the App; on the other hand, we will search for any phylogenetic relationship potentially underlying the contact-induced isoglosses of the East Iranian *Sprachbund*. A family tree and a Sprachbund are not mutually exclusive. Exploring the isoglosses under a phylogenetic framework, and especially performing the Maximum Parsimony analysis, will leave us better equipped to evaluate which isoglosses could have developed in a tree-like manner. We can also compare the consequences of the competing hypotheses on the same material: As elaborated below, the many different solutions to the subgrouping of the East Iranian languages build on positive evidence from different isoglosses. However, drawing the conclusion based on one or a few isoglosses leave the development of others in the dark. Our analysis will further the understanding of the relative chronology of the changes and add quantitative argumentation the study of the EIr. languages as a Sprachbund.

1.2. Aim of the study

While East Iranian (EIr.) is traditionally cited as a sub-branch of the Iranian (Ir.) branch of Indo-European, and further subdivided into a northern and southern group, recent research has shed serious doubt on this model.

As pointed out by Sims-Williams (1996: 651): "if one reconstructs "proto-Eastern Iranian" in such a way as to account for all the features of the group, it proves to be identical to the "common Iranian" reconstructible as the ancestor of the whole Iranian family." This means that there are no innovations that can be attributed to the time span separating Proto-Iranian (PIr.) from Proto-East Iranian, and it is impossible to establish isoglosses that characterise East Iranian as a whole. Indeed, features traditionally assumed to define East Iranian either do not hold for all of East Iranian (i.e. the isoglosses are not inclusive), or are also found in some West Iranian languages (i.e. the isoglosses are not exclusive). For some isoglosses, both even apply at the same time (Korn 2019: 248f.).

Rather, East Iranian can be regarded as a *Sprachbund* (Sims-Williams 1996: 651; 1989a: 165). The isoglosses, rather than being diagnostic for phylogenetic subgrouping, can then be considered as typical features of East Iranian.

Something similar holds for the sub-groupings of East Iranian. As highlighted by the fact that the existing suggestions differ dramatically (Figure 1), there is no consensus which languages should be assigned to which group, with the majority of East Iranian being South-Eastern under one approach and North-Eastern under the other.[1] Some have made Parachi and Ormuri one sub-branch (Southeastern) with the remaining languages constituting the other branch (Northeastern) (a), while others have held Sogdian, Yaghnobi, and Ossetic to be one sub-branch (North-Eastern) against Khotanese, Bactrian, Pashto, and the Pamir languages (South-Eastern) (b). Even others include Bactrian in the Northeastern branch against Khotanese, Parachi, Ormuri and the Pamir languages (c). In a variation of (a), Parker (2023: 44) adds a few subgroups to the Northeastern branch: Munji-Yidgha, Sanglechi-Ishkashimi and North Pamir consisting of Yazghulami and Shughni-Roshani-Sarikoli (d). A more complete presentation of the Pamir languages (e) can be found in Wendtland (2009: 173) and Sokolova (1966: 124).

One recent computational phylogenetic study of the relationship between 161 Indo-European language varieties finds some EIr. branches, but groups West Ir. closer to most EIr. languages than Ossetic, Bactrian and Khotanese to the rest of EIr. (Heggarty et al. 2023a; 2023b: 57–8).[2] This analysis only includes 10 EIr. languages, and not all subgroups are directly comparable to the traditional solutions, but it is notable that it groups Bactrian and Ossetic against the rest, groups Sogdian and Yaghnobi closely – and Wakhi and Sarikoli (the two only Pamir languages included in the study).

These widely differing suggestions based on essentially the same data and a largely shared set of features is the result of very different evaluations of those features in terms of which ones may be used as isoglosses, the remaining ones being declared irrelevant for the subgrouping. Depending on the individual author, there are very different assessments of the diagnosticity of the various features. There is thus a certain amount of arbitrariness.

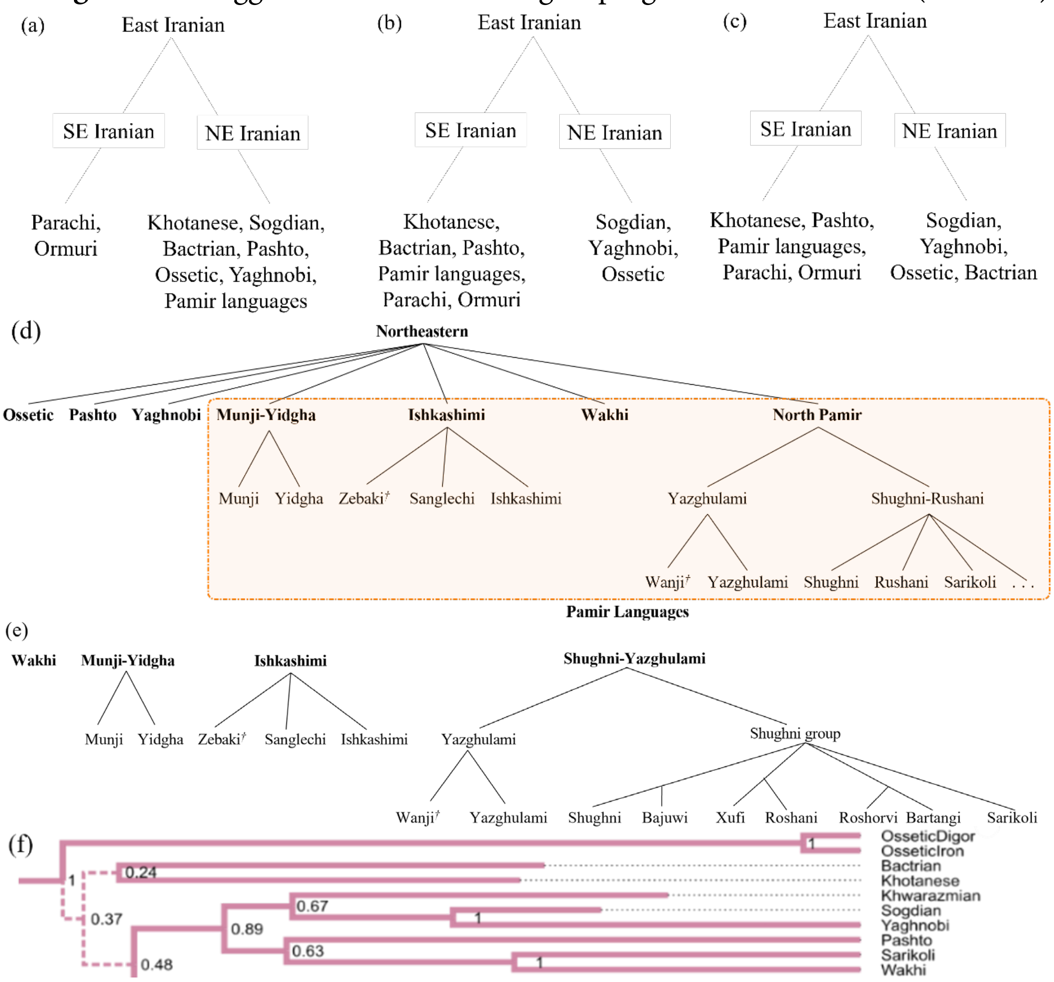
The arbitrariness just mentioned is the starting point for our study. We intend to use the features traditionally used for defining East Iranian and/or its sub-

[1] For a summary, see Wendtland (2009: 172f.).

[2] The methodology and data differ significantly from our study. Heggarty et al. (2023a) is a Bayesian analysis of shared cognates in 170 “basic” meanings. These can be found in the IE-CoR dataset: https://iecor.clld.org/. Some discussion of the choice is published in Scarborough (2020). See also **1.4.2.** below.

branches as a test-case for what happens when we apply an orthodox phylogenetic method to data from a suggested Sprachbund. We are curious to see if any isoglosses are compatible with an underlying family tree, and to compare the competing hypotheses on the same terms.

Figure 1: Suggestions for the subgrouping of East Iranian (selection)



(a) Morgenstierne (1926: 27–39) (followed by others). (b) As advocated by Oranskij (1979: 177ff.) (c) The “traditional” tree in Korn (2016: 402). (d) Parker’s (2023:44) Northeastern clade. (e) Sokolova’s (1967: 124) presentation of the Pamir languages. (f) Subsection of the Maximum Clade Credibility tree of Heggarty et al. (2023b: 57–8). Numbers (between 0 and 1) next to the nodes are posterior probabilities (i.e. the certainty) of these clades.

1.3. Data

This study uses isoglosses generally used for defining East Iranian and its subgroups. In order to avoid bias, we rely on Wendtland 2009 (with concluding table p. 185) for the isoglosses and to a large extent also for the data.

Likewise following Wendtland 2009, the languages discussed in this study include:

- Middle Iranian: Khotanese, Sogdian, Chorasmian, Bactrian,
- New Iranian: Shughni,[3] Sarikoli, Yazghulami, Ishkashimi, Yidgha and Munji, Wakhi, Yaghnobi (West and East separately), Pashto, Ossetic, Ormuri, Parachi.

Unless noted otherwise in Section 2 below, the data and etymologies for the individual languages mentioned are also from Wendtland (2009), who provides ample references.[4] As she focusses on the most salient results in terms of isoglosses, data for the remaining languages were added from various sources (for additional sources for individual words and morphemes, see the individual isoglosses in Section 2).

A study like this will always benefit from a larger dataset, but it is valuable in itself to evaluate the isoglosses usually presented as evidence for a Sprachbund in a phylogenetic framework. Moreover, isoglosses presented in the literature tend to be skewed in favour of a single analysis. For example, according to Skjærvø (1989a: 375) the development of PIr. **ću̯* into a (geminated) sibilant **š(š)* and not **sp* "strongly suggests that the ancestors of Khotanese and Wakhī were closely related". However, this isogloss hinges on the single word **aću̯a-* 'horse' (Khot. *aśśa-*, Wakhi *yaš*; Av. *aspa-,* OP *asa-*). If this etymon was borrowed between the diverging Iranian languages, the argument no longer holds. This study is a first attempt at an overarching approach and can serve as basis for future expansion.

[3] Wendtland occasionally distinguishes the Xufi, Roshani (Rušanī) and Roshorvi (Orošorvī) dialects from "standard" Shughni. Upon closer inspection, the codings of the dialects would be virtually identical for our purposes (see **m19**) – or we would not have the resources to distinguish them in the detail we need. **m1** is the only character where the distinction would be relevant. Therefore, we have collapsed them into one Shughni language encompassing the whole group.

[4] We have supplemented Wendtland's material with general references postdating the publication of her paper (Khotanese: Emmerick 2009; 2024; Skjærvø 2022; Sogdian: Yoshida 2009; Wendtland 2011; Chorasmian: Durkin-Meisterernst 2009; Ėdel'man 2008; Bactrian: Gholami 2014; Pamir languages: Edelman & Dodykhudoeva 2009a; Shughni: Edelman & Dodykhudoeva 2009b; Parker 2023; Sarikoli: Kim 2017; Wakhi: Bashir 2009; Pashto: Robson & Tegey 2009; Ormuri: Efimov 2011; Parachi: Kieffer 2009).

1.4. Method

1.4.1. Computational phylogenetics

As mentioned in **1.1.**, this study has the two-fold aim of testing the App as well as presenting a phylogenetic analysis using maximum parsimony of East Iranian.[5] This approach (see below) is based on the family tree model.[6]

This is not straightforward for East Iranian (see Section **1.2.**). However, computational assistance can help select the best among the $192*10^{15}$ rooted bifurcating trees that are theoretically possible with the 17 languages surveyed. The resulting trees can help us evaluate the relative chronology of changes and the intermediate protolanguages.

1.4.2. Character-based cladistics

Maximum Parsimony searches for the most parsimonious (i.e. economical) way to get from the **root**, Proto-Iranian (see below), to the **leaves**, the attested languages. This implies placing innovations on a tree with as few changes as possible. For instance, unless there are arguments to the contrary, it will be more economical to place an innovation found in several languages higher up in the tree (thus only once) rather than having the innovation twice in the sub-branches. The approach mimics what a historical linguist would (i.e. Sims-Williams 1989a; Wendtland 2009) and is thus an appropriate method for the languages and the data studied here. Of course, it is also necessary and possible to take loss into account. Although it has been deemed old-fashioned and simplistic (Pellard, Ryder & Jacques *in press*: 14), it has recently been defended for computational reasons (Canby et al. 2024; Ringe 2022: 59).

Owing to the fact that Old Iranian (and the closely related Old Indic) languages are well-attested, we can reconstruct the ancestral state of the languages studied. Moreover, typological considerations often clarify via which intermediary stage(s) a given output is likely to have developed. The direction of phonological and morphological changes is thus not arbitrary (X > Y is probable while Y > X would be highly unlikely). This is referred to as "directionality" in what follows. Directionality is easily incorporated in a dataset for Maximum Parsimony.

---

[5] For the details of the software and the mathematical aspects of Maximum Parsimony, see the description of the App (Canby 2024, with refs.).

[6] It is beyond the scope of this chapter to defend the choice of the tree model over the wave model, and to discuss the mathematical and computational aspects of the methodology in detail. The family tree model, simplistic as it is, still plays an important role in linguistic reconstruction (Goldstein 2022; Olander 2018; Jacques & List 2019).

While it would be possible to use other computer-assisted methods on our dataset, these would be less adequate. **Maximum Compatibility** works similarly to Maximum Parsimony, but rather than using the change within the isoglosses determine the tree, it relies on the compatibility of each isogloss. Since we do not expect many entire isoglosses to fall easily into subgroups, we would risk basing our tree on only a fraction of our material. For instance, not all outcomes of **θr-* develop neatly along the branches of a tree. Accordingly, this entire isogloss would not inform a Maximum Compatibility analysis on the tree structure. In a Maximum Parsimony analysis, the isogloss is considered even if only parts of the outcomes fit on a tree.

Bayesian analysis is seen as a reliable method because it takes statistical uncertainty into account. In linguistics, it is mainly applied to lexical data.[7] In our view, lexical data are less relevant for studying genetic relationships (Clackson 1994: 23–25). This is particularly true for Iranian, as these languages stayed in contact during their diversification; they are thus likely to share lexical items in a non-tree-like manner. Bayesian analysis is less adequate for phonological and morphological data because directionality would be inferred by the model, i.e. a Bayesian approach would not permit us to integrate our knowledge of the directionalities of changes directly. In more practical terms, the result of a Bayesian analysis is very far from the linguistic input. While it incorporates statistic insecurity, the model is opaque. We aim to use our computational analysis as a supplementary tool, not to replace existing methods. With a Parsimony analysis, we can easily examine the effect of each isogloss on the resulting tree (Canby 2024: 1).

1.4.3. From isoglosses to states of characters

Our approach requires that the data are presented as a matrix of "characters" (the isoglosses) and "states". This format has two advantages outside the computational analysis as well: it avoids getting distracted by positive evidence, ignoring potential counterarguments, and it requires taking decisions in difficult cases.

[7] Scholars arguing for the use of such data in linguistic phylogenetics prefer the term "Lexical Cognacy" (Greenhill, Heggarty & Gray 2021: 234ff.; Greenhill & Gray 2012), since the method examines whether the languages share a given cognate etymon in a given meaning ("semantic slot"). Bayesian analysis also requires data with binary features. Our dataset is not binary, but multi-state (for instance, **θr* yields a number of different results in the various East Ir. languages), but this would not be an obstacle a priori (one could convert the isogloss into a group of features expressed in terms of presence vs. absence of a specific outcome of **θr*) (Canby et al. 2024: 6).

A **character** can be any "linguistic property which languages can instantiate in a variety of ways, and languages which instantiate the character in the same way are assigned the same **state** of that character." (Ringe, Warnow & Taylor 2002: 71) We reformulate the isoglosses (i.e. those of Wendtland (2009) and Korn (2016: 405), see Section **1.3.**, with some adjustments) into innovations that could have taken place in the ancestor of the languages sharing the same output.

In character-based cladistics, each **character** is **independent**, and we should formulate our isoglosses so that exhibiting a change in one isogloss is not made possible by a change in another isogloss. However, many of our characters could be seen as interconnected, e.g. the lenition of Proto-Iranian word-initial **b d g* could be seen as one shift. On the other hand, there is a considerable degree of variation in the outcomes across the languages, e.g. the change **d-* > *l* has no counterpart in the changes of **b-* and **g-*, so that it seems justified to treat them separately. The App does not allow us to mark **codependent** characters, and all isoglosses are mathematically treated as independent. Our data is too complex to live up to this ideal, but we have tried to resolve the issues it causes by various strategies (see **1.4.4.** and **1.4.5.**).

The outcome in each language is expressed as a numerical **state**, and the change leading to the outcome is called **transition**. Ideally, each state represents "an identifiable unique historical stage of development" that could be shared between the languages sharing it.[8] We label the Proto-Iranian[9] state (the input, "**root**") *0*, and the various states (of the "**leaves**", i.e. the daughter languages) *1*, *2* etc.[10] For the analysis, the numbers merely identify the various states. In the few cases where we have been unable to assign a state to a language, "?" is used. This is replaced by a unique character (see below) in the analysis.

For a character to be (parsimony) **informative**, the App must be able to fit its states onto a tree (which does not imply that it needs to fit on all possible trees).

[8] In-depth explanations of how linguistic innovations are coded as numerical states can be found e.g. in (Ringe, Warnow & Taylor 2002: 71–9; Nakhleh et al. 2005: 410–8).

[9] Although the existence of a Proto-Iranian language has been doubted (Tremblay 2005), attested Old Iranian in combination with the later languages is a solid basis for reconstructing at least a proxy of the ancestor of the attested Ir. languages and for determining the ancestral state of the various isoglosses.

[10] The states are arranged from minor to more marked innovations (e.g., in the case of **xt*: partial voicing (*γt*) is state 1, followed by voicing of the whole group (*γδ*) as state 2 and then more marked changes. For the morphological isoglosses, inherited forms come first, followed by innovations). This sequence is purely for the convenience of readers, though; a higher number of the state does not imply that the result has gone through the states with lower number (although in some instances this may indeed be the case).

Languages showing an output identical to Proto-Ir. have the state *0*, but for the configuration of a family tree, shared archaisms are insignificant. In addition, if all languages shared the same outcome, a character would be **uninformative** since it cannot tell us anything about the configuration of a possible family tree. When two or more languages share a state, this translates to them potentially having undergone the same innovation in their prehistory, but our approach does allow for the possibility that a given innovation has occurred twice (or more) independently, i.e. we do not force common ancestry for shared innovations.

Outcomes that are specific to one language ("**unique states**") are in principle also uninformative, because a truly unique innovation could have happened at any node. However, most of the unique innovations in our dataset cannot, in fact, have occurred anywhere in the tree. This is the case for many phonological states. For instance, Ossetic is the only language in our dataset to exhibit the state 2 (*rt*) as the outcome of PIr. **θr*. Oss. *rt* is more likely to descend directly from the root (i.e. **θr*) than from one of the states (*hr, sr, c, š* etc.) shown by the other languages. Even though the Ossetic state is technically unique, our approach allows it to provide information for the computational analysis and thus for the configuration of the tree, just as it would in a non-computational phylogenetic analysis.

Conversely, for many morphological innovations, unique states are truly uninformative, since we have no way of knowing which, if any, innovations the language underwent before it obtained its unique state. For instance, the Sarikoli nominal plural morpheme *-xεγl* certainly replaced an inherited inflectional ending at some point, but which one exactly and at what stage is impossible to know.

1.4.4. Polymorphism and the strategies to avoid it

The EIr. languages often show more than one state of a given input ("**polymorphism**"). Unfortunately, the App handles **polymorphism** rather poorly. For instance, Khotanese shows both *č* and *c* as the outcome of PIr. **č*. If we coded this as though Khotanese exhibited *č* and *c* simultaneously, the App would automatically replace this polymorphism with a monomorphic state.[11] We would have no control over the allowed directionalities of this automated state and would thus lose valuable information.

Instead, we have reduced polymorphic outcomes by various strategies. Where feasible, only one output per language is considered, either the most archaic stage or, if this is too fragmentarily attested, the best attested stage. For instance, Proto-Iranian **θ* might be preserved in Bactrian in the one instance of *ιθαο* 'thus' if this

[11] We would be able to choose between a unique state for every instance of polymorphism, a unique state for every unique combination or the majority state.

derives from PIr.*$i\theta\bar{a}$; this is disregarded here because in all other instances *θ yields *h* (Wendtland 2009: 176 fn. 38).[12] If a development has taken place during the attested history of a language, this too is disregarded, as it cannot have occurred in a potential common ancestor. This is the case for the Bactrian noun inflection, where a two-case system is well attested in the inscriptions, but the two cases have coalesced by the period attested in the manuscripts. In the few cases when the outcomes are not **effectively polymorphic**, we have disregarded unique morphs that would not have any influence on the configuration of the tree anyway.

Where a twofold outcome seemed sufficiently important to be considered as such, we have formulated it as the property of a single state, i.e. we have assigned a monomorphic numerical state to a "polymorphic" or diverse outcome in the language. This is the case for the Khotanese outcomes of **č* mentioned above (see further **2.3.4.**) where we have assigned an intermediate stage *č, c* (1) to Khotanese rather than assigning it *č* (0) and *c* (2) simultaneously. Such states are treated like any other state, with costs for the changes leading to them and to the changes they may develop into.

When this approach was impossible because the polymorphism was more than two-fold or more languages were polymorphic, we have resolved the polymorphy by splitting the isogloss into several characters. An example of this is the verbal third plural ending. Pashto simultaneously preserves an ending going back to 3PL.ACT *-*nti* but also introduces the ending the third singular. Since the two innovations (the generalisation of *-*nti* over middle ending and the analogical merger with the third singular) can cooccur, we decided to split this into two characters, rather than treating the Pashto outcome as a likely intermediate step.

In those cases where such a split created codependent characters, we have downweighed them accordingly so that the same innovation is not counted multiple times (see **1.4.5.**).

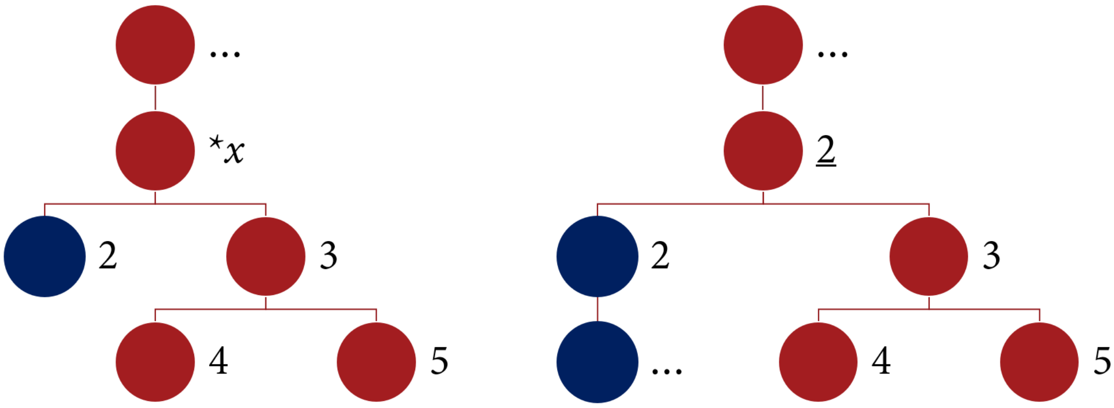
1.4.5. A strategy for handling unattested intermediate stages

Unfortunately, the App currently does not support including states not occurring in the dataset as intermediate steps. This is a drawback, since some innovations can best be analysed as derived from a reconstructed stage which is coincidentally neither reconstructed for PIr. nor preserved in any of the languages included. We have handled such cases by using an otherwise unique derivative of such an intermediate step as a proxy for the unattested innovation. Although it causes a mismatch between the numeric states and the outputs of the languages, it

[12] Sims-Williams (2007: 218) suggests that <θ> is a historic writing for /h/.

is viable because it leads to topologically identical trees. An example of this strategy is the 3pl-endings, where the Ossetic state is used as a proxy for the innovation underlying multiple other states. Most languages generalise either the middle or the active ending. The generalisation of the active gives rise to new innovations, some of which could also be informative. Ossetic has innovated further on its own (see **2.3.4.**), but it does share the generalisation of the active with many other languages. Therefore, we have used the state exhibited by Ossetic as a proxy for the shared innovation, knowing that this is not directly what is attested in Ossetic. This principle is visualised in Figure 2.

Figure 2: Two topologically identical coding strategies applied to the same material.



Since all attested states can be derived from the same unattested innovation (**x*), the state of a unique derivative (marked blue) can be used as a proxy for the innovation. In this example, Ossetic is represented by the blue circle. The Ossetic state (*2*) and the states *3*, *4* and *5* are all derivable from the shared innovation **x* (the generalisation of *-*nti*). The specific Ossetic outcome (*-uncæ, -ync* < *-*n-θ*-) is not shared by other languages, and although it is also a further derivation of the unattested innovation, the state can be used as a proxy for the innovation without changing the tree-structure.

1.4.6. Weights and costs

The App allows us to rank the relevance of the various developments studied by assigning **weights** and **costs**. Weight is the relative importance of entire characters, whereas cost applies to the transitions to the states of each character.[13] We use this to mimic "typological significance" in traditional phylogenetic terminology.

[13] The App allows three types of characters: "standard" (i.e. states are reversible, e.g. 0 > 1, 1 > 0), "irreversible" (e.g. only 0 > 1 is allowed), and "custom". For the latter, the possible directions of change are manually defined, and all allowed transitions can be assigned differentiated costs.

We have kept both weights and costs relatively low for two reasons. First, weights and costs are multiplied, and we wish to balance multiple trivial innovations with few obscure ones (other studies on comparable material, e.g. those of Ringe et al., make phonology and morphology infinitely important than lexicon, essentially leaving the latter irrelevant for the configuration of the tree). Second, our isoglosses are not very significant *per se*, rather the individual changes are – significance of an innovation does not follow from the significance of the isogloss. Our principle for weights is that phonological characters are weighted 1, morphological characters pertaining to restructuring of morphemes in the inherited system 2, and innovations of new categories 3 (see Section 2 for the argumentation of individual cases). While our weights and costs are admittedly to a certain extent subjective, not assigning differentiated weights and costs would amount to ranking all characters and transitions equally, which would be a standpoint that hardly anyone would take.

In those cases where we resolved extensive and important polymorphy by splitting polymorphic characters into multiple monomorphic ones, we have downweighed them accordingly. For instance, **m3** (the nominal plural paradigm DIR *-t,* OBL *-ti*) is directly dependent on **m2** (the innovation of the plurals from abstracts in *-*tā*). The former is downweighed to *1*, and the latter is kept at *3*, not to inflate a single isogloss unnecessarily.

Assigning relatively low **costs** to individual transitions allows us to include data that would usually be considered insignificant for linguistic subgrouping. For instance, phonetic changes that are not phonologically relevant have often been disregarded because it cannot be ruled out that they took place in the prehistory of a language but were reversed before the earliest attestation (Hoenigswald 1966; Clackson 2022). Potential examples of such changes include **č* > *c*, which could thus be reversed if there is no /c/ in the language. However, instead of omitting it from the analysis, we have assigned low costs to both **č* > *c* and its potential reversal **c* > *č* so that the transition can still contribute to the larger picture. Indeed, the non-phonological status of seemingly reversable changes such as **č* > *c* is often only theoretical, as loanwords often provide new instances of *č*. even if the sound may be peripheral phonologically at least at least for a certain time.

We have hard-coded the cost of all potential transitions between all states for each character on a scale of 1 (a typologically likely change) through 5 (a typologically implausible change). For the phonological developments, we have consulted Kümmel (2007). Standardising transition costs for morphological innovations is less obvious. We have followed the principles that easily explainable analogies and generalisations are assigned cost 1, innovations of new suffixes cost 3, and highly unlikely unparalleled innovations are assigned cost 5. Preserving a state, i.e. not changing state and not innovating, does not have a cost. Changes we consider impossible, such as phonemes reemerging once lost, or a plural paradigm of two cases reverting to the Proto-Iranian multiple case endings, for instance, are disallowed in the analysis. However, we have done so sparingly and preferred to assign relatively high costs to questionable transitions.

Figure 3: Scale of the transition costs used in the coding

cost	**description**	**typology**	**examples**
1	very common	many parallels cross-linguistically	voicing of intervocalic **t*, generalisation of one of two competing endings
2	plausible	occasional parallels	word-initial lenition of *p-*, merger of two cases
3	possible	isolated parallels	*d* > *r*, innovation of new morpheme in an existing function
4	odd, but not impossible	few or uncertain parallels	loss of intervocalic **t*, innovation of complex suffix-conglomerate
5	very odd and entirely implausible	no parallels known	**du̯-* > Armenian *erk-*, extensive reductions in case-system

Many of our characters are unlikely to have developed from root to leaf state without intermediate steps. We have taken this into account when assigning the costs so that the path reconstructed by experts is always a viable, but not the only, option. For example, it is theoretically possible to derive the Parachi nominal plural ending *-ān* from PIr. GEN.PL **-ānām* directly, but it is much more likely that it is a result of a gradual restructuring of the case system (see **2.3.1.**).

There are no limits on the number of changes a character can undergo from the root (state *0*) to the states of the attested languages. However, as every transition adds to the total cost of the tree, the method will not assume unnecessary steps (which includes reversals) when searching for the tree in with the lowest cost across all changes required for all characters.

We do not view shared **loss** as a significant innovation for establishing genetic relationships. However, not assigning any costs to loss would push innovations up the tree, as it would not add to the total cost to lose the same item multiple times independently. We have cautiously kept all transitions leading to loss at cost 1.

2. The characters and the outcomes of the isoglosses in the individual languages

2.1. Isoglosses and characters

This section presents the isoglosses used for the present study. Once again, the data compromises the traditional isoglosses as summarised by (Wendtland 2009). They are presented in the same order. For the most part, we rely on Wendtland's discussion and references, and we only add more material or further discussion when we differ. This is most straightforward for the phonological isoglosses (**2.2.**) – after all, the outcome of a phoneme is simpler and comes much closer to a numerical state. While some EIr. languages preserve a considerable amount of Proto- and Old Ir. morphology, other morphosyntactic categories disappear, and new morphology arises. Capturing the shared features of such morphological restructuring is much more complex. Therefore, the discussion of morphological isoglosses (**2.3.**) takes up a larger part of this section.

The section is structured as follows: First, the isogloss and the reasoning for the encoding is presented. Some features are split into several isoglosses (see also **1.4.3.**). Then the states (i.e. the outcomes occurring in the attested languages) are presented in a table which gives the character id (an identifier for the App), the feature (a brief identifier of the character for human readers), the weight of the feature (**1.4.6.**) and the numerical states corresponding to the linguistic outcome (**1.4.2.**). Below this table is a brief description of the principles of the restrictions on directionalities and the individual transition costs. Summing all of these up in prose would be tedious and pointless – the overall principles are outlined in **1.4.6.**, and the full list of allowed directionalities and their transition costs can be found in the appendix (see **7.**). The full coded dataset and all outcomes can also be seen in the appendix (see **7.**).

2.2. Phonological characters

2.2.1. p1: PIr. word-initial **č* (Wendtland §2.1)

Several EIr. languages show a change of **č* [t͡ʃ] to *c* [t͡s] and a parallel change of **ǰ* [d͡ʒ] to *j* [d͡z]. The clearest cases concern word-initial **č*, for which reason this context is used for the isogloss. We disregard the Bactrian output *s-* mentioned by Wendtland because the Manichean Bactrian manuscript points towards *ts-* as result of **č-* (Sims-Williams 2011a: 171). Khotanese likewise has *ts-* generally, and *č-* only in palatalising contexts until umlaut phenomena obscure the original distribution. This is added as a potential intermediate state.

States:

id	feature	weight	0	1	2
p1	*č-	1	č	č/c	c

Directionality and costs: All changes of state are trivial, allowed and assigned cost 1 since there is no phonemic merger. However, *c* > *č* is slightly rarer and therefore assigned cost 2.

2.2.2. p1–3: PIr. word-initial voiced stops (Wendtland §2.2)

Most EIr. languages show a fricativisation of word-initial voiced stops. The results of word-initial **b*, **d*, **g* are treated individually here since the outputs are not entirely parallel in the various languages. Languages that show *b-*, *d-*, *g-* may do so either by preserving an archaism (no change having occurred) or as a result of a backformation of *β, δ, γ* that they would share with other languages. The output *l* seen in Bactrian, Pashto, Yidgha and Munji is likely to be a secondary result of *δ*, i.e. these languages probably shared the development to *δ* found in most other languages before *δ* changed to *l*. We have not included the alleged Sogdian dialects showing **l* instead of *δ* (*pace* Wendtland 2009: 175; Sims-Williams 1989a: 168, n. 6). These are not attested as such, but the development is inferred from circumstantial evidence such as loanwords in Persian and Uyghur palaeography (Sims-Williams 1981: 353; Livshitz 1970: 262; Henning 1958: 61).

States:

id	feature	weight	0	1	2
p2	*b-	1	b	β	
p3	*d-	1	d	δ	l
p4	*g-	1	g	γ	

Directionality and costs: Fricativisation of word-initial stops is slightly rarer (cost 2) than word-initial fortition (cost 1). A change of *d* > *l* directly is less expected (cost 3) than *δ* > *l* (cost 2). *l* > *d* is assigned cost 3.

2.2.3. p5-6: PIr. consonant groups of fricative + *t* (Wendtland §2.3)

The voicing of the consonant clusters **xt* and **ft* is the most widespread phonological isogloss and the only innovation common to all the principle East Iranian languages (Sims-Williams 1989a: 167; 1996: 650). Khotanese shows *t*, which is reduced in the later stages of the language. Efimov (2011:38, 84, 87) presents the outcome of both **xt* and **ft* as zero in Ormuri, but *duka/dua* 'daughter'[14] and *ho/wo* 'seven' speak for *k* as one output of the former and for a vocalic reflex of the latter.

States:

id	**feature**	**weight**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
p5	*xt	1	xt	γt	γδ	yd	t	k/∅	∅	
p6	*ft	1	ft	βt	βδ	vd/ud	ut	b	w/u	∅

Directionality and costs: regressive devoicing of a single phoneme and (γt > xt) and loss of *γ* in *γt* are assigned cost 1; reduction of *xt* > *t*, and lenitions of *δ* > *y* are assigned cost 2; loss of intervocalic unvoiced stops, lenition of *xt* > *yd* and intervocalic fortitions are assigned cost 3. Partial voicing devoicing of an intervocalic voiceless cluster are assigned cost 4, and spontaneous loss of consonant groups and swaps of *t* and *k* are assigned cost 5. Phonemes reemerging once lost are disallowed.

2.2.4. p7: PIr. word-internal *-*θ*- (Wendtland §2.4)

Many Ir. languages show a change of **θ* to *h*, but the preservation of **θ* has been considered as an EIr. isogloss. Only some languages preserve **θ*, while many innovates further. Only word-internal **θ* is included here, unlike in Wendlandt 2009, because the examples of word-initial (prevocalic, but see **1.3.** and **2.2.4.**) **θ*- are very rare, and the outcomes do not necessarily match those of the more securely attested word internal *-*θ*-.

In Bactrian, the regular output is *h* (see **1.3.**). For other reflexes, see Wendtland (2009: 177-8). The outcome in Wakhi is unclear: it is usually said – without examples – that *θ* is preserved (Skjærvø 1989a: 371, 375; Steblin-Kamenskij 1999: 21,32,38). The example(s) in Wendtland (2009:177) and Paxalina (1987a: 429) are all initial, and Morgenstierne (1938: 445) notes that *θ* might be secondary. We follow Steblin-Kaminskij (1971: 7; 1999: 38) and code the outcome as "?", which does not impact configuration of the tree.[15]

The outcome *l* found in Pashto could indicate a change of **θ* to **δ* prior to the change of the latter (see **2.2.2.**) to *l*. The App does not allow us to derive the outcome

[14] We note that *duka* might go back to **duxta-kā-* (Efimov 2011: 60), in which case Ormuri should be assigned ∅.

[15] The form *γaθ* < **gūθā* 'faeces' quoted by Cheung (2015: 52) as Wakhi is in fact Shughni. Wakhi has *gi* in this meaning (Rastorgueva & Ėdel'man 2007: 294–5). The mistake is probably due to Morgenstierne's (1938: 21*) confusing use of commas.

l through a state not expressed in any language. We have therefor assigned cost 2 to *θ > l*, mimicking the cumulative cost of *θ > *δ* (cost 1) and **δ > l* (cost 1) in p3.

States:

id	feature	weight	0	2	3	4	5	6	7
p7	*-θ-	1	**θ**	t	s	x̌	h	l	y

Directionality and costs: Aspirates and sibilants becoming fricatives are assigned cost 1. Rare sound changes like *th > s* or *h* are assigned cost 3. Exceedingly rare changes like *s > t* are assigned cost 4. Cost 5 is reserved for sound changes that we do not know parallels of and consider highly unlikely.

2.2.5. p8-9: PIr. **θr* (Wendtland §2.5)

PIr. **θr* yields very divergent results depending on the language. In some languages, the output of **θr* is identical to that of **θ* and **r* individually; this particularly applies to *hr, sr* and *x̌(V)r*, which are found in languages that also change **θ* to *h, s* and *x̌*, respectively. The same is true for Western Yaghnobi, which shows **θ > t, *θr > t(V)r*, maybe also in Wakhi (*θ > ?, θr > tr*), but not so for Sogdian where **θ* is preserved, and **θr* yields *θr, rθ* and *š* word-internally, but *tr-, š-* and never ***θr-* initially.

In others, though, a merger to a different consonant such as *š* or *c* occurs; this is reminiscent of Old Persian, which shows a change of **θr* to a sound conventionally transcribed *ç*.

The word-initial and word-internal outcomes are mostly parallel in each language, although the use of epenthetic vowels may differ due to syllable structure constraints. But some languages show a marked contrast. Bactrian has several cases of *hr* (as one might expect from a combination of **θ* and **r*); the more frequent outcome *r* is undoubtedly a later development, probably independent from the outcome *r* in other languages.

States:

id	feature	weight	0	1	2	3	4	5	6	7	8	9	10
p8	*θr-	1	**θr**	t(V)r	(V)rt	tr, š	dr	c	š	x̌(V)r	s(V)r	h(V)r	(V)r
p9	*-θr-	1	**θr**	tr	r(V)t	θr, rθ, š	c	š	s(V)r	r(V)s	hr	r	

Directionality and costs: The lenitions and reductions follow the patterns outlines above (esp. 1.**4.5.-6.**). We consider metathesis possible in both directions

(*C(V)r* <> *r(V)C*) and have assigned cost 2 to **θr, tr > rt*, 3 to most other and 4 or more to methases in combination with other phonemic changes. The Sogdian outcome *tr, š* can be result of many different preceding steps, but once original **θr* has merged with *tr* and *š*, the original distribution can never be recovered.

2.3. Morphological characters

2.3.1. The plural of nouns (Wendtland §3.1)

Concerning the plural forms, both the case system and the morphemes employed to express them exhibit variation. These parameters do not necessarily depend on each other, but they cannot be seen in isolation either, as the choice of a specific inherited formant hinges on the case system at the time of this innovation.

While the more archaic EIr. languages show several case forms in the singular and plural, many other varieties have undergone a merger to a two-case-system (direct vs. oblique case) or lost case distinctions altogether. In both instances, the endings may either be inherited or innovated. In two-case-systems, inherited endings of the direct case usually go back to the nominative or nominative-accusative (nominal or pronominal), while the oblique plural may be a continuation of the genitive (**-ānām*) or of the cases containing **-b-* (esp. DAT-ABL.PL **-bi̯ah*). Languages without case distinction have generalised the oblique just mentioned as plural suffix or have innovated entirely new suffixes.

In addition, there are several novel formations, using abstract or collective suffixes (**-tā*) as plural markers. These can function as case suffix (i.e. as NOM/DIR.PL, in a paradigm with an inherited OBL.PL) or as plural marker (*-t-* with plural case endings attached to it). In the latter case, there may be an agglutinative system (*-t-* with SG endings suffixes).

The innovations within the inherited system and the novel formations are treated as several isoglosses to avoid largescale polymorphy. This is unfortunate because the innovations tie into the restructuring of the inherited case system. However, as proven by Sogdian, which has a complex plural system,[16] the

[16] The Sogdian system is very complex. In the Ancient Letters, some words – "light" and "heavy" stems alike – preserve plural endings (e.g. *β'γ* 'god': NOM.PL *β'γ'* < **-ā*, OBL.PL. *β'γ'nw* < **-anu* < GEN.PL**-ānam*), and preserve plural morphosyntax (e.g. *β'γ' 'γr'nt'nt* 'the gods were angry'). Other words instead form a "collective" which is semantically plural but morphosyntactically F.SG (Sims-Williams 2021). The Ancient Letters coincidentally only attest "heavy" stem collective nouns (NOM-ACC *-t*, GEN-LOC-ABL *-ty*), but the "light" stems (NOM-ACC *-t'* < **-tā*, GEN-LOC-ABL *-ty'* < **-tyā* < **-ah+tāi̯ā(h)*) most likely existed as well (Sims-Williams 2021: 176). The "light" and "heavy" stem collectives survive in later Sogdian as the regular plural (Yoshida 2009: 288–9). All original plural endings are eventually lost (Nicholas Sims-Williams, p.c.). In later Sogdian, the "light" stems

innovations were gradual, and multiple novel forms could co-exist with the inherited system. Therefore, it is not unjustifiable to split the isoglosses into several characters. We have split the plural suffixes into four characters: **m1** describes the innovations in the inherited case system, **m2–3** the grammaticalization of the collective as an (additional) plural, and **m4** the introduction of the plural marker **-išt-*.

2.3.2. m1: the innovations of inherited nominal plural system

PIr. had multiple endings across the different ablauting stem classes. For this study, the following are the most relevant: NOM **-a̯iah*, GEN **-ānām* and **-īnām*, DAT-ABL **-bi̯ah*, INSTR **-biš.*[17]

The Khotanese case system is copious and complex and can only derived directly from the ancestral state. Khotanese is the only EIr. language to continue a NOM-ACC system (albeit only in the singular). In the plural, it continues the NOM **-ai̯ah* as *-ä, -a* (etc.), and both variants **-ānā̆m* and **-īnā̆m* of the GEN as GEN-DAT *-anu*, *-änu* among the multiple oblique cases (Gercenberg 1981: 251; Skjærvø 2022: 125–6). At present, we take no stance on whether the short vowel of the final syllable is in fact an archaism of PIE age. The preservation of the inherited endings as part of the LOC and the analysis of *-yau* (INSTR-ABL.PL and VOC.PL) as continuation of **-b-* is especially important since it tells that Khotanese cannot be derived from a more reduced system.[18]

The other Middle and New EIr. languages have reduced the number of cases further. The Sogdian plurals of the Ancient Letters and the "light stems" preserve the most case distinctions. NOM **-ai̯ah* is continued as the vocalic ending of the NOM-ACC (DIR). Similar to Khotanese, GEN **-ānām* and **-īnām* are preserved as an oblique case. There have included traces of another oblique case, namely the equation of VOC.PL *-β* in *xwt'ynβ* 'sirs!' (< **xu̯atāu̯anaibis*) and Khot. *-yau* as the old INSTR.PL in vocative function (Sims-Williams 2023: 34; 1991: 183).

(deriving from the collective in **-tā*) still behave as F.SG and take the verb in the 3SG though it is difficult to recognise because they are usually attested in the "ergative construction". The "heavy" stems have become agglutinative (DIR.PL *-t* < **-t-Ø*, OBL.PL**-tī* < **-t-ī*) (Sims-Williams 1982). Additionally, animate "light" stems show the plural suffix **-išt-* (Sims-Williams 1989b: 183). See also Kim (2025: 4)

[17] In what follows, vocative forms and a few additional archaic endings specific to the inflexion of family terms have not been taken into account. See also Skjærvø (2017) and Kümmel (2018) for a fuller picture.

[18] It is debated exactly what this ending goes back to, but all proposals contain an equivalent of PIr. **-b-* from the DAT-ABL or INSTR – or a contamination of both (Del Tomba 2024: 261, n. 23; Skjærvø 2022: 125; Kim 2024: 9–19, esp. 91–2).

The Choresmian two-case system is similar to the Sogdian "light stems" (DIR from NOM and *-*īnām*/-*ānām* as OBL), but there is no trace of an extra oblique case in *-*b*- (Ėdel'man 2008: 28). The same applies to the Bactrian inscriptions, but *-*īnām* has been lost. DIR -*ε* (continuing the NOM) is distinguished from OBL -*ανο* (continuing the GEN *-*ānām*). This is given up in the manuscripts where case is lost, and -*ανο* becomes the sole plural marker (Gholami 2009: 133; Sims-Williams 2007: 40).

Pashto has a two-case system, but the formants are more complex than in most other EIr. languages because of stem classes and ablaut. The endings of the OBL go back to *-*ānām, *-īnām* and *-*b* across different stem classes.

Munji and Yidgha show vocalic endings in the DIR (our *-*ai̯ah* vel sim.), and OBL -*āf* which goes back to *-*b*-. They do not preserve the genitive plural.

Shughni does not have case distinctions, but the plural ending of Shughni (Xufi and Roshani) -*ēn* go back to the GEN.PL Notably, the closely related Roshorvi has -*īf* (< *-*b*-) in its place (Payne 1989: 428). We have included both endings as Shughni.

The origin of the Ishkashimi plural ending -*o* is unclear. Since **b* yields *v* (Skjærvø 1989a: 377), we have coded it is a descendant of the *-*b*-cases. It is clear from the fact that it surfaces as a consistently accented -*ó* that it cannot do so directly.[19]

Ormuri has also lost cases. It continues the old nominative as well a both variants *-*ānām* and *-*inām* as plurals. It is unclear if the Ormuri ending -*ān* is inherited or borrowed from Pashto or Persian (Kieffer 1989: 450). The ending -*in*, on the other hand, is attached to inherited words (Kieffer 2003: 101). Even in Pashto, -*ān* could be of Persian origin. For the sake of simplicity, since the endings *could* be inherited, we have treated them as such. Parachi also loses case distinction and continues *-*ānām* > -*an* as the only inherited plural. The ending -*ā́* seen next to it is a borrowing from Persian (Kieffer 1989: 450).

In the remaining languages, the plural systems are made up of a mixture of inherited and innovated endings. Since we do not want to count the same innovations twice, formants coded elsewhere are mentioned here but ignored under this character.

Yazghulami preserves the *-*ānām* as the OBL but innovates a new direct case from the collective in *-*tā*. This is innovation treated under **m2** below. Wakhi -*əv* and Sarikoli -*ef* both continue *-*b*- and function as the oblique case marker, but the languages have innovated new direct endings. Only the inherited endings are considered here. Since the enigmatic Sarikoli DIR.PL -*xεyl* is not shared with any other language, it can be left out without consequences in the analysis. Wakhi shares -*išt* with the Sogdian animate light stems (this is coded under **m4**).

[19] According to Paxalina (1987b: 502), it goes back to a collective marker *-*ăha* (which she also sees in the sister-dialect Sanglechi -*āy*).

The case systems of Ossetic and Yaghnobi cannot be directly derived from innovations within the inherited case endings of the plural since they have become agglutinative and add singular case endings to the plural suffix (Kim 2025: 5). Therefore, they are assigned the state “?”. The innovations in their systems are coded under other the following characters.

States:

It is important to stress that features in parentheses are accounted for elsewhere and have not influenced the weighting of the costs concerning the state they apply to. They are only mentioned for the sake of completion of this overview.

Id	**feature**	**weight**	**0**		**1**
m1	nominal plural endings	2	NOM-ACC *-ayah, GEN *-ānām, *-īnām; DAT *-byah, INSTR *-biš, LOC *-su, *-šu, stem classes and ablaut.		NOM *-ayah vel sim., GEN *-ānām,*-īnām; OBL *-b-, stem classes and ablaut
	2	**3**	**4**	**5**	**6**
	DIR *-ayah; OBL *-ānām, *-īnam,*-b-; stem classes and ablaut	DIR *-ayah; OBL *-ānām, *-īnam	DIR *-ayah, OBL *-ānām	OBL *-ānām, (neo-NOM)	PL *-ayah, *-ānām, īnām; stem classes and ablaut
	7	**8**	**9**	**10**	**11**
	PL *-ānām	PL *-ānam, *-b-	DIR *-ayah, OBL *-b-	OBL *-b-, (neo-NOM)	PL *-b-

Directionality and costs: Once the cases have merged in function, and once endings have been lost, they cannot resurface in their original distribution. The loss of a single formant in the same function is assigned cost 1. Restructuring the case system (from multi-case to two-case or to-case to no case at all) is assigned cost 3. Reductions requiring restructuring of the case system and massive loss of endings are assigned cost 4, and highly unlikely reductions whereby a single form is preserved in a radically different function are assigned cost 5. The change from 9 to 10 has been assigned the unusual cost *0*, because the oblique endings is kept exactly the same, and the different innovations in the direct case are coded in **m4**.

2.3.3. m2-4: the "collective" as a plural

In a number of Ir. languages, inherited PL endings have been replaced by a new PL marking, which originates in abstract suffixes (*-*tā*) (Gershevitsch 1961: 177–80).[20]

It even proved impossible to describe the addition of the new plural (originally collective) as a single character if we want to capture all potentially shared innovations. Therefore, we have split this complex in two characters: The first being the overall grammaticalization of the abstract suffix *-*tā*- as a collective and subsequent plural, the second being the innovation of a specific paradigm shared by the Sogdian "heavy stems" and Yaghnobi.

m2 encodes the innovation of a plural from the collective. The Sogdian Ancient Letters attest an early stage where the collective in -*tā* is still morphosyntactically F.SG but semantically identical to the plural, and "thus the process whereby the collective was to take over the function of the plural had already begun" (Sims-Williams 2021: 185). The Sogdian "light stems" only show the plural semantics. We use Sogdian as a proxy for the shared state.

From here, we derive too further innovations. Yaghnobi and Ossetic (and the Sogdian heavy stems) use the newly formed suffix *-*t*- as an agglutinative plural marker and add singular case endings to it (Bielmeier 1989: 483; Thordarson 1989: 469). Yazghulami instead incorporates the newly formed plural suffix *-*t*- as the DIR.PL -*aθ* into the existing case system (with OBL.PL -*an* < *-*ānām*).

m3 encodes the innovation of the paradigm DIR.PL *-*t,* OBL.PL*-*ti* specific to the Sogdian "heavy stems" and Yaghnobi.

Presumably, **m2** and **m3** are dependent on each other. This violates a key principle of character-based cladistics. However, treating them as one character would cause Sogdian to be polymorphic. Assigning all states (the ancestral system of the Ancient Letters, the archaic system of the "light stems" and the system of the "heavy stems" shared with Yaghnobi) simultaneously in order to incorporate as much nuance as possible would essentially render Sogdian irrelevant in the analysis since polymorphic states are automatically replaced. In order to keep the information – and to control the allowed directionalities – we have chosen to reformulate this polymorphism into several characters where monomorphic states represent the polymorphic outcomes, i.e. single digits represent multiple morphs. This does mean that a shared state inferred by the software would have to be manually validated afterwards. To repair the damage inflected by doubling the character, we have downweighed **m3** to weight 1.

[20] We wish to thank Nicholas Sims-Williams and Antje Wendtland for their elaborations on these matters.

Finally, **m4** encapsulates the innovation of the enigmatic plural suffix **-išt-* which serves as the DIR.PL *-išt* in Wakhi, and as the base of the DIR.PL *-yšt* and OBL.PL *-yšty* of the animate Sogdian "light" stems. The unique and therefore parsimony uninformative Sarikoli ending *-xεyl* is included here, so that assigning a cost to the change of state of 9 > 10 under **m1** does not force the innovation of **-išt* to be counted twice.

States:

id	Feature	weight	0	1	2	3
m2	PL < abstr. *-tā-	2	no	*-tā F.SG (coll.) in PL meaning	-t- as agg. PL suffix	-t as NOM.PL
m3	PL DIR -t, OBL -ti	1	no	yes		
m4	Neo-plural	2	no	*-išt-	-xεyl	

Directionality and costs: for m2 the innovation itself is assigned cost 3. States 2 and three are only derivable through state 1 and are also assigned cost 3, as are changes between states 2 and 3. Returning to state 0 is difficult is more difficult – if it happened at a stage like archaic or later Sogdian, it would definitely be possible. It is, however, difficult to imagine a language giving up its plural system in entirety. Since these characters cooccur with inherited states of **m1**, we have assigned cost 1 to all reversion to 0. Taking part in the innovation of m3 is assigned cost 3. Losing the paradigm again is assigned cost 1 (with the caveat that this should be validated against the other plural characters). Taking part in the innovation of m4 is assigned cost 5 because it is etymologically unclear and the functional pressure to replace the plural endings is very low.

2.3.4. m5: The verbal ending of the 3rd plural (Wendtland §3.2)

In most East Ir. languages, the 3rd plural verbal ending continues the active ending (**-Vnti*), which is often reduced to *-n(d)*, sometimes the *n* is lost. Some languages have preserved both endings or have generalised the middle ending (**-Vr*) as the 3PL.

Khotanese preserves the distinction in diathesis just like in Proto- and Old Iranian. Yaghnobi *-or* and Chorasmian *-ri* continue the old middle ending, but function as a plural ending regardless of diathesis. Ossetic *-uncæ* and *-ync* are likely to derive from **-nθ-*, with **θ* taken over from the 2PL (Kim 2023: 160–3). This state would be unique to Ossetic, but it could only derive from a generalisation of **-nti*. Since this prestage does not occur in any of the languages, and we cannot add unattested intermediate stages, we have assigned Ossetic state 2 (the generalisation of the active ending) (see **1.4.5.**).

The reason for including the various phonetically reduced results of *-*nti* is that we cannot exclude that some outcomes could be reflexes of word-final *-*r* as well.

States:

id	feature	weight	0	1	2	3	4	5
m5	inherited 3PL-endings	2	ACT *-nt(i) + MID *-r	3PL *-r (only)	3PL *-nti (only)	-nd	-n	-t

Directionality and costs: Generalising the middle is slightly less obvious (cost 3) than the active (cost 1) as the only 3PL ending. The phonetic reductions and potential rhotacism follow the principles outlined above. Once the endings – and the functions – have merged, the original distribution cannot recur.

2.3.5. m6: The merger of the verbal 3SG and 3PL endings (Wendtland §3.2)

Pashto and Ormuri also show a vocalic ending in the 3PL, although both languages preserve -*n*. To resolve this polymorphy (see **1.4.4.)**, we have treated this ending as a separate character. The vowel quality is likely to be analogical to the 3SG (Efimov 2011: 185). According to Kieffer (1989: 450), the Ormuri ending is likely to have been adjusted to the 3SG on the model of Pashto. In Pashto, the regular ending is -*i* in the third person regardless of number, but Wanechi preserves a distinction 3SG -*i*, 3PL -*in*. These may go back to *-*ai̯ati* and *-*ai̯anti*, respectively (Grjunberg & Ėdel'man 1987: 93).[21] We have included this feature as a potential shared innovation that could have started in a common prestage of Pashto and Ormuri.

States:

id	feature	weight	0	1
m6	analogical 3PL ending	2	inherited 3PL ending	3PL = 3SG

Directionality and costs: Once the endings have merged, either phonetically or due to a morphological innovation, they cannot find their original distribution. The innovation has been assigned cost 3. because it was probably phonetically motivated.

[21] Morgenstierne (1942: 105) notes the archaic and poetic Pashto form *kāndi'ə* 'does; they do' which is functionally 3SG and 3PL and "evidently" goes back to 3PL **kāṇnde* < **kr̥nantai*. This is further evidence for an earlier functional merger of the 3SG and 3PL.

2.3.6. m7: Pronoun 2nd plural (Wendtland §3.3)

The 2PL pronoun shows several innovations. Most languages restructure the original paradigm in interesting ways, but the most remarkable feature is the "specific formation" of a 2PL pronoun based on the stem of the 2SG shared by Bactrian and some Pamir languages. This innovation may be motivated by a sound change **šm* > **m* which caused the 2PL and the 1PL to fall together. If **šm* became **m*, the 2PL **šmāx* (< PIr. **šmāka-*, cf. Av. GEN *xšmākəm*) would be identical to 1PL **māx* (< PIr. **ahmāk-*, cf. YAv GEN *ahmākəm*, OPers. *amāxam*) (Korn 2016: 416f.; 2019: 266; Skjærvø 2017: 524–5).

Reformulating this isogloss into a character requires some effort. It is crucial that the states described below aim for the latest potential common prestage, and many languages have smoothed out the paradigm individually afterwards. Additionally, we do not reconstruct the development of the pronominal paradigm as a random generalisation of single oblique case form, but as a gradual restructuring of the case system. The states show what must have been preserved.

In the Khotanese paradigm, nom. *uhu* probably derives from the NOM **i̯ūžam*, and the OBL enclitic is preserved as *ū* (Emmerick 1989: 220).

Other languages preserve one oblique case form as the dir. next to an enclitic form. The preserved case form is not always the same: Yidgha and Munji continue the old DAT **i̯ušmabi̯a* as the non-attributive case (Munji *mof*, Yidgha *mŏ̌f, mă̌f*), and Munji also attests a continuation of the enclitic (which Yidgha does in other persons) (Skjærvø 1989b: 414). Ossetic (*symax, sumax*), Sogdian (*'šm'x(w)*) and Yaghnobi (*šumox*) generalise the GEN **šmāxam* as the DIR. Parachi *wå* continues the enclitic **=u̯ah* (cf. Av. *vå̄*)[22] – the reason for preserving only the enclitic form could have been the coalescence of **šm* and **m* (Morgenstierne 1938: 62).

The following states are assigned to the languages that share the innovation of a new paradigm or form based on the 2SG (orth. **tū/tau̯a-*, encl. **-tu̯ā*) motivated by the merger of the 2PL and 1PL by the sound change **šm* > **m*. Because we cannot force our states through unattested intermediate states (see **1.4.5.** and cf. **2.3.4.**), the otherwise isolated state exhibited by Chorasmian is used as the proxy for the common prestage. The Choresmian form *hβy* is based on the 2SG clitic (*βy* <**-tu̯ā*) and an unknown particle *h-* (Wendtland 2009: 180; Ėdel'man 2008: 36; Bogoljubov 1962: 12–13). Chorasmian cannot have taken part in the subsequent evolutions shared by the following languages.

[22] Synchronically, this is also continued in the enclitic personal pronoun *-ó(w)* < **-a-wah* (Morgenstierne 1929: 63; Kieffer 1978: 90).

Pashto *tāse* must be derived from orthotonic 2SG pronoun **tū/taṷa* and an unclear particle *-se* (Korn 2016: 416; 2019: 266; Wendtland 2009: 180, nn. 71–2). Since the particle is not shared by other languages, but the innovation of basing the paradigm specifically on the orthonotic 2SG is, the state exhibited by Pashto is used as a proxy for this innovation.

Bactrian and most Pamir languages share an innovation based on the former proxy state, namely a form that looks like a combination of the orthotone 2SG pronoun and the 1PL pronoun: Bactian *τωμαχο*, Shughni *tama*, Sarikoli *tamaš*, Ishkashimi *tьmьx*, Yazghulami *təmox* < **taṷa-māx* < **taṷa+ahmākam* (Sims-Williams 2007: 271, 268; Payne 1989: 423).

Finally, Ormuri and Wakhi are isolated and assigned unique states: Ormuri *tyōs* is probably borrowed from Pashto (Morgenstierne 2003: 84), and Wakhi *sa(y)-iš(t)* is unclear (Steblin-Kamenskij 1999: 310).

States:

id	feature	weight	0	1	2	3	4
m7	pronoun 2PL	3	NOM *yūžam, GEN *šmāxam, DAT *yušmabya, clitic *⸗wah	*yūžam, *⸗wah	*yušmabya, *⸗wah	*šmāxam, *⸗wah	*⸗wah
			5	**6**	**7**	**8**	**9**
			2PL *šmāxam > = 1PL *māxam, replaced by form based on 2SG	2PL based on 2SG orthotone *tū/tawa	2SG orthotone + 1PL	probably LW	unclear etym.

Directionality and costs: Once lost, a form cannot reemerge. Transitions into the unique states 8 and 9 have been assigned a cost 5 – not because we deem loanwords improbable everywhere, but because we do not want them to interfere with the configuration of the tree. Their impact is neutralised by the equal cost: they fit equally bad on every tree, and since it is impossible to return from the Wakhi and Ormuri states to states exhibited by other languages, the innovation is forced towards the leaves. We acknowledge that replacing a single form with a single borrowed form is more likely than losing an entire inflectional paradigm in favour of one borrowing, but in order not to give too much leeway for the "unknown unknowns", we have neutralised the costs this way.

Generalising just the clitic form is not trivial (cost 2 from a paradigm with two forms and cost 4 from the root). Because getting from the root to states 5, 6 and 7 requires many intermediate steps, and we only operate with five differentiated costs, we have assigned the costs so that 0>3>4>5>6>7 is comparable to 0>7 directly. This is not ideal, but the alternative would have been a much-less fine-tuned grading of the innovation.

2.3.7. m8: Deictic demonstrative pronouns (Wendtland §3.4)

Instead of only two deictic demonstrative pronouns (‘this’ vs. ‘that’), several East Ir. languages have three. The stems employed are not identical, though. The matter is complicated by the fact that the paradigms are suppletive in Old and most of Middle Iranian. Later languages have abandoned the suppletivity and/or rearranged the stems into new praradigms.

The innovations can be viewed as recombinations of the following four demonstrative pronouns. We refer to them according to their deictic functions: D0 for the most unmarked one, D1-3 according to their personal deixis.

Figure 4: The PIr. stems of the demonstrative pronouns

	DIR	OBL
D0	**ha-*	**ta-*
D1	**ai̯a-*	**a-/ima-*
D2	**aiša-*	**aita-*
D3	**hau̯-*	**au̯a-*

Khotanese has built entirely new pronouns based on the stems *ttā-/ttu-* < **ta-* (D0 OBL) and *ṣä-/ṣā-* < **aiša-* (D2 DIR). The function of the deictic pronouns in Bactrian is not clear (Gholami 2014: 128). What is clear is that the demonstrative pronouns continue at least two stems: *το* and *τι* go back to the **ta-* (D0 OBL), and both stems of D1 are continued in *ειο* and *ειμο*. It is debated what *ειδο* represents. We follow Sims-Williams (2007: 210) in interpreting *ειδο* as **aita-* (D2 OBL). Wendlandt, on the other hand, thinks that it goes back to a compound of the two other preserved stems, **ai̯a-* (D1 DIR) and **ta-* (D0 OBL). Since Bactrian does not share either distribution with another language in this study, the coding would not effect the results much – but preserving more stems of course pushes Bactrian up the tree.

Ormuri near-deictic *a-* continues **ha-* (D0 DIR), and the first element of *a-fō* ‘that’ has been traced back to **au̯a-* (D3 OBL), possibly contaminated with **ha-* and/or- **au̯a-* (Efimov 1991: 292–3). Because the Kaniguram dialect also has *hō* ‘that’ probably from **hau̯-* (D3 DIR), we have included both stems of D3 in the coding. Pashto *ha-γa* and *dā* go back to **ha-* (D0 DIR) and **aita-* (D2 OBL).

The Sogdian three-term deictic system continues the both stems of D1, D2 and D3: *yw, ʾmw; šw, ʾtw; (ʾ)xw, ʾw(w)* (Sims-Williams 1994; Wendtland 2006). This system could be the predecessor of the following 3- and 2-term deictic systems (*pace* Skjærvø 1989a: 375–6).[23]

In Shughni, both stems of D3 are attested as *yu, yā* and *wi, wam*, and **ai̯a-* (D1 OBL) and **aita* (D2 OBL) as *yam, mi; yid, di*. Similarly, Wakhi, Ishkashimi, Munji, Yidgha and Sarikoli all continue only the oblique stem **a/ima-*, **aita-* and **awa-* (D1, D2 and D3). The forms can be seen in Payne (1989: 430).

In Chorasmian, the forms have been rebuilt and merged with largely unknown prefixes. However, the direct stems **ai̯a-*, **aiša-* and **haw-* (D1, D2 and D3) and **aita-* (D2 OBL) are visible underneath in what is synchronically a system of two deictic pronouns and one neutral one (Durkin-Meisterernst 2009: 344; Ėdel'man 2008: 36–7).

The remaining languages have recombined the inherited stems and formed novel paradigms of two-term deictic systems.

In Yazghulami, the stems *du* and *way* go back to **aita-* (D2 OBL) and **au̯a-* (D3 OBL). A last form, *yu*, is debated. Ėdel'man derives it from **aiša-* (D2 DIR). We have accepted Wendtland's (2009: 180) analysis, deriving it from **hau̯-* (D3 DIR), since a mixture of the deictic stems into one paradigm would be less likely (but not impossible).

Parachi and Ossetic have two-term deictic systems based on D1 and D3. Parachi *(h)ō* is traced back to **hau̯-* (D3 DIR), and *(h)ē* presumably to the genitive of the stem **a-* (D1 OBL), *hē* < **a-si̯a-* (cf. OAv. *ahiiā,* YAv. *ahē*) (Wendtland 2009: 182). Both dialects of Ossetic continue the same stem as the proximate *a-* 'this' (D1 OBL). In Digor, the distal pronoun is *ie* (< **ai̯am-* D1 DIR) in the nominative and *uo-* (**au̯a-*, D3 OBL) in the oblique cases, whereas Iron *u-* 'that' (D3) is used for both, and it either continues **hau̯-* (D3 DIR) directly or as a conflation of **hau̯-* and **au̯a-* (D3 OBL) (Thordarson 1989: 472). In either case, both stems must have been present in Proto-Ossetic. Thordarson even points out that the use of the different lexical elements in the two dialects could point to an earlier stage with three deictic stems.

Finally, Yaghnobi continues both stems of D2 and D3: **aiša-*, **aita-* > *iš, it*; **hau̯-*, **au̯a-* > *ax*, *aw*.

[23] By entering the languages as distinct leaves and not forcing ancestry, our study takes for granted that none of the languages descend directly from each other. However, their latest common ancestor may be identical (in terms of states) to an attested language. Wendtland (2009: 180; 2011: 353–4) remarks that the – admittedly rare – 2sg-deictic pronoun (*-*š*-/-*t*-) disappears in the later Sogdian language – unlike in Yaghnobi, where it is preserved.

States:

Id	feature	weight	0	1	2
m8	**deictic pronouns**	3	4 suppl. stems: *ha- *ta- *aya- *a/ima- *aiša- *aita- *haw- *awa-	New paradigms Ø ttā-, ttu- Ø Ø ṣä-/ṣā- Ø Ø Ø	Unclear Ø το, τι ειο ειμο Ø ειδο? Ø Ø

3	4	5	6	7
2-term deixis a- Ø Ø Ø Ø Ø hō afō	2-term deixis ha-γa Ø Ø Ø Ø dā Ø Ø	3-term deixis Ø Ø yw ʾmw, ʾmn šw ʾtw (ʾ)xw ʾw(w)	3-term deixis Ø Ø Ø(yu?) yam, mi Ø yid, di yu, yā; wi, wam	3-term deixis Ø Ø Ø *a/ima- Ø *aita- Ø *awa-

8	9	10	11	12
2-term deixis Ø Ø ie- (D) a- Ø Ø u- (I) u- (I?), uo- (D)	2-term deixis Ø Ø Ø h(ē) Ø Ø h(ō) Ø	2-term deixis Ø Ø Ø (yu?) Ø Ø du yu way	3-term deixis (?) Ø Ø nyn, nʾn; hy Ø nyš nyd nʾw Ø	2-term deixis Ø Ø Ø Ø iš it ax aw

Directionality and costs: As for directionality, the key principle is that a language obtain stems once it has lost them. We have kept all other logical possibilities open: As long as the stems are attested, we have deemed it possible to reshuffle them into all attested deictic systems. The loss of a single stem within the system is assigned cost 1, the loss of one deictic degree but retention of most stems is assigned cost 2, reduction from 3 to two deictic stems is assigned cost 3, unlikely and asymmetrical – but technically possible – reductions are assigned cost 4, and highly unlikely rapid reductions are assigned cost 5. Further, we have taken the potential combinatory costs into account and deducted a point in cases where multiple steps is reconstructible, so that a reduction from state 0 is not automatically more economically attractive compared to a more well-defined gradual restructuring.

We realise that this coding may be unfortunate since we cannot exclude that one or more of the lost stems survives outside the deictic system (e.g. as an article).

2.3.8. m9-12: Prefixed pronouns 1st/2nd person (Wendtland §3.5)

Some Eir. Languages innovate a set of prefixed pronouns of the first and second person. They are formed with prepositions and special forms of the personal pronouns. The fact that also some Western Iranian languages have similar prefixed pronouns (Korn 2019: 259–60) does not make it impossible that some of the Eir. languages innovated this trait.

Although the isogloss as it is presented by Wendtland concerns fused forms of prepositions and personal pronouns, it includes forms that are functionally and formally distinct. In terms of potentially shared innovations, the most important axis is not just *if* the forms are fused, but also *how* they are fused. The PIr. state, also preserved in many daughter languages, is the free combination of the adpostion and the orthotonic personal pronouns. As noted by Wendtland, some fused forms only occur in the SG (Sogd., Chor.) or PL (Munji). Bactrian attests both numbers, but different combinations of clitics, prefixes and persons. This is disregarded as the data from historical languages could be incomplete. It seems that the forms mostly occur with the same set of prepositions (**hača* 'from', **hada* 'with', **abi* 'on', **upari* 'for'), we have treated these as different characters. We recognise that the innovation possibly occurred with one preposition first, and that the other followed suit, and have weighted them accordingly.[24]

The main formal distinction is whether it is the preposition or the pronoun that is reduced. In Sogdian and Bactrian (1PL; 2PL unattested), the preposition is regularly elided, and it is combined with clitic pronouns (e.g. Sogd. 1SG *c'm'* < **hača-mā* and Bactr. 1PL *ασ-αμαχο* < **hača-ahmāxam*). In Sogdian (optionally), Chorasmian (always, when prefixed, but not suffixed) and in Bactrian (only SG), a new pair of clitic personal pronouns only occurring when fused to prepositions is innovated: 1SG **hača-mā(d)-ka* > Sogd. *c'm'k(H)*, Chor. *c-m(y)k*, Bactr. *ασα/ο-μαγο*, 2SG **hača-θu̯ā(d)-ka* > Sogd. *c'f'k(H)*, Chor. *c-fyk*, Bactr. *ασ(α/ο)-φαγο* (Gholami 2014: 102).

Conversely, the preposition can be elided into a prefix and combined with the orthotonic pronoun, e.g. Munji 2PL *žāmof* < **(ha)č(a)-i̯ušmābi̯a*. The long vowel *ā* of the fused prepositions might be analogical from other prepositional compounds, e.g. OP *hačāma* (Sims-Williams 2011b: 30 n. 15). Under the assumption that the fusion happened before the paradigmatic restructuring of the 2PL pronoun (see **2.3.6.**), this could be formally – but not functionally – identical to possessive pronouns of Pashto (e.g. 2PL.POSS *stāse* < **(ha)č(a)-tū(u̯a)-se*) and the Yazghulami direct

[24] Munji and Chorasmian seem to have spread the innovations to other propositions independently, e.g. Munji *nāmox* < **ana-*, Chor. *pš-myc* < **pati-*).

pronominal objects (e.g. 2PL.ACC *š=təmox* < *(*ha*)*č*(*a*)-*tau̯a-šmākam*) (Morgenstierne 2003: s.v.; Rastorgueva & Ėdel'man 2007: 302; Jamison 2022: 32).[25]

Fused forms with **hača* have the widest distribution both in terms of languages, states and functions: Other than the two lines of innovation outlined above, it also becomes the case-like postposition *-jsa* in Khotanese.

The etymology of many prepostitions in the Pamir languages is debated. Wakhi *də/tə* could go back to **hada*, but it is possibly contaminated with **antara* (Steblin-Kamenskij 1999: 153).[26] The same could, perhaps, go for Munji *da*; although **d* usually yields *l*, influence from **antara* could have resulted in *da*. Accordingly, we have included Munji *dāmox, dāmof* as descendants of preclitic **hada*-. Similarly, "it is probable that the marker *va*, *vo* (etc.) in Yidgha-Munji and Sanglechi also derives from **abi* rather than from **upa* as generally assumed" (Sims-Williams 2011b: 30). Therefore, we also treat Munji *vāmox, vāmof* with the other descendant of **abi.* We thus arrive at a slightly different point of departure than Wendtland's (2009:182-3), see Figure 5.[27]

[25] According to Ėdel'man (1990: 234), there are similar fused forms in Shughni. However, these forms are not found in (Edelman & Dodykhudoeva 2009a: 781–2; 2009b: 794–5; Skjærvø 1989a: 380; Payne 1989: 430–3). Ėdel'man's comment might refer to another language in the Shughni-group. We are grateful to Jaroslava Obrtelova for sorting out the Shugni details for us. In most Pamir languages, **hača*- functions as an object marker, but it is not specifically fused with personal pronouns (Payne 1989: 434).

[26] Wakhi, *pər* is also ambiguous and could be the descendant of **para* and/or **upari*, and *pə* could go back to **pati*, perhaps contaminated with **upa* (Steblin-Kamenskij 1999; Ėdel'man 1990: 242).

[27] The table and examples can be supplemented with additional forms from (Gholami 2014: 102–3; Sims-Williams 2007: 191, 229, 273; Vinogradova 2000: 82; Livšic & Xromov 1981: 455; Humbach 1989: 198; MacKenzie 1990: 111; Ėdel'man 2000: 99–100; Durkin-Meisterernst 2009: 345).

Figure 5: Prefixed personal pronouns

ADP	**Sogdian**		**Chorasmian**		**Bactrian**		**Munji**	
	1p.	2p.	1p.	2p.	1p.	2p.	1p.	2p.
hača								
SG	*c'm'(k(H))*	*c'fk(H)*	*c-m(y)k m'-c(y)*	*tw'-c, -β/f-c, c-fyk*	ασα/ο-μαγο	ασ(α/ο)-φαγο	–	–
PL	–	–	*mn'-c*	–	ασ-αμαχο	–	*žā-mox*	*žā-mof*
hada								
SG	*δ'm'(k)*	*δ'f*	*-m'-θ*	*-f-θ*	αλ(α/ο)-μαγο,	αλ(α/ο)-φαγο	–	–
PL	–	–	–	–	–	–	*dā-mox*	*dā-mof*
abi								
SG	Ø	Ø	*f-myk*	*f-fyk*	αβα/ο-μαγο	αβ(α/ο)-φαγο	–	–
PL	Ø	Ø	–	–	αβ-αμαχο	–	*vā-mox*	*vā-mof*
upari								
SG	*pr'm'k*	*pr'β'k*	*pr-m(y)k*	*pr-fyk*	Ø	Ø	Ø	Ø
PL	–	–	–	–	Ø	Ø	Ø	Ø

In order not to inflate the influence of this isogloss disproportionately, we have downweighted codependent characters. Fused forms with **hača*-, being the potential cradle, is weighted 3, and the three subsequent shared prepositional fusions are weighted only 1.

The states are assigned as follows, using the more complex **hača* as example. To avoid polymorphy, we have assigned one state to Chorasmian, Bactrian and Sogdian who fuse the preposition with a (mostly) clitic pronoun and form the new series. The exact distribution and combinations of types of fused forms across persons and numbers is not exactly clear. The other innovative route of Yazghulami, Munji and Pashto whereby the preposition is fused with an orthotonic pronoun is assigned another state (the unique innovations of Pashto and Shughni can safely be ignored for the purpose of the configuration of the tree). Khotanese has its own state. Languages continuing the preposition as a simplex in combination with orthotonic pronouns share the state *0* with PIr., and languages in which the preposition is lost share a state.

Despite our best efforts it is entirely possible that we have missed an attestation of an adposition somewhere or misinterpreted an etymology.[28] Since all four prepositions are reconstructed for PIr., and loss is not weighted much, this will have no significant impact on the analysis. At any rate, the existence of the prepositions in question is also guaranteed by multiple attestations as preverbs.

States:

Id	**feature**	**weight**	**0**	**1**	**2**	**3**	**4**
m9	*hača+ pron. 1/2 pers.	2	*hača+ orth. pers.pron.	*hača+ clitic pron.; *hača-mā-ka-	*(ha)č(a) as procl./pref. +*orth. pron.	*hača suffix-like postp.	∅
m10	*hada+ pron. 1/2 pers.	1	*hada+ orth. pers.pron.	*hada+ clitic pron.; *hada-mā-ka-	*(ha)d(a) as procl./pref. +*orth. pron.	∅	
m11	*abi+ pron. 1/2 pers.	1	*abi+ orth. pers.pron.	*abi+ clitic pron.; *abi-mā-ka-	*(a)b(i) as procl./pref. +*orth. pron.	∅	
m12	*upari+ pron. 1/2 pers.	1	*upari+ orth. pers.pron.	*upari+ clitic pron.; *upari-mā-ka-	∅		

Directionality and costs: In order not to base subgroups on shared losses of fused prepositions and pronouns, all losses are assigned cost 1. Once lost, reverting to the other states is disallowed. Undergoing one of the two types of fusion (0 > 1, 0 > 2) is assigned cost 3. We presume that the preposition must have been preserved as a simplex relatively late in all languages, which means that reverting to the PIr. state or swapping between the two types of fusion are all assigned cost *2*. Grammaticalising the adposition as a case-like postposition is assigned cost *1*, unless it requires losing fused forms which is assigned cost *2*. Reversing this by making the case-life suffix a preposition or a prefix is assigned cost 4.

2.3.9. m13-16: Demonstratives with pre- and suffixes (Wendtland §3.6)

Sogdian and Shughni share an innovation whereby demonstrative pronouns can be prefixed and suffixed at the same time. However, it is difficult to pinpoint exactly where the innovation began. The two languages share the same three-term deictic pronouns combined with the prefix **hača-* and the suffix *-*antara*: Sogd.

[28] Other than the much-cited handbooks and chapters, we have relied on (Rastorgueva & Ėdel'man 2000: 72–3; 2007: 301–2, 316–7) and (Paxalina 1959: 39; Efimov 2011: 222–3; Emmerick 2024: 316). We note that Morgenstierne (1974: 57) derives the preposition *par/pər* from **parātara* rather than **upari*.

cyw'nt, Shughni *azamand* < **hača-ai̯am-antara*. However, Sogdian has only this combination, whereas Shughni also attests prefixed demonstratives without the suffix.

From here, it gets increasingly difficult to describe the innovations as shared, mostly because the formations happen to be attested with different adpositions and suffixes, although the starting point might very well be shared. In **m14**, we have treated the spread of the feature to other prepositions, still combined with the deictic demonstratives and *-*antara* (Sogdian attests **hada*-, **ana*-, **upari*, and Shughni **tara*-). Conversely, in **m15**, we have conjoined the spread of the feature to other suffixes as well (Sogdian attests an etymologically opaque *-*da* (> -*'γδ*) and Shughni *-*arda*).

In an effort to stay true to the attested states while simultaneously capturing the potentially shared innovation of **m13**, we have treated the Shughni and Sodgdian situations as two different states. Accordingly, we do not insist on the route of innovation; whether Shughni gained prefixed forms independently, Sogdian lost them, or they both innovated entirely independently are possible options.

Since the tendencies of **m14** and **m13** cannot be seen in isolation from **m13**, we have downweighted these to *1* and kept **m13** at the standard weight *3*.

States:

Id	**feature**	**weight**	**0**	**1**	**2**
m13	*hača-DEM	3	Absent	*haca-DEM & *haca-DEM-antara	*haca-DEM-antara
m14	productivity of the pattern adposition + DEM-*antara-	1	Absent	*hada-,*upari-, *ana-, *tara- + *DEM-antara	
m15	productivity of the pattern adposition + DEM + suffix	1	Absent	*hača, *hada, *upari, *ana, *tara- + DEM + *-da-, *arda-	

Directionality and costs: In order not to force subgroupings based on joint loss of spurious features, we have assigned all losses, in this case reverting to the PIr. state *0* cost *1*. is assigned cost *3*.

For m13, we have assigned cost *1* to changes between the innovative states *1* and *2*. This may seem disproportionately low, but it is necessary to keep the cumulative costs for these states low because we cannot force the two diverging states through and unattested common predecessor, and because we do not wish to insist on knowing the route of innovation. However, this solution leaves it – all other things equal – more parsimonious to presume that the innovations are independent: cost *3* from the root to either of them plus *any* cost between then will be greater than *3*.

2.3.10. m17-20: Local adverbs with suffixes (Wendtland §3.7)

Finally, local adverbs can also be suffixed in some EIr. languages. Like the isoglosses treated above, the systems are parallel but not entirely identical. They are all incorporated into the deictic pronominal systems, but not all languages attest all combinations. We deem this a historical coincidence. To avoid polymorphism, we have split the isogloss into characters according to the suffix, and we have weighted them accordingly: The standard weight *3* is assigned to m1*7*, the most widespread as a proxy for the cradle of the innovation, m1*8* and m1*9* have been downweighted to *1*,

Our material and analysis differ slightly from Wendtland's. Wendtland seems to connect the suffix of Sogdian *mrδ*, *trδ*, *'wrδ*, Bactrian *μαρο, οαρο*, Xufi[29] *amard, adard, udard* with Ossetic *ardæm, ūrdæm* all from **-arda* 'side'. However, it is preferable to connect at least the Bactrian forms with the Khotanese forms *mara* and *vara* going back to **ima-θra* and **au̯a-θra* (Sims-Williams 2007: 231, 242; Gercenberg 1981: 268, 299). The Khotanese forms are interesting because they must have been coined and lexicalised before the restructuring of the deictic pronouns (the exception to this is proximal *ttara* < **ta-θra*).

It is impossible to connect all these suffixes with one reconstruction: **-θra* would not yield Shughni *-rd*, and **-arda* would not yield *-(α)ρο* in Bactrian. Note, however, that ***-rδ* and not *-rd* would be the expected outcome in the Shughni-group (Morgenstierne 1974: 66). Sogdian is difficult, as both suffixes would probably yield the attested *-rδ*. However, there may be some evidence in favour of both suffixes having been present in the Sogdian prehistory: Christian Sogdian in principle distinguishes /θ/ from /δ/. This dialect attests *kwrθ* 'whither' < **ku-θra-* (cf. Av. *kuθra*), but also *wysprd* 'everywhere' < **vispa-arda-*. At face value this points to both suffixes surviving next to each other. We have taken these forms as circumstantial evidence for the productivity of both patterns in "pre-Sogdian" and thus counted *mrδ*, *trδ*, *'wr*δ doubly, as evidence of **ima-arda-* and **ima-θra-* etc. simultaneously. Unfortunately, there is some fluctuation of word final *δ* and *θ* in Christian Sogdian (*cywyd*, *cywyθ*), which weakens this argument substantially. It is thus possible that *wysprd* goes back to **vispa-θra-* with either orthographic, phonetic or morphological blending of the suffixes **-θra* and **-arda*. Additionally, it may be that **kuθra* was lexicalised earlier and thus cannot be connected with the forms quoted here.

[29] Wendtland only mentions Xufi forms, but other dialects of the Shughni group also attest at least one suffixed local adverb of this type: Shughni *yůdard*, Roshani *údari* (Parker 2023: 288; Sokolova 1966: 382; Morgenstierne 1974: 66).

States:

Id	feature	weight	0	1
m16	DEM + *arda-	3	Absent	fused form: *ima-arda-, *ta-arda-, *awa-arda-, etc.
m17	DEM-*θra-	1	Absent	fused form: *ima-θra-, *ta-θra-, *awa-θra-, etc.; *ku-θra-
m18	DEM-*da-	1	Absent	fused form: *ima-da-, *ta-da-, *awa-da-, etc.
m19	DEM-*da-aida- (?)	2	Absent	fused form: *ima-da-aida-, *ta-da-aida-, *awa-ta-aida-, etc.

Directionality and costs: Like the preceding sections, loss of a marginal formation has been assigned cost *1* as to not force subgroups based on shared losses of them. Taking part in the innovation has been assigned *2* (for **-arda*, grammaticalizing a noun as a suffix), *2* for **-θra* and **-da* (generalising inherited and frequent suffixes) and *3* for the opaque conglomerate **-da-aida-*.

3. Results

3.1. Results of the Maximum Parsimony analysis

When performing the Maximum Parsimony analysis in *LinguiPhyR*, the software returns multiple trees with almost equal scores: Two (best) trees score 655, three second-best score 656, one third-best score 657, two forth-best score 658 and three fifth-best score 659 and so forth.[30]

The scores of the best trees might seem high, but they have no inherent value and simply reflect the sum of all transitions of all characters required to get from

[30] To arrive at these results, we ran the analysis multiple times. *LinguiPhyR* is a very convenient tool for running a search with *Paup**, but the standard settings are not ideal for our dataset: Beginning the search at a randomly generated tree and running 25 iterations. The user decides whether the App should keep only the single best trees or all trees (up to 100) below a certain score. With 17 languages it is impossible to run an exhaustive search, and it is possible that the truly best trees hid among the billions coincidentally not surveyed. To perfect the search, we did some manual tweaks: first, we repeated the search in the App, keeping all trees below a parsimony score close to the best (in this case, 640 (not accounting for weights)). Then, we used this result to repeat the analysis running *paup** directly from the terminal. In this search, we replaced the random starting point for the search with the 100 best trees found using *LinguiPhyR*, ran 10000 iterations instead of 25, and repeated the process keeping only trees below a number close to the all-time best score. For the last of these optimisation runs, we ran 25000 iterations and kept all trees below a score of 641 (without weights). This yields 33 trees. The result of this search was then uploaded to explore using the tools of the App.

the Proto-Iranian root to the leaf states on this tree. In other words, there is no golden standard, and running the analysis with a different scale of weights and costs would give completely different scores. The scores can, however, be compared directly with each other. This quickly reveals something about the phylogenetic signal: A clear signal will yield a few trees that are clearly better than their alternatives, and the difference between the scores will be great. A more muddled signal will yield a plethora of different trees with almost the same score.

One character, **m13**, is parsimony uninformative – but this is a technicality; while no two languages share the same state, the two innovative states are derivable from each other and there is still potential for candid phylogenetic information in the coding (see **2.3.9.**).

23 of the 28 characters are incompatible with these best trees – meaning that they cannot be fitted onto the tree without inferring backformation or parallel developments. 3 characters (**m1**, **m6** and **m8**) are consistently compatible with these best trees (this issue will be discussed in **4.3.3.**).

Most characters enforce the best tree in some ways; although it may be impossible for a character to develop from the root to every single leaf without backformation or parallel innovations, it may very well enforce one – or several – subgroups in other areas of the tree. On the best trees, 18 of the 28 characters provide the key innovation on the configuration of the tree, and some characters provide multiple subsequent innovations at different internal nodes in the tree.

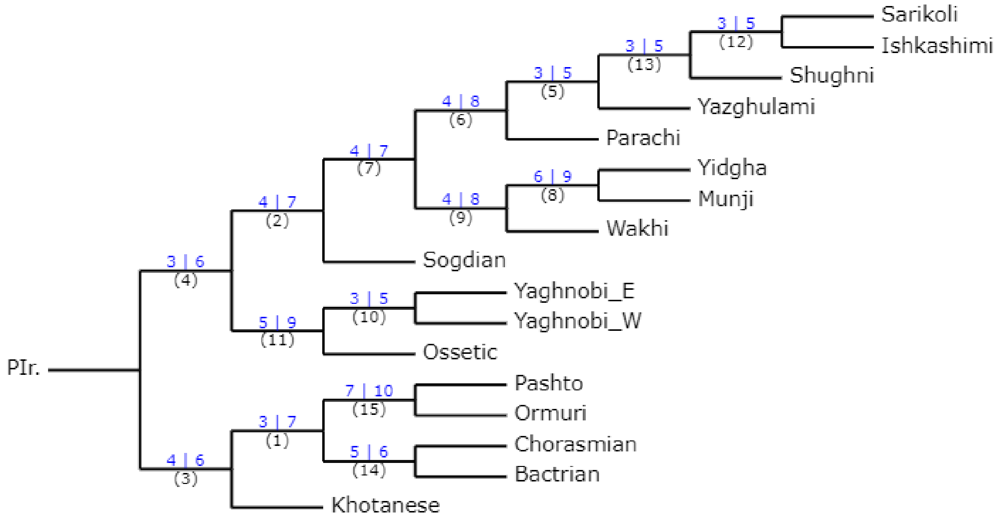
In Figure 6 below, a representative sample of the best trees is presented. The trees are to be read as follows: The bracketed black number below a line is the identifier of the split occurring directly to the left of it. This is known as the *edge id*. The blue numbers above the above this identifier show the support for this edge. The left number tallies the number of characters supporting it, and the right number is the same support, but adjusted for weights. Currently, there is no easy visualisation of the edge support adjusted for costs as well.

The trees shown are: One of the two best trees (Tree 1), two of the three second-best trees (Trees 2 and 3), the third-best tree (Tree 4) and one of the two fourth-best trees (Tree 5).[31] Although these five (seven if counting the variation described) trees

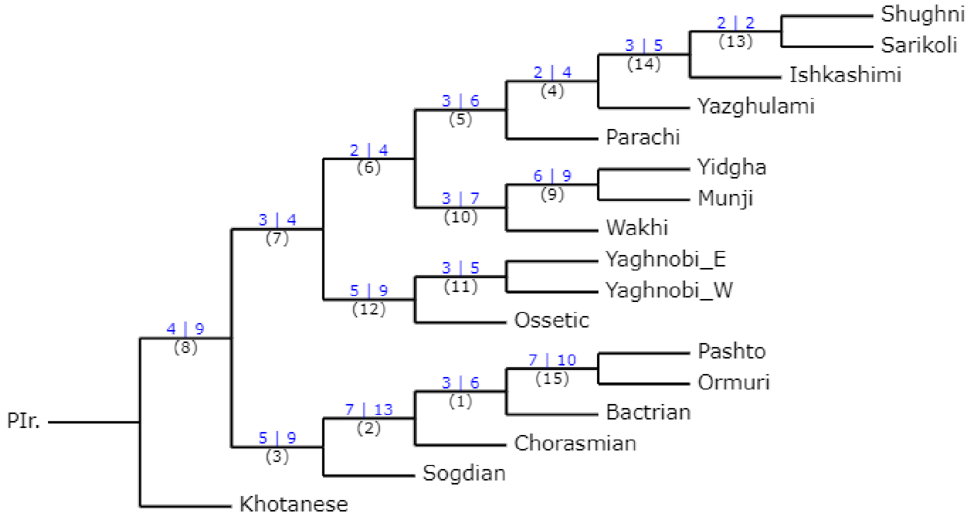
[31] The reason for giving only one of the best and one of the second-best trees here is that they find the same groups but in complementary distribution. The two best trees only differ on the relative position of Sarikoli, Ishkashmi, Shughni and Yazghulami. In the tree not depicted here, these languages are grouped as in Tree 2. Similarly, the second-best tree not depicted here has the same overall distribution as Tree 2, except for the relative position of Sarikoli, Ishkashmi, Shughni and Yazghulami, where it agrees with Tree 1. The second fourth-best tree is identical to Tree 4 but groups Yazghulami-Shughni against Wakhi—Sarikoli-Ishkashimi.

only make up a small section of the reasonable possibilities, they serve as a useful comparison of the relevant subgroups.

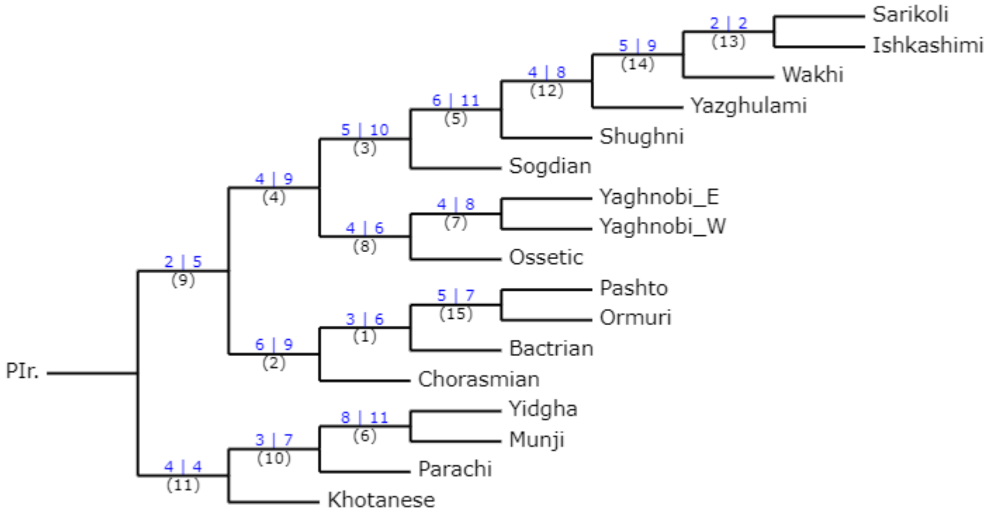
Figure 6. A sample of the best trees



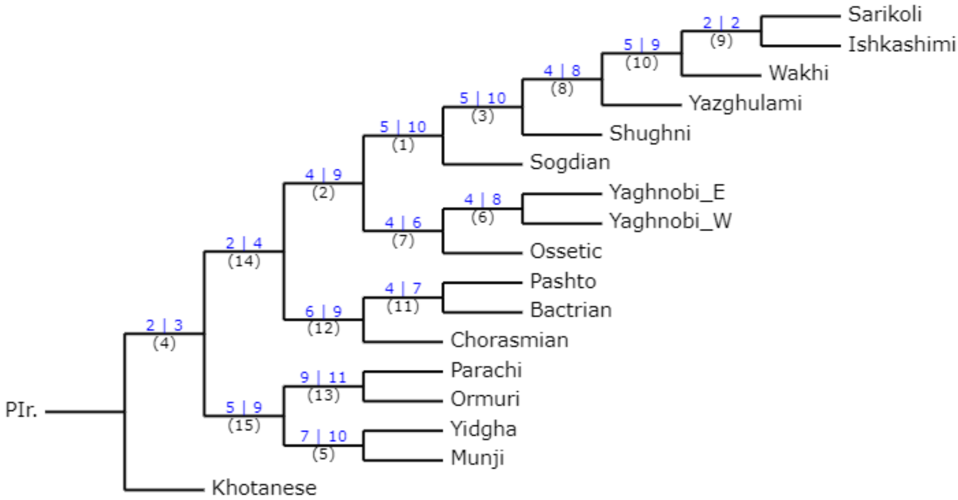
Tree 1



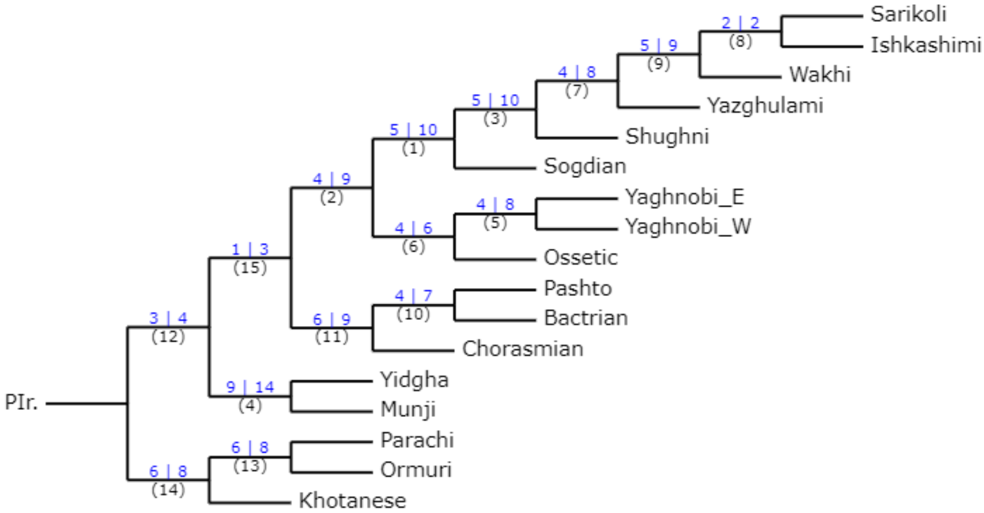
Tree 2



Tree 3



Tree 4



Tree 5

3.2. Consistency and robustness of the results

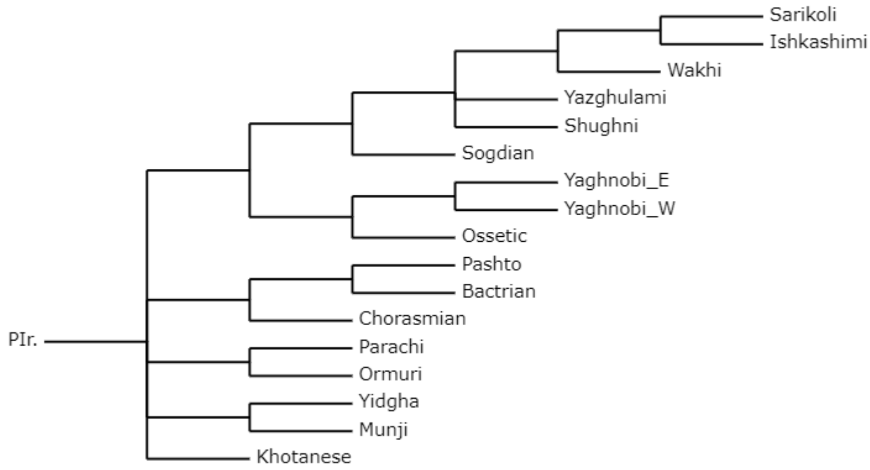
At first glance, these five trees look radically different although they have very similar scores. While especially the higher-order subgroups diverge, all trees consistently find some of the same lower-order subgroups: Yidgha and Munji are consistently grouped together, as are East and West Yaghnobi, and these two are always returned as the sister of Ossetic. The best trees presented all find a close relationship between Shughni, Sarikoli, Ishkashimi and Yazghulami, although their relative positions within this clade are not consistent. Most trees find Wakhi to be a sister of Yidgha-Munji rather than the aforementioned group, and they differ substantially on the relationship between Parachi, Yidgha-Munji-Wakhi and the other Pamir languages. Strikingly, the best trees group Pashto and Ormuri together, which leaves Parachi among the Pamir languages, because all trees also find close ties between Pashto, Bactrian and Chorasmian. The latter two form a clade in most of the best trees.

In most trees, Khotanese is either the outlier – or the first to split off within an outlier-branch. The position of Sogdian is the most unclear shown by the fact that it shifts from one higher-order subgroup to the other across the best trees.

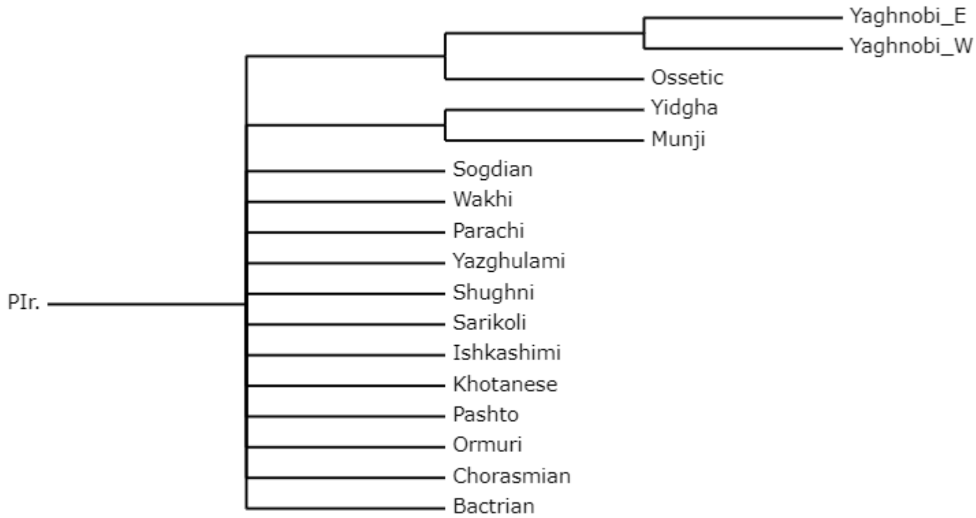
Our best trees are not very different in terms of cost: Only a single point distinguishes the two that tie for first place from the three that tie for second place, and only one point distinguishes these from the third best and so on. Given our weights and costs (see **1.4.6.**), the difference separating these trees is minuscule, a single shared trivial phonetic innovation could change the picture. It is thus apparent that there is no clear phylogenetic signal in the data.

The robustness of the results can be checked by comparing the consensus trees in Figure 7. The Strict Consensus tree shows only the subgroups which all trees of the analysed file agree upon. In this case, the 33 trees of the final run (see fn. 30 above) only agree on two groups, the rest being presented as a rake-like starburst-dispersal. The Majority Consensus tree, however, reveals that our trees are not irreconcilably different: Following the uncertainty of the higher order groups, it only yields one trifurcating split. this must be considered a great success given the input data.

Figure 7. The consensus trees



Majority Consensus tree



Strict consensus tree

To test our manual inference on the results, we repeated the search with three modified datasets: One with only transition costs and the characters weighted equally, one with character weight but equal transition costs, and one with neutralised weights and costs only relying on our custom directionality restraints. Results of these analyses are shown in Figure 8. All of these are largely similar to the trees found in our primary analysis.

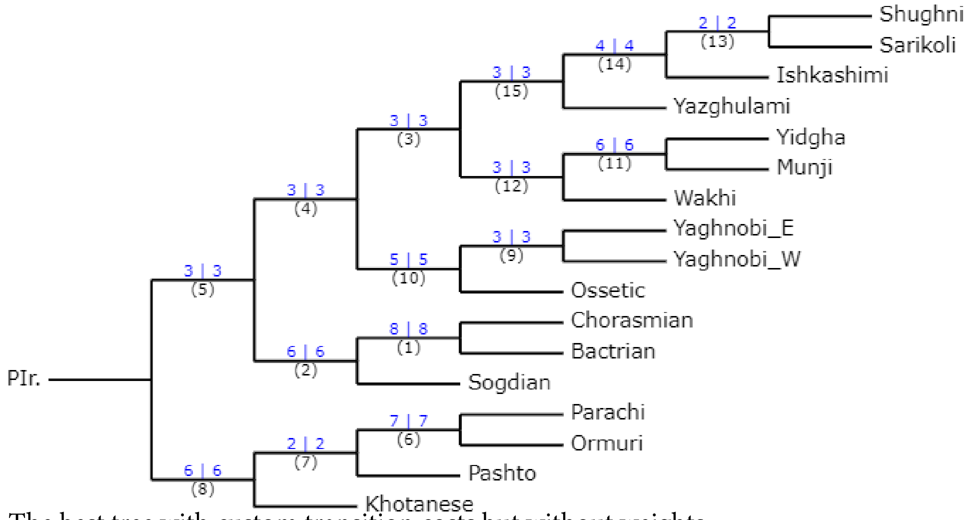
When removing the weights but keeping the transition costs, the best tree and the Majority Consensus tree group Parachi and Ormuri together – as a sister of Pashto. Similarly, they show a clade of Sogdian, Bactrian and Chorasmian. They find the same relationship between Ossetic and Yaghnobi and between Yidgha-Munji and Wakhi as above. Removing the weights neutralises the differences across types of characters (innovations in phonology and morphology become equal), and most importantly, the somewhat interconnected characters involving fused prepositional forms become disproportionately important.

Keeping the weights but equalising the transition costs makes more of an impact. The software finds four equally good (or bad) trees. However, they do share quite a lot of subgroups, only the relative placement of these differs. Parachi-Ormuri tend to be grouped with Pashto or Khotanese, and Yidgha-Munji is also found as a relatively early split-off (though from which branch is unclear). In these trees, Wakhi goes with Shughni, Sarikoli, Ishkashimi and Yazghulami rather than with Yidgha-Munji. As in the unweighted trees, the analysis tends to group Sogdian, Bactrian and Chorasmian. In this analysis the emphasis on the different types of evidence is kept, but all innovations are counted equally. All shifts, no matter how significant or trivial, are weighted equally.

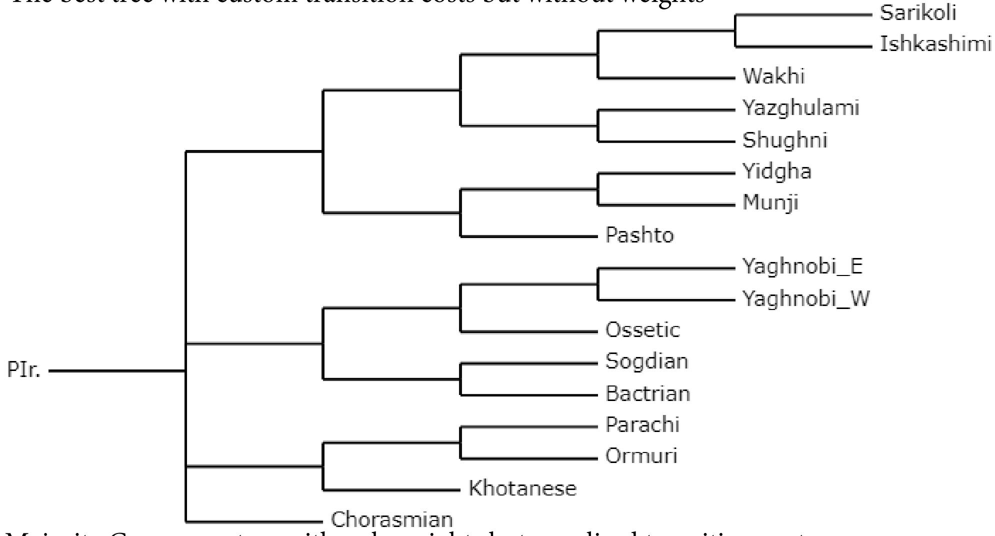
Removing both costs and weights once again yields similar results. These trees tend to group Pashto with Yidgha-Munji and Parachi-Ormuri with Khotanese within a branch. As in the other unweighted trees, Sogdian, Chorasmian and Bactrian form a clade, as do Ossetic and Yaghnobi – and the remaining Pamir languages.

Although these deviant results do not reflect our best linguistic analysis, it is very informative as to which groupings lie in the sheer number of changes, and which lie in the quality of shared isoglosses. We saw now point in experimenting with other character types since our directionalities are quite generous. Letting the surface distribution rather than the likelihood of the individual transitions determine the tree would be unfeasible. Relying only on directionalities and less of the type of change tends to group more conservative languages together – especially those that exhibit conservative traits in their pronominal stems (on this issue, see **4.4.3.**).

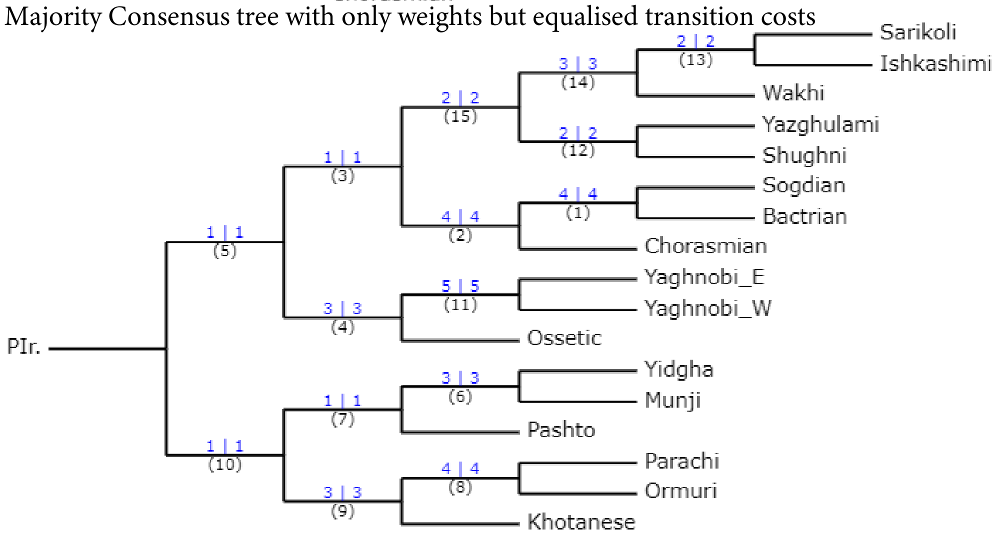
Figure 8: The best trees of the alternative analyses



Tree 6: The best tree with custom transition costs but without weights



Tree 7: Majority Consensus tree with only weights but equalised transition costs



Tree 8: One of two best trees without weights and with neutralised transition costs

4. Discussion

4.1. Structure of the discussion

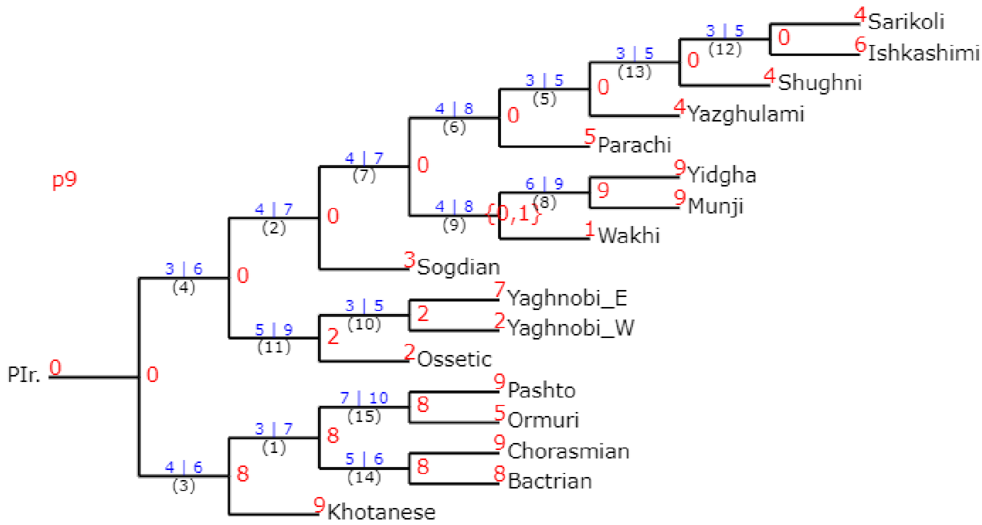
The aim of this section if to compare the results of our Maximum Parsimony analysis with the existing proposals for the subgrouping of the EIr. languages. However, before this comparison can be undertaken in any meaningful way, it is crucial to scrutinise our results to fully understand what the basis of the comparison is. Therefore, we begin by examining the inferred development of the individual characters and the innovations supporting some of the subgroups of the best trees. The computational analysis itself is nothing but a numbers' game, but the advantage of our approach is that the impact of all choices in data selection, character coding and especially directionality and assigned weights and costs is readily available for evaluation. The method supplies a valuable tool for comparative analysis for linguists, not a final say in the debate.

4.2. The inferred development of the Phonological characters

Although we have allowed for a great deal of phonetic backmutation to be possible, scattered archaisms force the PIr. state to be inferred for many intermediate nodes in the tree in many cases. While this may look unrealistic, it is not too different from what would have to be reconstructed "by hand" – this is the logical consequence of not being able to reconstruct an EIr. proto-language different from PIr. itself.

For instance, in the case of the development of **-θ-*, **θr-* and **-θr-* which develop into *-t-*, *t(i)r-* and *-rt-* in Western Yaghnobi, but *-s-*, *s(a)r-* and *-rs-* in Eastern Yaghnobi, the differences between the two dialects are traced back to the preservation of the PIr. state (Skjærvø 1989a: 375). Since no one sees Yaghnobi as an outlier, the consequence of this is to reconstruct the PIr. situation for all intermediate nodes until Yaghnobi finally splits off. See also Figure 9.

Figure 9: The inferred development of the character **p9**, *-*θr*- on Tree 1.



Conversely, some phonetic changes are assigned costs that are so low that the software finds it more likely that a few languages traditionally considered conservative outlier within EIr. because they preserve the PIr. state rather underwent independent backformations. This is the case for the voiced stops in Parachi and Ormuri (Sims-Williams 1996: 650). Since we deem word-initial fortition of **β δ γ* > **b d g* slightly less costly than lenition of **b d g*, the Parsimony analysis infers **b*- and **d*- (but not **g*- on the trees in which Khotanese is the outlier) to have shifted to **β δ* already on the "Proto-Iranian edge". What this means is that it is more economic to infer the developments for all of East Iranian with subsequent backformation in the few branches otherwise seen as conservative – because they take part in other and more salient innovations.

Particularly one character, the first one, the development of **č*, has costs so low in either direction that it contributes very little to the configuration of the tree. For almost all branches, the PIr. state is inferred to have been preserved until, finally, multiple independent shifts of **č* > *c* took place on the leaf edges. In some way, this is compatible with the traditional notion that it is a shared isogloss of the EIr. languages to have preserved **č*, although this has very little phylogenetic information.

Figure 10: The inferred development of character **p4**, **d*- on Tree 1.

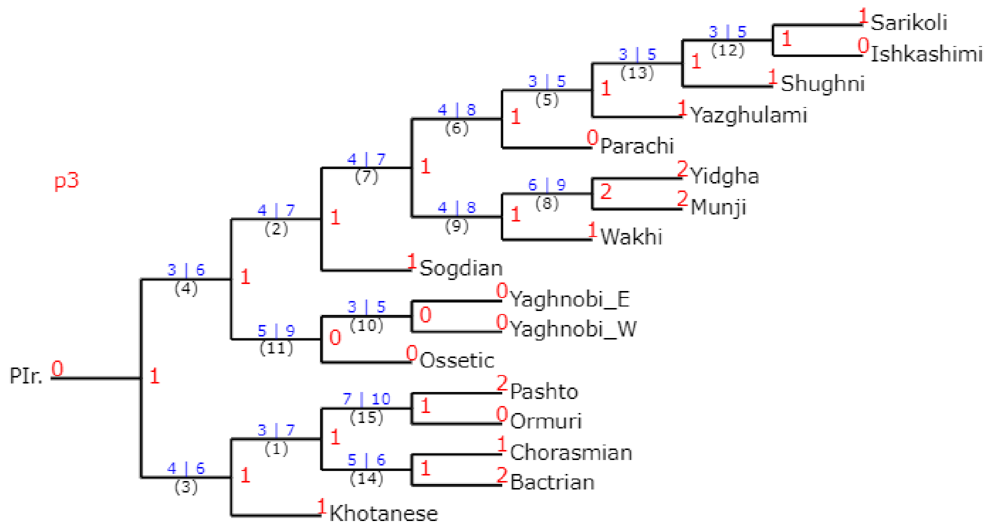
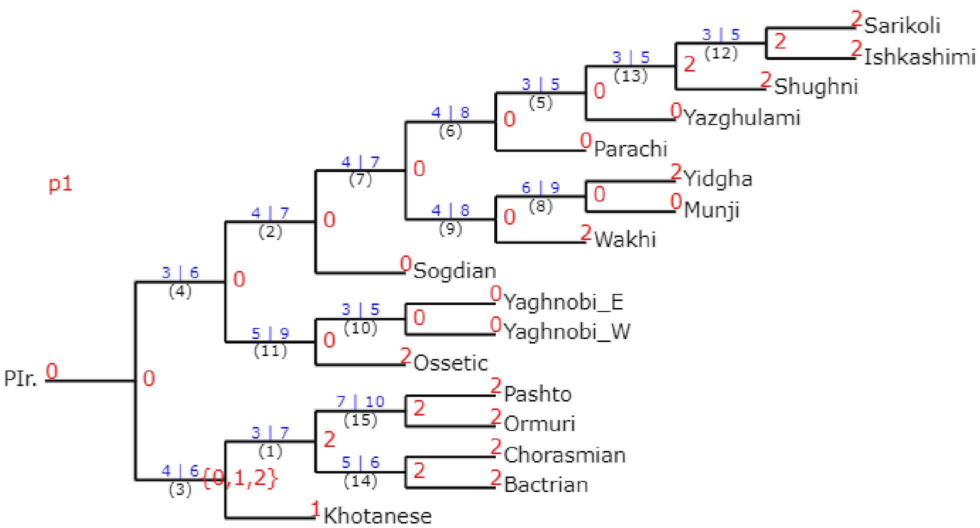


Figure 11: The inferred development of character **p1**, **č*- on Tree 1.



Our strategy of assigning high costs to very unlikely phonetic changes rather than restricting them has some unfortunate consequences. In a few cases, the software infers some rather bizarre sound changes. These are, all other things taken into account, the most economical solution to the App – but not the human reader. For instance, in tree 1, the App infers the development **ft* > **u/vd* > *βδ* for Bactrian and Chorasmian, which is probably not the most attractive route en face but done so because of the inferred relationship with Khotanese (where the outcome is *ut*). While word-internal *-*θr*- is inferred to have been preserved very late on one major branch, it is inferred to have shifted to **hr*- in the other. Again, this fits Bactrian-Chorasmian -*hr*- and Khotanese -*r*- reasonably, but the inferred change of *-*hr*- > -*š*- in Ormuri raises caution. However, this change is actually reconstructed – **θr*, **fr* and *xr all coalesced in *hr which became *š* in Ormuri (Logar) and *ř* in Kaniguram (Efimov 2011: 47).[32]

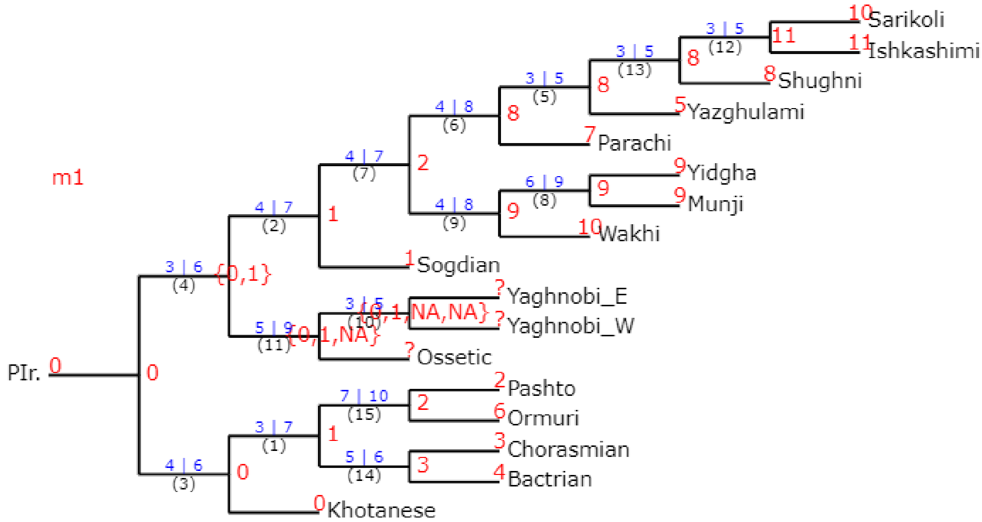
4.3. The inferred development of the Morphological characters

4.3.1. Innovations within the inherited system

Regarding the inferred development of the morphological characters, two main tendencies need to be addressed critically. First is the difference between inherited and innovated morphs, and second is the influence of our weight and cost system on the results and the consequences of the inferred relative chronologies and inferred independent innovations. When examining the inferred development of the characters, there is a clear difference between innovations that occurred within the ancestral system and innovations that incorporate new material into the inherited system. The former (**m1**, **m6**, **m8** and **m5**) behave much like the phonological isoglosses described above. For the nominal plurals, the system is inferred to have remained relatively conservative until after the break-off of Khotanese (which is no great surprise as this is the only language preserving the inherited case system somewhat intact). The software cannot infer what happened in the Ossetic-Yaghnobi-clade and its predecessor (which is also unsurprising, given that we have coded these as "?" since they no not preserve the relevant endings at all). From there, the App infers two identical but parallel innovations, namely the reduction of the inherited NOM-ACC-system to a Sogdian-like system with a direct case, a genitival case in **-ānām/-īnām* and some oblique case in **-b-*. Because Pashto is found to be deeply embedded on the one branch, and this language preserves **-ai̯ah, *-ānām, *-īnām* and **-b-* as plural markers (although not in their original functions), all formants must have been preserved late on this branch. Similarly, because various Pamir languages generalise **-b-* or **-ānām* as their sole plural ending, and these are found to form a clade, the final reduction of the formants must have been recent.

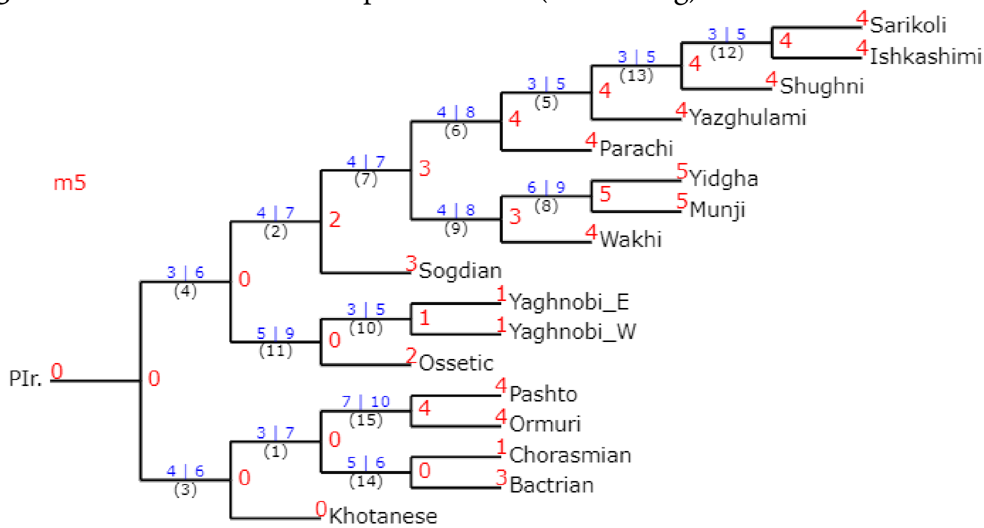
[32] The real issue here is whether this reflex coalesced with PIr. **hr-* or not. If it did, the inferred development is impossible without intermediate steps. The synchronic instability of /h/ in Ormuri makes this impossible to assess (Efimov 2011: 83–4).

Figure 12: The inferred development of **m1** (nominal pl.) on Tree 1.



The same applies to the verbal 3PL ending where both formants, originally going back to the active and middle, respectively, of the verbal 3PL ending are found to have been present surprisingly late in the EIr. languages. On the one hand, this looks strange as Khotanese is the only language to preserve both variants (though not strictly in active/middle function). On the other hand, this must be the logical consequence of Yaghnobi having preserved the ending "unknown to literary Sogdian but found in Khotanese (3 plur. subj. *-āru/o*) and Chorasmian (3 plur. subj. *-'*r^i and impf. *-'*r^a)" (Skjærvø 1989a: 376) – if these languages are not to be grouped as a relatively conservative outlier clade. Consequently, the generalisation of **-nt-* over **-r* should have taken place 4 times independently: In Bactrian, "Pashto-Ormuri", Proto-Yaghnobi and the "Sogdian-Pamir" clade.

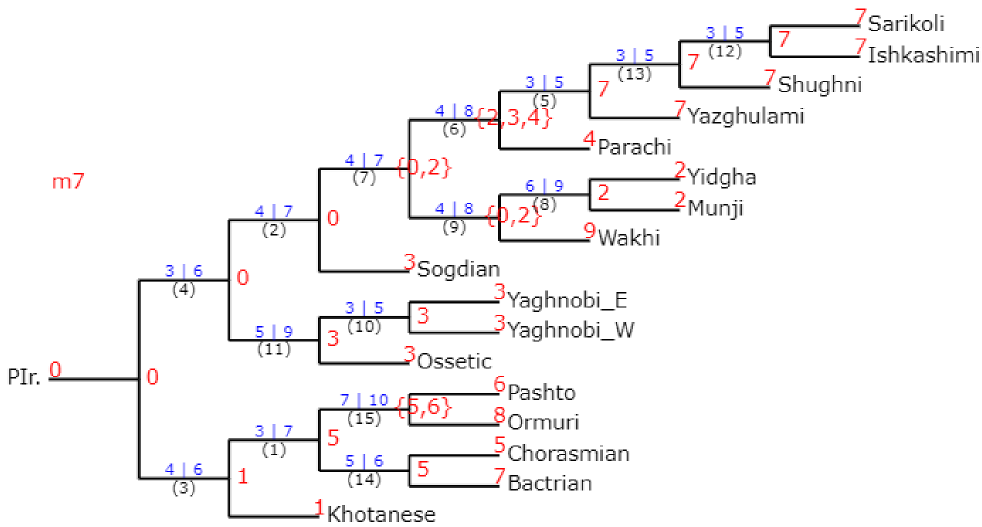
Figure 13: The inferred development of **m5** (3PL ending) on Tree 1.



The pronominal characters are found to have preserved archaic states relatively late. For the second plural pronoun, the clitic *⸗u̯ah must be preserved until the break-off of Parachi. As our analyses mostly find Parachi to be deeply embedded because of traits shared with the Pamir languages, this means that *⸗u̯ah must be deeply embedded. This has the unforeseen consequence that the chain of innovations whereby the 2PL was replaced by the 2sg stem should have occurred several times independently. Especially strange is Bactrian and Shughni-Yazghulami who share (the 2PL synchronically being identical to a compound of the 2PL and the 1 PL) the same state but very different routes to obtain it.

The inferred states of many intermediate nodes should definitely be taken with a large grain of salt. Since the software only operates with the numerical states and we cannot feed it hypothetical intermediate systems, preserving the Proto-Iranian state means nothing more than *not* having undergone a divisive innovation whereby this stage cannot be reached again. Thus, if we were to reconstruct the latest common state shared by Sogdian, Wakhi, Yidgha-Munji, Parachi and Shughni-Yazghulami, it would be more reasonable to state that this node must have preserved **i̯ušmabi̯a*, **šmāxam* and *⸗*u̯ah* rather than claiming that the entire paradigm was preserved. This is no major obstacle, it only shows that our analysis is a tool, not a final say.

Figure 14: The inferred development of **m7** (2PL pronoun) on Tree 1.



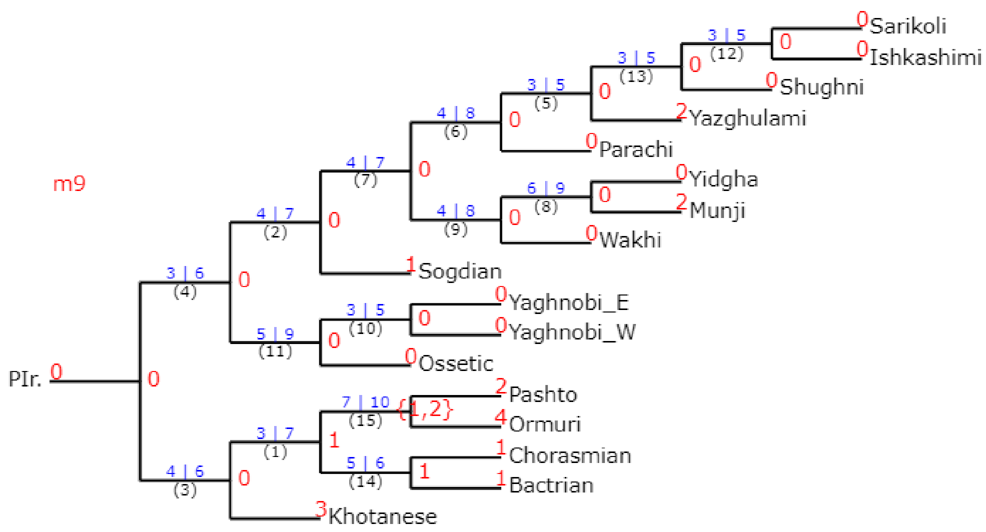
The deictic pronominal stems also behave this way on our trees. The ancestral state is preserved fairly late on one branch with the consequence that the diverging attested systems seem to have sprung out of nothing independently in Bactrian, Chorasmian, Pashto and Ormuri. On the other branch, the Sogdian system (where both stems of three out of four stems are preserved in a three-term deictic system; see **2.3.7.**) is inferred as the ancestor of Ossetic, Yaghnobi, Parachi and all Pamir languages.

4.3.2. Innovations involving new morphemes

Conversely, the isoglosses where we the innovation lies in the grammaticalisation of new features and not in restructuring of inherited inventory behave differently. Part of this is self-explanatory: The software must infer the innovations at some point between the PIr. root and the leaves. Since these isoglosses are is not shared by all languages sampled, the innovations are unlikely – but not impossible – to be inferred for the PIr. edge. The surprising part is how these isoglosses fall on our best trees. If the characters were easily compatible with these, the innovations would simply be inferred for the latest shared node of the languages sharing the innovation, i.e. the innovation would be reconstructed for the latest common ancestor. For the most part, this is not what we find. These isoglosses fall out in two ways. Either they are so incompatible with our best trees that the software infers multiple independent but identical innovations, or they are inferred to have occurred relative early in the evolutionary history and lost again multiple times.

The fused prepositional forms all belong to the former category, and the novel plural formations to the latter. The prepositional forms are scattered across a variety of languages, but there is no clear pattern – Sogdian often takes part and shares innovations with various branches (Bactrian and Chorasmian in **m9–m12**, Shughni in **m13–15** and Bactrian (and Khotanese) in **m17–19**), and Yazghulami and Munji share a pattern in **m9–12**. But as not all languages attest all the encoded combinations, even this pattern in vague.

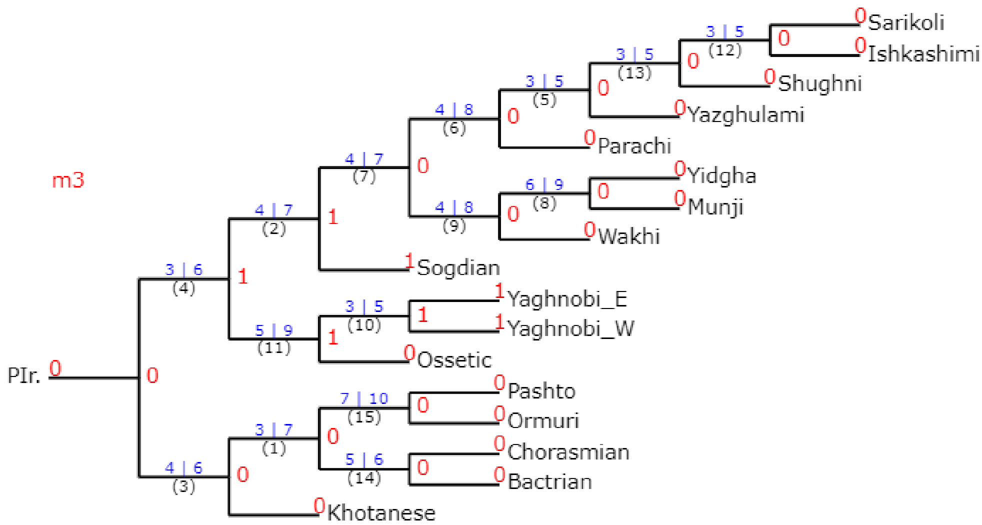
Figure 15: The inferred development of **m9** (**hăca* + 1/2PL pron.) on Tree 1



Among the latter category are the nominal plurals derived from the collective in *-*tā* and the neo-plural in *-*išt*-. These are inferred to have arisen already after the break-off of Khotanese – and lost multiple times independently. It is highly dubious that these endings ever existed in the prehistory of almost all EIr. languages. The reason for this bizarre result may lie in the isogloss itself, or it may be a result of our coding strategy: If the isogloss did in fact not develop in a tree-like fashion, it will look excessively strange being forced onto a tree. This, however, leaves the obvious question open if it is indeed these isoglosses – or the ones compatible with our best trees – that reveal the underlying tree-structure. Our software answers this by comparing the cumulative costs, and thus our weights and transition costs are what make the difference when the isoglosses point in multiple different directions.

Many of these results might have more to do with our codings than with linguistic reality: Since we do not want to build our tree on shared loss of features, we have downweighed the loss of newly created morphemes which has the consequence that it forces some innovations, notably the plural in *-*išt*-, up in the tree, because it is more parsimonious to posit multiple independent losses than independent (incompatible) innovations.

Figure 16: The inferred development of **m3** (PL in *-*tā*-) on Tree 1



4.3.3. Issues of coding and weighting

The inferred developments of these isoglosses reveal some drawbacks of our coding strategy. The first type of morphological isoglosses – the ones that are simultaneously compatible with all our best trees *and* inferred to have remained relative conservative over the course of the development of the EIr. languages turn out to be the most important ones for the configuration of out trees. Not necessarily because we have weighted them disproportionately, but because we have restricted their directionalities the most. Consequently, we may have forced the software to return trees compatible with these isoglosses, because it is possible to accommodate most other isoglosses for which more backformations are allowed onto those trees. Some of these inferred parallel developments, especially phonological ones (chiefly **č* > **c*) may indeed have happened multiple times, but some highly specific morphological ones (especially in the pronominal isoglosses) are quite unlikely to have occurred multiple times in the evolutionary history of the EIr. languages.

A further complicating factor might be that out software assesses innovations as the replacement of one numerical state with another. In grammar, the process of innovation hardly ever looks like that. Competing forms arise, and one is generalised over the other, gradually. While the resulting outcome is the same, and it is therefore not unreasonable to code only the final result, the reality of the innovation might be masked by this and might in turn not be very well representable on a tree.

A final grave issue is the simple fact that weights and costs multiply. Although we have been aware of this throughout the process and adjusted the weights accordingly, a careful recoding of all weights and costs might yield better results. Loss of morphological features is assigned a transition cost of 1, and gaining a feature cost 3. Accordingly, in the analysis with costs and weights, gaining a morphological feature has a total cost of *9* (*3*3*), and morphological loss cannot cost less than *3* (*3*1*). This leads to the unrealistic situation that any shared new feature will be projected unrealistically far back, because gaining it once will have the same total cost as losing this feature three times independently (*3*3*1=9*). Similarly, morphological loss unintentionally becomes as important as a "possible" (more unusual than "trivial" and "plausible") phonetic change. This is a difficult issue to solve – it would be untrue to linguistic reality to posit very high costs to loss of inherited feature. Additionally, this strategy would force our subgroups to be based on the linguistically common process of loss. On the other hand, lowering the costs of morphological innovations, which are considered the strongest evidence of linguistic genetic affinity, would have the unintentional consequence that the same innovation should have happened multiple times independently. Some innovations, like the restructuring of the case system of the nominal plural or the generalisation of one 3PL verbal ending over the other could perhaps be downweighed, since these innovations rely heavily on archaisms (coincidentally preserving the same endings).

On top of these issues is the question of areal features and contact-induced change. While we have sorted those features out where the shared isogloss cannot be formulated in such a way that a single preform or system could be reconstructed for a common prestage, the gradual restructuring of the nominal case system or the rise of the three-term deictic systems might indeed have an areal component to them. In some cases, this would be indistinguishable from a true genetic innovation: The starting point for the neighbouring speech communities in question was ultimately the same Proto-Iranian language, and it is hardly surprising that many innovations involve inherited material. Areal influence might indeed cause etymologically identical morphemes to develop the same functions after the break-up of the clade.

Ironically, a much more detailed understanding of the very relative chronology of splits and innovations that we are trying to shed light on, would massively improve the basis for our analysis.

4.3.4. The inferred relative chronology and the characters enforcing the subgroups

In this section, we will discuss some of the evidence in favour of the inferred subgroups. Note that this is not just a description of shared features – these can be found in section 2 – but the post-analysis break-down of the inferred shared innovations that the subgroups in the best trees rest on. Additionally, we will address some of the discrepancies between our analyses and some clades usually assumed in the literature. In this respect, it is as interesting to us why we do not find some of the more secure clades as why our software groups the languages as it does.

There are very few exclusive innovations across our data. This is not a problem per se. As explained above the software is a tool to sort conflicting phylogenetic signals. But it does make for an interesting comparison with the existing suggestions since our data is based on the summary of isoglosses traditionally invoked for describing relationships among the EIr. languages.

4.3.5. Pashto-Ormuri, not Parachi-Ormuri?

Our analyses only find Parachi and Ormuri to form a group – as usually assumed in the literature – when performed without differentiated weights. Instead, the software quite consistently groups Ormuri with Pashto.

In our dataset, Parachi and Ormuri only share one state exclusively, namely *-c-* as the outcome of **-θr-*. Because Ormuri instead exclusively shares the states of two characters, one phonological (**ft* > **u̯*) and one morphological (the vocalic 3PL ending) with Pashto, the software tends to group these two together instead (unless the weights are removed). Ormuri and Pashto further share the development of **č* > *c* (which Parachi does not take part in), and perhaps **xt* > **k* before eventually being lost. On the other hand, **b d g* are preserved in Parachi and Ormuri. These phonological features are archaic at first glance, but because of the very low cost assigned to word-initial fortition, they do not force a subgroup and are instead inferred to be independent backformations.

In morphology, these three languages often disagree. For the nominal plural, Pashto is the most conservative – although the case system is very remodelled, all the formants **-ai̯ah, *-ānam/*-īnām* and *-*b*- are preserved. In Ormuri, only *-*b*- is lost, and in Parachi only *-*ānām* remains. In the demonstrative system, all three languages have two-term deictic systems, but they do not share the same stems. Because Pashto and Ormuri agree on preserving D0 DIR **ha*- these are forced up the tree. Ormuri and Parachi agree on preserving D3 DIR **hau̯*-, but the App can easily place the Parachi system as a development from a later stage (where D0 is lost altogether). In the 2PL pronoun, Parachi has generalised the enclitic, Ormuri has replaced whatever it preserved with a Pashto loanword, which is in some way built to the orthotone 2SG **tū, tau̯a*-.

None of the languages show traits of the plurals in *-*t*- or *-*išt*-, but the cumulative evidence – and the low costs assigned to loss – makes the App infer that Pashto-Ormuri never took part in these innovations, but that Parachi did and subsequently lost the formants. This is hardly true.

There is no doubt that there has been massive influence from Pashto on Ormuri. Even in our small dataset, we have singled out the 2PL pronoun which is considered a borrowing. Importantly, Kieffer (1989: 451) suggests that the vocalic ending (merging with the 3SG) of Ormuri was created on the model of Pashto. If this is the case it should be considered a loan rather than evidence of a prehistoric unity. Similarly, he suggests that the -*n* of the animate nominal plural ending is due to contact with Persian or Pashto. If we reran the software with all of these traits considered contact-induced rather than genetic innovations, we would definitely get differing results – but it is not clear that this analysis would yield a Parachi-Ormuri subgroup.

Although they are structurally similar, there are important differences among the attested states, e.g. the development of **xt* to Ormuri *k* or *Ø*, but to *t* in Parachi which do no show signs of joint innovations. It is very noteworthy that the App infers that Parachi underwent no less than 10 independent innovations on its own edge.

4.3.6. Yidgha-Munji and Wakhi

Yidgha and Munji share the states of five characters exclusively, namely the outcomes of **θ*- and **θr*- > *x̌-, x̌Vr-,* the nominal plural case system (*-*ai̯ah*, *-*b*- > -*i, -āf*), generalising only **i̯ušmabi̯a* as the 2PL pronoun (as *mŏf*, *măf*) and the reduction of 3PL *-*nti* > *-*Vt* (*-at, -εt*). Additionally, the App infers that they share the development of **d-* > **δ* > **l-* separately form the identical outcome in Bactrian, and that they should have lost the preposition **upari* at a common prestage.

Wakhi and Yidgha-Munji do not share any states to the exclusion of others, but they do share exactly the same system of deictic pronominal stems (as in Ishkashimi and Sarikoli). Further, they share the retention of many outcomes that are inferred to have been innovated at earlier stages and preserved as such in their immediate ancestor in our best trees: **č, *β, *δ, *γ, *γδ, *βt* and generalising 3pl *-*nti*. The software also infers that they share the loss of local adverbs formed with *-*arda*, and that Yidgha and Munji took part in the innovation of the DIR.PL *-*išt* and subsequently lost it. The former is insignificant if true, the latter is highly dubious.

The App further infers that Wakhi and Yigha-Munji share the generalisation of *-*b-* as the OBL.PL marker (with a conservative DIR.PL in Yidgha-Munji and the novel *-*išt* in Wakhi) and the generalisation of the 3PL ending *-*nt* (parallel to many other branches). Since the 2PL pronoun was borrowed in Wakhi, the software (nor anyone else, to be fair) cannot determine what the borrowing replaced. It infers that it was either the **i̯ušmabi̯a* of Yidgha-Munji or something more archaic. Because of the diverging outcomes of *-*θ-* and **θr*, the App cannot infer whether the common prestage preserved the PIr. state or that of Yidgha-Munji. However, if the clade existed there are many more options.

Wakhi and Yidgha-Munji are tied to the (other) Pamir languages by the Sogdian-like deictic pronominal system but left outside the "Pamir-Parachi" clade by *not* taking part in the joint loss of fused forms of **hača-* (Munji) and the plural *-*išt-* (Wakhi), which are arguably not the strongest arguments against these languages belonging in a Pamir clade.

4.3.7. (North) Pamir

There are no exclusively shared states among Sarikoli, Ishkashimi, Shughni and Yazghulami, but they nonetheless share a lot of traits. All share **č* > *c*, **b* > **β* and **g* > *γ*. The software also infers that they shared **d* > **δ*, and that this was reverted to *d* in Ishkashimi (similarly, Cheung 2015: 49). They all partake in the (partial) voicing of **xt* and **ft*, though they differ on the final outcome. **θr-* turns to *-(*V)r* in Sarikoli, Ishkashimi and Shughni, but to *c* in Yazghulami, whereas *-*θr-* turns to *c* in Yazghulami, Shughni and Sarikoli, but *-s(V)r-* in Ishkashimi.

For the nominal plural, the App infers that all of the languages descend from a system with the same formants as inherited in Sogdian and Pashto (DIR *-*ai̯ah*, obl. *-*ānām*, *-*īnām*, *-*b-*). The App infers that – after the break-off of the Wakhi-Yidgha-Munji clade generalising *-*b-* as the OBL.PL marker – the remaining

“Parachi-Pamir” languages lost the case system but kept all formants as plural endings, like in Shughni. However, the scenario inferred in Tree 1 is very unlikely, in which Sarikoli first lost the case distinction by generalising *-*b*- as a general plural ending and then reassigned it to the OBL.PL when introducing the novel DIR.PL ending -*xεyl*. This is an artefact of our custom costs. For the 2PL pronoun, Yazghulami, Shughni, Ishkashimi and Sarikoli all show a form seemingly built on the 2SG and 1PL – identical to, but independent from the Bactrian innovation.

As mentioned above, Parachi is grouped with these for because of morphological isoglosses, but it is a strange island as to the inferred phonetic developments. The Wakhi-Yidgha-Munji clade is slightly more conservative in many phonological characters, but it does take part in the same general tendencies (**b d g* > **β δ γ*, voicing of **xt* and **ft*). All of the Pamir languages, including Yidgha-Munji, Wakhi and (in our best trees) Parachi, share the generalisation and reduction of *-*nti* > *-*n* and most importantly deictic systems derivable from the three-term Shughni inventory. Shughni, obviously, preserves this state, whereas it is reduced in the other languages: Wakhi-Yigha-Munji and Sarikoli-Ishkashimi preserve the obl. stems of D1, D2 and D3. Parachi and Yazghulami do not have three-term systems: Yazghulami continues at most the obl. of D2 and both D3-stems (but see **2.3.7.**).

This is one of the innovations that causes Sarikoli to form a clade with the Shughni-group, and Parachi to be grouped with the Pamir languages instead of with Ormuri. Although their deictic systems look almost identical, the pronouns are not traced back to the same stems.

4.3.8. Sogdian, Yaghnobi and Ossetic

Remarkably, all our best trees – and most consensus trees – find Yaghnobi to be a sister of Ossetic rather than Sogdian as usually stated. It is important to stress that our analysis relies heavily on the isoglosses presented by Wendtland 2009. This paper is not particularly concerned with the relationship between Yaghnobi and Sogdian, and Wendtland also addresses the fact that Sogdian shares isoglosses with Shughni. It is possible that important additional isoglosses would change the picture of our analysis, but Wendtland – and therefore we – do treat most of the arguments usually presented in favour of this subgroup (Skjærvø 1989a: 375; Sims-Williams 1996: 650).[33]

[33] The only innovations in favour of a Sogdian-Yaghnobi relationship listed there we have not included are the present-future stems in -*išt*- and the use of the pronoun *x*- as a copula.

Surprisingly, the only isogloss exclusive to Eastern and Western Yaghnobi is their deictic pronominal system. It could be stressed here that while there are – important – differences between the two dialects of Yaghnobi, no innovations are shared between only of them and Sogdian or Ossetic. It is indeed worth noting that Sogdian and Yaghnobi share no exclusive innovations in the summarising table of Wendtland (2009: 185). Even under the traditional assumption, there are grave differences: In phonology, Yaghnobi is either seen as more conservative (preserving **-θ-, *θr-, *-θr-, xt, (*)ft* (Skjærvø 1989a: 375), or it is seen as having undergone backmutations developing from the Sogdian state of affairs (Sims-Williams 1996: 650; Xromov 1972: 123–7). In morphology, the generalisation of *-*nti* in Sogdian but *-*Vr* in Yaghnobi for the 3PL is a further divide between the two.

The similarities in the pronominal system as presented by Skjærvø (1989a: 375) are noteworthy, but the distribution and function of the deictic stems are in no way unique to these languages – the Sogdian system being equally "ancestral" to the system of all Pamir languages, Chorasmian and Parachi as it is to Yaghnobi. Among the unique states shared by Sogdian and Shughni are the innovation of circumfixed demonstrative stems (**m17–19**). No other language takes part in this innovation. While much can be said about our specific coding of the isogloss and weighting of the characters, it is a striking and unique feature of these languages. Sogdian and Wakhi share the novel plural suffix *-*išt*-.

While Sogdian and Yaghnobi only share a single state exclusively (the plural paradigm DIR *-*t*, obl. *-*ti* of the Sogdian heavy stems) in our dataset, Yaghnobi and Ossetic share the agglutinative nature of the plural suffix *-*t*- to the exclusion of Sogdian. However, this detail is also more a matter of coding than of linguistic reality. Sogdian also shows agglutinative tendencies (Kim 2025: 4; Sims-Williams 1982), but because of the system of the ancient letters, these cannot have arisen by the latest shared state of Sogdian and any other language. If we had instead chosen to view the innovation as a binary, Sogdian, Ossetic and Yaghnobi would share the same state. This would follow the same pattern as the 2PL pronoun, where Sogdian, Yaghnobi and Ossetic share an exclusive state (generalisation of **šmāxam*). This finding is different form the usual view that Yaghnobi is a direct continuation of a Sogdian dialect.

4.3.9. Chorasmian, Bactrian and Pashto?

Our weighted analyses somewhat consistently find Chorasmian and Bactrian on a branch with Pashto and Ormuri. In some trees, Bactrian and Chorasmian form a clade. Chorasmian and Bactrian only share one state exclusively, and it is clearly a historical coincidence that only these two, and not Sogdian as with the other prepositions, show fused forms of **abi* and the 1/2PL pronoun. The inferred Bactro-Chorasmian clade would be supported by the outcomes of **xt* > *γt* and **vd* (from **ft*) > *βδ*. It would also be compatible with the changes **č b d g* > **c β δ γ* inferred one step higher up in the tree. They also share almost the same nominal plural endings: Both have created a two-case system with DIR *-*a̯iah*, obl. *-*ānām/īnām*. Chorasmian continues both *-*ānām* and *-*īnām*, Bactrian only *-*ānām*; and importantly, none of them *-*b*- as opposed Pashto (and Sogdian, outside the inferred clade). They also both take part in remodelling the 2PL pronoun on the 2SG, but not in the same way. This last innovation is the only isogloss that places them as the sister clade of Pashto and Ormuri in Tree 1.

Also non-computational analyses have found a relationship between Pashto, Bactrian, Chorasmian. These form, together with Ishkashimi, the earliest Sprachbund in the analysis of Cheung (2015: 57). A clade consisting of Bactrian and Pashto would, with our dataset, be based on **d* > **δ* > **l*, generalising 3PL *-*nti* and the vague notion of sharing the orthotonic 2SG pronoun in the very different 2PL stems.

As to the perhaps peculiar placement of Pashto-Ormuri with Khotanese, Chorasmian and Bactrian, it is not very well supported, and all inferred innovations are either trivial or shared elsewhere: **ft* > **vd*, **θr*- > **dr*-, *-*θr*- > *-*hr*- and the reduction of the 2PL pronoun to **i̯ūžam*, *⁼u̯ah*. However, the preservation of multiple formants of the nominal plural and the pronominal stem D0 force them together as the outliers not taking part in the innovations of Sogdian and the Pamir languages. Unlike Ormuri, Parachi does not preserve any archaism that forces it into this group and is instead derivable from a clade shaped by the Sogdian states for the software.

4.3.10. The position of Khotanese and the higher-order subgroups

In the two best trees (Tree 1), the first divide is between a clade of Khotanese, Pashto-Ormuri and Bactro-Chorasmian opposing a clade of Sogdo-Pamir and Ossetic-Yaghnobi. The former group rests on weak phonological traits that merely exist in that shape because of the lack of nuance that comes from the App not being able to reconstruct, only infer numbers based on our weights and costs: **ft*, **θr*- and *-*θr*- should have become **u̯d*, **dr*-, *-*hr*-. Most of these fit Khotanese or other conservative branches well but require some frankly impossible changes elsewhere (**dr*- becoming **hr*- without merging with existing **hr*-, for instance). The App also infers the Khotanese state of affairs for the 2PL pronoun, **i̯ūžam*, *⁼u̯ah* – which then requires the extra assumption, which we have built into the allowed directionalities, that not only **šm*- but also **žm*- would become **m* and trigger the replacement of the 2PL stem with that of the 2SG to avoid coalescence with the 1SG. The latter group is inferred because of the low cost of losing morphological innovations again. The App infers that the plurals in *-*tā*, the paradigm *-*t, -ti* and the loss of the demonstrative stem D0. This clearly demonstrates the weakness of our approach.

In Trees 2 and 4, Khotanese is a complete outlier. Four developments are inferred already on the PIr. edge: **b d xt* > **β δ γt* and the innovation of the deictic-demonstrative adverbs formed with the suffix *-*θra* (those of e.g. Khot. *vara*, Bactr. *οαρο* and Sogd. *'wrδ*). Khotanese should then have proceeded to undertake every single innovation independently in the cases where it does not directly continue the PIr. state. The Non-Khotanese branch would have innovated **g* > **γ*, restructured the nominal plural case system (but keeping all formants attested in Sogdian) and innovated the plurals in *-*tā* and *-*išt*-. The morphological inferences again only show the consequences of loss of new formants being relatively low.

In Tree 3, Khotanese, Parachi and Yidgha-Munji form a clade based on the innovation of **βt* (from **ft*) > **u̯t*, **θ* > *h* – and the loss of the preposition **upari*. This is hardly compelling evidence for anything. The "core"-group would share only the generalisation of **šmāxam* as the 2PL pronoun. This is only one isogloss, but in turn a lot more linguistically plausible.

Finally, in Tree 5, the Software finds a clade of Khotanese and Parachi-Ormuri based on six inferred developments that can be dismissed out of hand: **xt* > **t/Ø* and *-*θ*- > *-*h*-, the reduction of the paradigm of the 2PL pronoun two **i̯ūžam*, *⸗u̯ah* – and the loss of the prepositions **hada, *abi* and **upari* underlying the innovations of **m10–12**. The fact that three prepositions are not attested tells nothing about genetic relations. We know that the outcome of **xt* is more complicated in Khotanese (Wendtland 2009: 175, fn. 28), that Parachi has only **t*, and that Ormuri has nothing – or, perhaps, *k*. Thus, what the App infers cannot be remotely true. As coded, it would require a change of **t* > **k* in Ormuri, which is highly implausible. **θ* could have become **h* in a predecessor of Parachi and Khotanese, but it is difficult to fit a development of **θ* > **h* > *y* in Ormuri into the picture since there is no consistent merger with the spurious **h*. The reduction of the pronominal paradigm could strictly speaking be true, it fits the outcome of Khotanese and Parachi, and Ormuri replaced its inherited ending anyway.

In all of these scenarios, Khotanese would have undergone massive innovation and restructuring after its break-off from the other languages. Accordingly, the innovative morphological features shared with Sogdian, Bactrian and Chorasmian (demonstrative adverbs in *-*θra*) and the fused forms of **hača*- would be completely independent. This scenario is, perhaps, not impossible: *-*θra* was an inherited suffix anyway, yet *vara* < **au̯a-θra*- proves that at least this form was lexicalised before the restructuring of the deictic stems in Khotanese. On the other hand, the development of **hača-* into the (fusing) case-like postposition *-jsa* is indeed likely to be independent.

While all of these scenarios are unsatisfactory, the fact remains that Khotanese is quite conservative in its morphology, and that the phonological innovations are either unique or easily repeatable. Thus, with this material, it is difficult to place the language any better.

4.3.11. Comparison with existing suggestions and evaluation of the results

It is immediately clear from the fact that we very consistently find Khotanese to have split off relative early that our analyses do not agree in full with any proposed subgrouping. That said, the position of Khotanese is mostly based on archaisms and unique innovations. Additionally, it is evident that out software would never return trees agreeing fully with the presented previous proposals (see **1.2.**) either because the language samples differ or because the models only depict the first-order subgroups and no subsequent branching of the 17 languages.

With that in mind, we can only compare the higher-order subgroupings with the first three suggested trees. We find none of them.

We never find Parachi and Ormuri to form an outlier group as in (a). We rarely even find them to constitute a subgroup. While they are phonetically conservative when it comes to the stop system, they are morphologically innovative. Ormuri further shares morphological innovations with Pashto, and the Parachi demonstrative system can be derived from that of the Pamir languages. We do not believe that this result is credible.

We also never find the North-Eastern branch as in (b). We consistently find Yaghnobi and Ossetic, not Sogdian and Yaghnobi to form a clade, but as elaborated above, this rests on some perhaps unideal coding choices. These languages do share important morphological innovations (the nominal plurals in **-tā*, the 2PL pronominal base **šmāxam*), but Sogdian also shares innovations with Bactrian and Chorasmian on the one hand (the fused forms of prepositions and the 1/2(SG)) and the Pamir languages (the plurals in **-išt-* with Wakhi, the circumfixed adverbs with Shughni – and the entire deictic system) on the other. The former does indeed resemble the North-Eastern branch of (c). If they were to oppose the remaining languages, these shared developments would either be of P(E)Ir. age and lost elsewhere, or there would be considerable overlap of contact induced spread of morphological features.

A general trait of the suggestions (a), (b) and (c) is that they group all the remaining languages into one major clade. Since we never find the same outgroups as these suggestions, we also cannot find the same major higher-order groups contrasting these. A particular difference is the connection of Khotanese and the Pamir languages. Except for Yidgha-Munji in Tree 3, the software never returns Pamir languages as descendants of the same node as Khotanese.

In the best trees (Trees 1 and 2 (and their variations described in fn. 31), the software does find the Pamir languages to constitute a genetic clade in contrast to what is usually held (e.g. Wendtland 2009: 173). As explained, Parachi tends to be connected to the Pamir languages, but ignoring that for a moment, the result of our best trees resembles a binary-branching version of the models (d) and (e). The main difference is that our best trees group Wakhi and Yidgha-Munji together, and that Ishkashimi belongs to the Shughni-Yazghulami-clade instead of being one of more sisters of it. Some of our best trees (Trees 3–5, not 1–2) place Shughni and not Yazghulami as the outlier in this group, but that only has to do with the innovations shared with Sogdian (making joint loss relatively less costly for Yazghulami-Wakhi-Sarikoli-Ishkashimi).

Although the number of EIr. languages samples is much smaller in the Bayesian study of shared vocabulary (f), and leaving aside the fact that this study finds Ossetic, Bactrian and Khotanese to have broken off before a joint clade of the remaining EIr. languages and the entirety of West Iranian, we can still compare some tendencies. In this model, Ossetic is the first Iranian language to break off after Avestan. Instead, our material embeds it deeper into the EIr. languages because of the shared innovations with Sogdian, Bactrian, Chorasmian and Yaghnobi. In (f), Bactrian and Khotanese form a poorly supported clade which should also have broken off before the remaining East Iranian and all West Iranian languages formed a clade. As we find Khotanese to show mostly archaic and unique traits, and Bactrian to have innovated a great deal with Sogdian and Chorasmian, our trees do not show this clade. As explained above, our analyses rather group Yaghnobi with Ossetic, but there are indeed innovations shared between Sogdian and Chorasmian (and Bactrian). This study does not include most Pamir languages, but it does find a very close relationship between the two it includes, Wakhi and Sarikoli – as a sister of Pashto. Our trees also group the Pamir languages closely, but not with Pashto. It also finds – although poorly supported – Bactrian and Khotanese to be sisters – which we do not. In this study, Sogdian and Yaghnobi form a clade, as also traditionally assumed, and Chorasmian is found to be the closest relative of this branch. Once again, we do not find these clades. Chorasmian rather goes with Bactrian than Sogdian, and we find Sogdian to have close ties with the Pamir languages (more on this below).

Although it was neither our goal to reconstruct an East Iranian proto-language nor to compare the languages with West Iranian, our methodology does infer the existence of a Proto East-Iranian different from Proto-Iranian proper: Since we used Proto-Iranian as a proxy for the ancestral state, and the software infers shared changes on the "Proto-Iranian" edge, it is – given exactly this data input and our costs and weights – more likely that *all* languages underwent the word-initial lenition of **b d* > * *β δ* and possibly **g* > * *γ*, maybe even that **xt* was partially voiced and the local adverbs in *-*θra* could be formed to demonstrative stems. However, this all depends on the position of Khotanese and whether the costs we have assigned to phonetic reversals are indeed plausible.

5. Conclusion

This study had two main objectives: To evaluate the application and the analysis of the Graphic User Interface (the App) *LinguiPhyR* on a set of languages usually considered to not have developed in a tree-like fashion and to test the traditional EIr. isoglosses in a phylogenetic framework.

Regarding the first objective, the shortest answer is that using the App was difficult, and that the performance on the dataset was poor. However, this is not necessarily the fault of the software. The preparation of the dataset is extremely tedious, but it does have the great advantage that all innovations and isoglosses can be directly compared on equal terms. This overarching approach is of immense value no matter the results of the analysis and the specific data format. Nevertheless, the foundational criterion that characters should be completely independent is difficult to meet with phonological and morphological data in general and with data from an already established Sprachbund in particular (cf. Khot *vara* < **au̯a-θra-* in **m17** with the loss of the stem **au̯a-* in **m8**).

The results of the analyses are not overly convincing. There are many reasons for this. First, it was never expected that forcing a Sprachbund into a tree-model would yield credible or consistent results. Second, our dataset consists only of a summary of the positive arguments in favour of various conflicting groupings. Third, our custom-coded directionalities, weights and costs might have had unforeseen negative consequences.

Staying very close to the outcome of the languages rather than coding only shared innovations at a reconstructed state made us able to use layers of innovation for each character just as traditionally done for isoglosses (e.g. **d* > **δ* > *l*). But relying heavily on custom-coded directionalities and differentiated costs introduced new issues. First, in many cases it is unclear by which route an outcome came to be. Often, we can only exclude what is not possible (e.g. deriving the Parachi PL *-an* from *-*b*- or the like). Second, since it is impossible to rely on unattested intermediate states, the inferred developments of the App end up more dubious with such fine-grained data than they would have with fewer macro-states (e.g. the inferred chronology of the innovations of the plurals in *-*t*-). Third, we may have unintentionally forced the software to base some parts of the trees on shared archaisms; in the sense that some (shared) innovations require the retention of some formants (i.e. many changes in **m1**, **m6** and **m8**). This need not be a faulty approach, but it would be necessary to balance these restricted isoglosses – especially the deictic pronominal system – with more material in the future. Fourth, the use of weights to differentiate the significance of different types of isoglosses and the use of costs to differentiate the likelihood of the individual transitions (innovations) simultaneously cannot be recommended. This strategy has caused inferred loss of morphological features to outweigh the even unusual phonetic innovations in our analyses which is problematic at best and devastating at worst.

As to the subgrouping languages of the EIr languages, the results are inconclusive, and we would probably need a larger dataset with more – also trivial – isoglosses and not just positive arguments in favour of conflicting solutions to say anything reliably.

On the other hand, the absence of a single recoverable tree does not prove that all shared features are the result of contact. Although it was not the objective, analysing all isoglosses at once with typologically informed costs confirms that it is likely that some innovations did take place in Proto-East-Iranian (e.g. **b d* > **β δ*, possibly **g* > **γ*, and perhaps even **xt* > **γt*), while others (e.g. **č, -θ-, *θr-, *-θr-*) are much less likely to have taken place at this stage. This, perhaps, calls the reliability of some of the isoglosses underlying some suggested prehistoric Sprachbünde into question. We can safely conclude that there is very little if any consistent phylogenetic information in the shared phonological innovations.

With that in mind, there are some conclusions to be drawn about the subgrouping of EIr. We do not find support for any of the previous suggestions – but this may be because morphological archaisms of Khotanese skew the results. In this data, there is very limited support for a Parachi-Ormuri clade; they mostly share insignificant phonetic archaisms that could be independent backformations, and Parachi is in most of its morphology also derivable from a Sogdian or Pamir system. Instead, our analyses reveal a close relationship between Ormuri and Pashto – which is more likely to be the result of our coding strategies than linguistic reality. Pashto is difficult to place. It shares some tendencies with Bactrian, and some archaic traits with Khotanese, Parachi and Ormuri.

Sogdian plays a special role in the configuration of the tree and must also be quite deeply embedded. Sogdian shares key innovations with three groups – some of which may be genetic, some areal. It shares the 2PL pronoun and the plurals in *-*t*- with Ossetic and Yaghnobi, many fused forms of prepositions and pronouns with Chorasmian, Bactrian and Khotanese – and other fused forms and demonstrative adverbs with Shughni.

Our material does not allow us to disprove the close relationship between Sogdian and Yaghnobi although we never find it; but what is clear is that the outcomes of the Pamir languages are just as derivable from, often even closer to, Sogdian than Yaghnobi is. However, there may be a reverse confirmation bias as we have not systematically included isoglosses in favour of this subgroup. While they are consistent, the evidence in favour of the Pamir languages as a clade, the grouping of Wakhi and Yidgha-Munji as well as the Shughni-Yazghulami group – including Sarikoli, is slender.

In conclusion, while our approach is tedious, and our results inconclusive, we have given a great deal of insight into the issues of custom-coding significance of individual innovations in a complex data set. We hope that this attempt will further the debate on the evolution of the EIr. languages, but we humbly observe that a lot more data is needed to take the next leap forward.

6. References

Bashir, Elena. 2009. Wakhi. In Gernot Windfuhr (ed.), *The Iranian languages*, 825–858. (Routledge Language Family Series). London & New York (NY): Routledge.

Bielmeier, Roland. 1989. Yaghnōbī. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 480–488. Wiesbaden: Reichert.

Bogoljubov, Mixail Nikolaevič. 1962. Ličnye mestoimenija v xorezmijskom jazyke. *Učenye zapiski Leningradskogo ordena Lenina i ordena Trudovogo Krasnogo Znameni gosudarstvennogo universiteta imeni Andreja Aleksandroviča Ždanova. Serija vostokovedčeskich nauk.*

Canby, Marc E., Steven N. Evans, Donald Ringe & Tandy Warnow. 2024. Addressing Polymorphism in Linguistic Phylogenetics. *Transactions of the Philological Society* n/a(n/a). 1–32. doi:10.1111/1467-968X.12289.

Canby, Marc Edward. 2024. LinguiPhyR: A Package for Linguistic Phylogenetic Analysis in R. *Journal of Open Source Software* 9(102). 6201. doi:10.21105/joss.06201.

Cheung, Johnny. 2015. The linguistic and geographic position of Pashto within the east iranian language group: 1. A phonological and lexical analysis. *Iranian*

Languages and Literatures of Central Asia. From the Eighteenth Century to the Present, 29–65. (Cahiers de Studia Iranica 57). Leuven: Peeters.

Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. (Publications of the Philological Society 30). Oxford & Cambridge (MA): Blackwell.

Clackson, James. 2022. Methodology in Linguistic Subgrouping. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 18–32. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.002 (23 September, 2022).

Del Tomba, Alessandro. 2024. Sound Change or Analogy? Palatalisation in the Plural Paradigm of the Khotanese i-declension. *Acta Orientalia Academiae Scientiarum Hungaricae* 77(2). Akadémiai Kiadó. 253–264. doi:10.1556/062.2024.00396.

Durkin-Meisterernst, Desmond. 2009. Khwarezmian. In Gernot Windfuhr (ed.), *The Iranian Languages*, 336–376. (Routledge Language Family Series). London & New York: Routledge.

Ėdel’man, Džoj Iosifovna. 1990. *Sravnitel'naja grammatika vostochnoiranskikh jazykov: Morfologija, ėlementy sintaksisa*. (Ed.) Vasilij Ivanovič Abaev. Moskva: Nauka.

Ėdel’man, Džoj Iosifovna. 2000. Xorezmijskij jazyk. In Valentina S. Rastorgueva, Džoj Iosifovna Ėdel’man & V. V. Moškalo (eds.), *Jazyki mira*, vol. 7. Iranskije jazyki. 3. Vostočnoiranskije jazyki, 95–104. Moskva: Indrik.

Ėdel’man, Džoj Iosifovna. 2008. Xorezmijskij jazyk. In Valentin Aleksandrovič Efimov (ed.), *Osnovy iranskogo jazykoznanija: Sredneiranskie i novoiranskie jazyki*, 6–61. Moskva: Vostočnaja Literatura.

Edelman, Joy I. & Leila R. Dodykhudoeva. 2009a. The Pamir languages. In Gernot Windfuhr (ed.), *The Iranian languages*. (Routledge Language Family Series). London & New York (NY): Routledge.

Edelman, Joy I. & Leila R. Dodykhudoeva. 2009b. Shughni. In Gernot Windfuhr (ed.), *The Iranian Languages*, 787–824. (Routledge Language Family Series). London & New York: Routledge.

Efimov, Valentin A. 1991. Ormuri. *Osnovy iranskogo jazykoznanija: Novoiranskie jazyki*, 247–315. Moskva: Nauka.

Efimov, Valentin Aleksandrovič. 2011. *The Ormuri language in past and present*. (FLI Language and Culture Series 6). Islamabad: Forum for Language Initiatives.

Emmerick, Ronald. 1989. Khotanese and Tumshuqese. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 204–229. Wiesbaden: Reichert.

Emmerick, Ronald. 2009. Khotanese and Tumshuqese. In Gernot Windfuhr (ed.), *The Iranian languages*, 377–416. (Routledge Language Family Series). London & New York (NY): Routledge.

Emmerick, Ronald E. 2024. *A Handbook of Khotanese, completed and edited by Mauro Maggi, John S. Sheldon, and Nicholas Sims-Williams*. (Beiträge Zur Iranistik 51). Wiesbaden: Reichert.

Gercenberg, L. G. 1981. Xotansakskij jazyk. *Osnovy iranskogo jazykoznanija: Sredneiranskie jazyki*, 233–313. Moskva: Nauka.

Gershevitsch, Ilya. 1961. *A grammar of Manichean Sogdian*. (Publications of the Philological Society). Oxford: Blackwell.

Gholami, Saloumeh. 2009. Ergativity in Bactrian. *Orientalia Suecana* 58. 132–141.

Gholami, Saloumeh. 2014. *Selected Features of Bactrian Grammar*. Harrassowitz. https://www.jstor.org/stable/j.ctvc2rmbs.9 (23 May, 2024).

Goldstein, David M. 2022. There's no escaping phylogenetics. In Laura Grestenberger, Charles Reiss, Hannes A. Fellner & Gabriel Z. Pantillon (eds.), *Ha! Linguistic studies in honor of Mark R. Hale*, 71–91. Wiesbaden: Reichert.

Greenhill, Simon J. & Russell D. Gray. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica* 29(4). 523–537. doi:10.1075/dia.29.4.05gre.

Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2021. Bayesian Phylolinguistics. In Richard D. Janda, Joseph, Brian D. & Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, 226–253. John Wiley & Sons. doi:10.1002/9781118732168.ch11 (28 September, 2021).

Grjunberg, Aleksandr Leonovič & Džoj Iosifovna Ėdel'man. 1987. Afganskij jazyk. In Vasilij Ivanovič Abaev, Valentina S. Rastorgueva & Mixail Nikolaevič Bogoljubov (eds.), *Osnovy iranskogo jazykoznanija: Novoiranskie jazyki*, vol. 2. Vostočnaja gruppa, 6–155. Moskva: Nauka.

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, et al. 2023a. Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* 381(6656). eabg0818. doi:10.1126/science.abg0818.

Heggarty, Paul, Cormac Anderson, Matthew Scarborough, et al. 2023b. Supplementary materials for "Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages." *Science* 381(6656). eabg0818. doi:10.1126/science.abg0818.

Henning, Walter Bruno. 1958. Mitteliranisch. In Bertold Spuler, H. Franke, J. Gonda, et al. (eds.), *Iranistik*, vol. 4. Linguistik, 20–130. (Handbuch Der Orientalistik 1. Abteilung). Leiden: Brill.

Hoenigswald, Henry M. 1966. Criteria for the subgrouping of languages. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects: Proceedings of the Conference on Indo-European Linguistics, held at the University of California, Los Angeles, April 25-27, 1963*, 1–12. Berkeley (CA) & Los Angeles (CA): University of California Press.

Humbach, Helmut. 1989. Choresmian. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 193–203. Wiesbaden: Reichert.

Jacques, Guillaume & Johann-Mattis List. 2019. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1). 128–167. doi:10.1075/jhl.17008.mat.

Jamison, Rachel Elizabeth. 2022. The enclitic =*da* and the marking of indicative and subjunctive mood in Yazghulami. Dallas: Dallas International University MA-thesis. https://diu.edu/academics/student-theses/.

Kieffer, Charles. 1989. Le parāčī, l'ōrmuṛī et le groupe des langues iraniennes du Sud-Est. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 445–455. Wiesbaden: Reichert.

Kieffer, Charles. 2003. *Grammaire de l'ōrmuṛī de Baraki-Barak (Lōgar, Afghanistan)*. Vol. 22. (Beiträge zur Iranistik). Wiesbaden: Reichert.

Kieffer, Charles. 2009. Parachi. In Gernot Windfuhr (ed.), *The Iranian Languages*, 693–720. (Routledge Language Family Series). London & New York: Routledge.

Kieffer, Charles M. 1978. Études Parāči 3a. Notes sur la phonologie et morphosyntaxe du parāči. *Studia Iranica* 7(1). 81–107.

Kim, Deborah. 2017. Topics in the syntax of Sarikoli. Leiden: Universiteit Leiden PhD-dissertation (Proefschrift). https://www.lotpublications.nl/topics-in-the-syntax-of-sarikoli (21 July, 2023).

Kim, Ronald I. 2023. The Prehistory of Ossetic Verbal Inflection (I): Present Indicative and Imperative. In David M. Goldstein, Stephanie W. Jamison & Brent Harmon Vine (eds.), *Proceedings of the 32nd Annual UCLA Indo-European Conference*, 153–71. Hamburg: Buske.

Kim, Ronald I. 2024. The Khotanese instrumental-ablative and locative plural endings. Paper presented at the 43rd East Coast Indo-European Conference (ECIEC), University of Georgia, Atlanta (GA). https://linguistics.uga.edu/sites/default/files/2024-07/Kim_ECIEC43.pdf (4 February, 2025).

Kim, Ronald I. 2025. Case Marking from Old Iranian to Ossetic: A Comparative and Typological Approach. *Transactions of the Philological Society*. doi:10.1111/1467-968X.12316. https://onlinelibrary.wiley.com/doi/10.1111/1467-968X.12316 (17 March, 2025). Online Version of Record.

Korn, Agnes. 2016. A partial tree of Central Iranian: A new look at Iranian subphyla. *Indogermanische Forschungen* 121(1). 401–434. doi:10.1515/if-2016-0021.

Korn, Agnes. 2019. Isoglosses and subdivisions of Iranian. *Journal of Historical Linguistics* 9(2). 239–281. doi:10.1075/jhl.17010.kor.

Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.

Kümmel, Martin Joachim. 2018. The morphology of Indo-Iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 3, 1888–1924. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter Mouton. doi:10.1515/9783110542431-032. https://www.degruyter.com/document/doi/10.1515/9783110542431-032/html (26 August, 2024).

Livshitz, Vladimir Aronovich. 1970. A Sogdian alphabet from Panjikant. In Mary Boyce & Ilya Gershevitch (eds.), *W. B. Henning memorial volume*, 256–263. London: Lund Humphries.

Livšic, Vladimir A. & Al'bert Leonidovič Xromov. 1981. Sogdijskij jazyk. In Vasilij Ivanovič Abaev, Mixail Nikolaevič Bogoljubov & Valentina S. Rastorgueva (eds.), *Osnovy iranskogo jazykoznanija: Sredneiranskie jazyki*, 347–514. Moskva: Nauka.

MacKenzie, D. Neil. 1990. *The Khwarezmian Element in the Qunyat al-munya*. London: SOAS.

Morgenstierne, Georg. 1929. *Indo-Iranian frontier languages I: Parachi & Ormuri*. (Instituttet for Sammenlignende Kulturforskning. Serie B: Skrifter 11). Oslo: Aschehoug. http://archive.org/details/in.gov.ignca.14416 (31 July, 2023).

Morgenstierne, Georg. 1938. *Indo-Iranian frontier languages II: Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi*. (Instituttet for Sammenlignende Kulturforskning. Serie B: Skrifter XXXV). Oslo, Norvège. https://archive.org/details/dli.pahar.2693/.

Morgenstierne, Georg. 1942. Archaisms and innovations in Pashto morphology. *Norsk Tidsskrift for Sprogvidenskap* 12(2). 88–114.

Morgenstierne, Georg. 1974. *Etymological vocabulary of the Shughni group*. (Beiträge zur Iranistik 6). Wiesbaden: Reichert.

Morgenstierne, Georg. 2003. *A New Etymological Vocabulary of Pashto*. (Beiträge Zur Iranistik 23). Wiesbaden: Reichert.

Nakhleh, Luay, Tandy Warnow, Don Ringe & Steven N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Transactions of the Philological Society* 103(2). 171–192.

Olander, Thomas. 2018. Connecting the dots: The Indo-European family tree as a heuristic device. In David M. Goldstein, Stephanie W. Jamison & Brent Vine (eds.), *Proceedings of the 29th annual UCLA Indo-European conference*, 181–202. (Journal of Indo-European Studies Monograph Series). Bremen: Hempen.

Parker, Clinton. 2023. A grammar of the Shughni language. Montréal: McGill University Doctoral dissertation. https://escholarship.mcgill.ca/concern/theses/0v8385834.

Paxalina, Tat'jana N. 1959. *Iškašimskij jazyk: očerk fonetiki i grammatiki, teksty i slovar'*. Moscou: Izdatel'stvo Akademii Nauk SSSR.

Paxalina, Tat'jana N. 1987a. Vaxanskij jazyk. In Vasilij Ivanovič Abaev, Valentina S. Rastorgueva & Mixail Nikolaevič Bogoljubov (eds.), *Osnovy iranskogo jazykoznanija: Novoiranskie jazyki*, vol. 2. Vostočnaja gruppa, 408–473. Moskva: Nauka.

Paxalina, Tat'jana N. 1987b. Iškašimskij jazyk. In Vasilij Ivanovič Abaev, Valentina S. Rastorgueva & Mixail Nikolaevič Bogoljubov (eds.), *Osnovy iranskogo jazykoznanija: Novoiranskie jazyki*, vol. 2. Vostočnaja gruppa, 474–536. Moskva: Nauka.

Payne, John. 1989. Pāmir languages. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 417–444. Wiesbaden: Reichert.

Pellard, Thomas, Robin J. Ryder & Guillaume Jacques. *in press*. The Family Tree model. In Adam Ledgeway, Edith Aldridge, Anne Breitbarth, et al. (eds.), *The Wiley Blackwell companion to diachronic linguistics*. Wiley-Blackwell. https://hal.science/hal-04571512v1/file/The_Family_Tree_Model.pdf (8 May, 2024). Preprint.

Rastorgueva, Vera Sergeevna & Džoj Iosifovna Ėdel'man. 2000. *Ėtimologičeskij slovar' iranskix jazykov*. Vol. 1. a–ā. Moskva: Vostočnaja literatura.

Rastorgueva, Vera Sergeevna & Džoj Iosifovna Ėdel'man. 2007. *Ėtimologičeskij slovar' iranskix jazykov*. Vol. 3. f-h. Moskva: Vostočnaja literatura.

Ringe, Donald A. 2022. What We Can (and Can't) Learn from Computational Cladistics. In Thomas Olander (ed.), *The Indo-European Language Family: A Phylogenetic Perspective*, 52–62. Cambridge: Cambridge University Press. doi:10.1017/9781108758666.004 (27 September, 2022).

Ringe, Donald A., Tandy Warnow & Ann A. Taylor. 2002. Indo-European and computational linguistics. *Transactions of the Philological Society* 100(1). 59–129.

Robson, Barbara & Habibullah Tegey. 2009. Pashto. In Gernot Windfuhr (ed.), *The Iranian Languages*, 721–772. (Routledge Language Family Series). London & New York: Routledge.

Scarborough, Matthew Joseph Charles. 2020. Cognacy and computational cladistics: Issues in determining lexical cognacy for Indo-European cladistic research. In Thomas Olander & Matilde Serangeli (eds.), *Dispersals and diversification: Linguistic and archaeological perspectives on the early stages of Indo-European*, 179–208. (Brill's Studies in Indo-European Languages and Linguistics). Leiden & Boston (MA): Brill.

Sims-Williams, Nicholas. 1981. The Sogdian sound system and the origin of the Uyghur script. *Journal Asiatique* 269. 347–360.

Sims-Williams, Nicholas. 1982. The double system of nominal inflectionin Sogdian. *Transactions of the Philological Society* 80(1). 67–76.

Sims-Williams, Nicholas. 1989a. Eastern Middle Iranian. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 165–172. Wiesbaden: Reichert.

Sims-Williams, Nicholas. 1989b. Sogdian. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 173–192. Wiesbaden: Reichert.

Sims-Williams, Nicholas. 1991. A Sogdian greeting. In Dieter Weber & Ronald Eric Emmerick (eds.), *Corolla Iranica: papers in honour of Prof. Dr. David Neil MacKenzie on the occasion of his 65th birthday on April 8th, 1991*, 176–187. Frankfurt am Main: Peter Lang.

Sims-Williams, Nicholas. 1994. The Triple System of Deixis in Sogdian. *Transactions of the Philological Society* 92(1). 41–53. doi:10.1111/j.1467-968X.1994.tb00427.x.

Sims-Williams, Nicholas. 1996. Eastern Iranian languages. *Encyclopædia Iranica* VII. 649–652. doi:10.1163/2330-4804_EIRO_COM_8633.

Sims-Williams, Nicholas. 2007. *Bactrian documents from northern Afghanistan. 2. Letters and Buddhist texts*. (Studies in the Khalili Collection 3. Corpus Inscriptionum Iranicarum. 2. Inscriptions of the Seleucid and Parthian Periods and of Eastern Iran and Central Asia. 4. Bactrian). Oxford & New

York (NY): The Nour Foundation in association with Azimuth Editions & Oxford University Press.

Sims-Williams, Nicholas. 2011a. Remarks on the phonology of the Manichaean Bactrian fragment (M 1224). In Elena K. Molčanova (ed.), *Leksika, ėtimologija, jazykovye kontakty: K jubileju doktora filologičeskix nauk professora Džoj Iosifovny Ėdel'man*, 244–251. Moscow: Tezaurus.

Sims-Williams, Nicholas. 2011b. Differential object marking in Bactrian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in Iranian Linguistics*, 23–38. (Beiträge Zur Iranistik 34). Wiesbaden: Reichert.

Sims-Williams, Nicholas. 2021. Plural and Collective in the Sogdian Ancient Letters. *Transactions of the Philological Society* 119(2). 174–181. doi:10.1111/1467-968X.12211.

Sims-Williams, Nicholas. 2023. *The "Ancient Letters" and other early Sogdian documents and inscriptions*. (Corpus Inscriptionum Iranicarum. 2. Inscriptions of the Seleucid and Parthian Periods and of Eastern Iran and Central Asia. 3. Sogdian 5). London: School of Oriental and African Studies.

Skjærvø, Prods O. 2022. Khotan 4. The Khotanese language. In Elton L. Daniel (ed.), *Encyclopaedia Iranica*, vol. 17, 117–133. Leiden: Brill. https://brill.com/edcollbook/title/62996 (4 February, 2025).

Skjærvø, Prods Oktor. 1989a. Modern East lranian languages. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 370–383. Wiesbaden: Reichert.

Skjærvø, Prods Oktor. 1989b. Yidgha and Munǰī. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 411–416. Wiesbaden: Reichert.

Skjærvø, Prods Oktor. 2017. The morphology of Iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 1, 503–549. (Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science 41). De Gruyter Mouton. doi:10.1515/9783110261288-034. https://www.degruyter.com/document/doi/10.1515/9783110261288-034/html (26 August, 2024).

Sokolova, Valentina Stepanovna. 1966. Šugnano-rušanskaja jazykovaja gruppa. In Vinogradov, Viktor Vladimirovič (ed.), *Indoevropejskie jazyki*, 362–397. (Jazyki Narodov SSSR 1). Moskva: Nauka.

Steblin-Kamenskij, Ivan M. 1999. *Etimologičeskij slovar' vaxanskogo jazyka*. (Peterburgskoe Vostokovedenie). St. Petersburg: St-Peterburgskij Gosudarstvennyj Universitet.

Steblin-Kamenskij, Ivan Mixailovič. 1971. Istoričeskij fonetika vaxanskogo jazyka. Avtoreferat. Sankt-Peterburg: St-Peterburgskij Gosudarstvennyj Universitet PhD dissertation.

Thordarson, Fridrik. 1989. Ossetic. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 456–479. Wiesbaden: Reichert.

Tremblay, Xavier. 2005. Bildeten die iranischen Sprachen ursprünglich eine genetische Familie oder einen Sprachbund innerhalb des indo-iranischen Zweiges? Beiträge zur vergleichenden Grammatik der iranischen Sprachen. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel: Akten der XI. Fachtagung der Indogermanischen Gesellschaft, 17.–23. September 2000, Halle an der Saale*, 673–688. Wiesbaden: Reichert.

Vinogradova, S. P. 2000. Sogdijskij jazyk. In Valentina S. Rastorgueva, Džoj Iosifovna Ėdel'man & V. V. Moškalo (eds.), *Jazyki mira*, vol. 7. Iranskije jazyki. 3. Vostočnoiranskije jazyki, 58–94. Moskva: Indrik.

Wendtland, Antje. 2006. Deixis im Sogdischen oder: warum wird man *wdγ* („dort") geboren und stirbt *wδ'γδ* („dort")? In Heiner Eichner, Bert G. Fragner, Velizar Sadovski, et al. (eds.), *Iranistik in Europa, gestern, heute, morgen*, 241–259. (Veröffentlichungen Zur Iranistik 34). Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Wendtland, Antje. 2009. The Position of the Pamir Languages within East Iranian. (Ed.) Agnes Korn. *Orientalia Suecana* 58. 172–188.

Wendtland, Antje. 2011. *Die Entwicklung von Demonstrativpronomen zu Artikeln im Soghdischen*. 1st ed. Harrassowitz Verlag. doi:10.2307/j.ctvbqs61b. https://www.jstor.org/stable/j.ctvbqs61b (17 September, 2024).

Xromov, Al'bert Leonidovič. 1972. *Jagnobskij jazyk*. Moskva: Nauka.

Yoshida, Yutaka. 2009. Sogdian. In Gernot Windfuhr (ed.), *The Iranian languages*, 279–335. (Routledge Language Family Series). London & New York (NY): Routledge.

7. List of appendices

Unmarked appendices can be found in the print and pdf versions, appendices marked with an asterisk (and a file typed in brackets) are (only) available here:[34]

https://doi.org/10.5281/zenodo.15044575

- Appendix 1: The data matrix
    - *1a: The data matrix proper (.csv)
- Appendix 2: The data matrix with outcomes
- Appendix 3: List of all states
- Appendix 4: Condensed list of allowed directionalities and costs of transitions
- Appendix 5: Costs of all allowed directionalities
    - 5a: phonological characters
    - 5b: morphological characters
- *Appendix 6: The 33 best trees of the final run (.trees)
- Appendix 7: A sample of the best trees of the final run
- Appendix 8: Inferred development of all characters on Tree 1 and Tree 2
- Appendix 9: A sample of the best trees of the alternative analyses
- *Appendix 10: Data matrices and best trees of the alternative analyses (.zip)

[34] Click or copy the following string to access https://zenodo.org/records/15044575?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImU5MTEzMTljLTk2ZTAtNGNlNy1hMWY4LThlODgwZWZlNzIzNCIsImRhdGEiOnt9LCJyYW5kb20iOiJkODJlYjYzZmY2MjNmNzU5MmZmZWQ2NjZiOGRlYmQxMCJ9.raHZk2xbAcFKg1L0wEyz36WHlkZkj4TvFybcjiDBkrBSE6wV4CL832TJ5CqNkQ2GGQMhg-NUfWzE9KV8L54i6w

id	feature	weight	chartype	Plr.	Bactrian	Chorasmian	Ishkashimi	Khotanese	Munji	Ormuri	Ossetic	Parachi	Pashto	Sarikoli	Shughni	Sogdian	Wakhi	Yaghnobi_W	Yaghnobi_E	Yazghulami	Yidgha
p1	*č-	1	custom: 0>1(c:1); 0>2(c:1); 1	0	2	2	2	1	0	2	2	0	2	2	2	0	2	0	0	0	2
p2	*b-	1	custom: 0>1(c:2); 1>0(c:1)	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1
p3	*d-	1	custom: 0>1(c:2); 0>2(c:3); 1	0	2	1	0	1	2	0	0	0	2	1	1	1	1	0	0	1	2
p4	*g-	1	custom: 0>1(c:2); 1>0(c:1)	0	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1
p5	*xt	1	custom: 0>1(c:2); 0>2(c:2); 0	0	2	2	2	4	2	5	2	4	6	3	3	1	2	0	2	2	2
p6	*ft	1	custom: 0>1(c:2); 0>2(c:3); 0	0	2	2	2	3	3	6	3	4	6	3	3	1	5	0	3	3	3
p7	*-θ-	1	custom: 0>1(c:2); 0>2(c:2); 0	0	4	0	6	4	3	6	1	4	5	0	0	0	?	1	2	0	3
p8	*θr-	1	custom: 0>1(c:2); 0>2(c:4); 0	0	9	6	10	4	7	6	2	6	4	10	10	3	1	1	8	5	7
p9	*-θr-	1	custom: 0>1(c:1); 0>2(c:2); 0	0	8	9	6	9	9	5	2	5	9	4	4	3	1	2	7	4	9
m1	nominal plural endings	2	custom: 0>1(c:1); 0>2(c:4); 0	0	4	3	11	0	9	6	?	7	2	10	8	1	10	?	?	5	9
m2	PL < abstract suffix *-tā-	3	custom: 0>1(c:4); 1>0(c:1); 1	0	0	0	0	0	0	0	2	0	0	0	0	1	0	2	2	3	0
m3	PL DIR -t, OBL -ti	1	custom: 0>1(c:3); 1>0(c:1)	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0
m4	Neo-Plural	3	custom: 0>1(c:5); 0>2(c:5); 1	0	0	0	0	0	0	0	0	0	0	2	0	1	1	0	0	0	0
m5	inherited 3PL-endings	2	custom: 0>1(c:3); 0>2(c:1); 0	0	3	1	4	0	5	4	2	4	4	4	4	3	4	1	1	4	5
m6	analogical 3PL-ending	2	custom: 0>1(c:3)	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
m7	pronoun 2PL	3	custom: 0>1(c:3); 0>2(c:1); 0	0	7	5	7	1	2	8	3	4	6	7	7	3	9	3	3	7	2
m8	deictic pronouns	2	custom: 0>1(c:4); 0>2(c:3); 0	0	2	11	7	1	7	3	8	9	4	7	6	5	7	12	12	10	7
m9	adposition *hača + pron. 1/2 person	2	custom: 0>1(c:3); 0>2(c:3); 0	0	1	1	0	3	2	4	0	0	2	0	0	1	0	0	0	2	0
m10	adposition *hada + pron. 1/2 person	1	custom: 0>1(c:3); 0>2(c:3); 0	0	1	1	3	3	2	3	0	3	0	3	3	1	0	3	3	3	0
m11	adposition *abi + pron. 1/2 person	1	custom: 0>1(c:3); 0>2(c:3); 0	0	1	1	3	3	2	3	3	3	3	3	0	0	3	3	3	0	0
m12	adposition *upari + pron. 1/2 person	1	custom: 0>1(c:3); 0>2(c:1); 1	0	2	1	2	2	2	2	2	2	2	0	0	1	0	0	0	2	2
m13	affixed dem:*haca-DEM (+ suffix)	3	custom: 0>1(c:3); 0>2(c:3); 1	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0
m14	affixed dem:productivity of the patternadposition + DEM-*antara-	1	custom: 0>1(c:3); 1>0(c:1)	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
m15	affixed dem:productivity of the patternadposition + DEM + suffix	1	custom: 0>1(c:3); 1>0(c:1)	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
m16	local adverbs with suffix:DEM + *arda-	3	custom: 0>1(c:2); 1>0(c:1)	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
m17	local adverbs with suffix:DEM-*θra-	1	custom: 0>1(c:2); 1>0(c:1)	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
m18	local adverbs with suffix:DEM-*da-	1	custom: 0>1(c:2); 1>0(c:1)	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
m19	local adverbs with suffix:DEM-*da-aida- (?)	2	custom: 0>1(c:3); 1>0(c:1)	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

id	feature	weight	Proto-Iranian		Bactrian		Chorasmian		Ishkashimi		Khotanese		Munji		Ormuri	
p1	*č-	1	0	č	2	c	2	c	2	c	1	č/c	0	č	2	c
p2	*b-	1	0	b	1	β	1	β	1	β	1	β	1	β	0	b
p3	*d-	1	0	d	2	l	1	δ	0	d	1	δ	2	l	0	d ø
p4	*g-	1	0	g	1	γ	1	γ	1	γ	0	g	1	γ	0	g
p5	*xt	1	0	xt	2	γδ	2	γδ	2	γδ	4	t	2	γδ	5	k,
p6	*ft	1	0	ft	2	βδ	2	βδ	2	βδ	3	ud	3	vd	6	u
p7	*-θ-	1	0	θ	4	h	0	θ	6	y	4	h	3	x̌	6	y
p8	*θr-	1	0	θr	9	hVr	6	š	10	r	4	dr	7	x̌Vr	6	š
p9	*-θr-	1	0	θr	8	hr	9	r	6	sVr	9	r	9	r	5	š
m1	nominal plural endings	2	0	NOM-ACC *-ayah, GEN *-ānām, *-īnām; DAT *-byah, INSTR *-biš, LOC *-su, *-šu, stem classes and ablaut.	4	DIR -ε, OBL -ανο	3	DIR -i etc., GEN -ān; -(i)na	11	PL -o	0	NOM -ä, -a G-D -ānu, -änu IN-AB -yau LOK -uvo' etc.	9	DIR -i OBL -āf	6	PL -i, -īn, -ān, etc.
m2	PL < abstract suffix *-tā-	3	0	no	0	no	0	no	0	no	0	no	0	no	0	no
m3	PL DIR -t, OBL -ti	1	0	no	0	no	0	no	0	no	0	no	0	no	0	no
m4	Neo-Plural	3	0	no	0	No	0	No	0	No	0	No	0	No	0	no
m5	inherited 3PL-endings	2	0	*-nt(i)~-Vr	3	-ινδο	1	-ri	4	-on	0	-īndä / -āre	5	-āt	4	-in
m6	analogical 3PL-ending	2	0	no	0	No	0	No	0	No	0	No	0	No	1	-e
m7	pronoun 2PL	3	0	Nom. *yūžam, Gen. *šmāxam, Dat. *yušmabya, clitic *=wah	7	τωμαχο, τομαχο, ταμαχο	5	hßy	7	tьmьx	1	NOM uhu, encl. ū etc.	2	mof, encl. -əfon	8	tyos, tos
m8	deictic pronouns	2	0	D0D *ha- / D0O ta- D1D *aya- / D1O *a-, *ima- D2D *aisha- / D2O *aita- D3D *haw- / D3O *awa-	2	uncl.: το, τι ειο/ειμο ειδο	11	nyn, n'n, hy nyš/nyd n'w	7	am, ma ad, da aw, wa	1	ttā-, ttu- ṣä-, ṣā-	7	ma- ya- wa-	3	a- hō, afō
m9	ADP *hača + pron. 1/2 person	2	0	*hača + pron.1/2 person	1	SG 1 ασα/ο-μαγο, 2 ασ(α/ο)-φαγο; 1PL ασ-αμαχο (2PL unatt.)	1	c-, -c ; SG 1 m'-c(y) etc., c-m(y)k, 2 tw'-c, -β/f'-c etc., c-fyk; PL 1 mn'-c	0	cə	3	-jsa	2	1PL žāmox, 2PL žāmof (no SG)	4	
m10	ADP *hada + pron. 1/2 person	1	0	*hada + pron.1/2 person	1	SG 1 αλ(α/ο)-μαγο, 2 αλ(α/ο)-φαγο, (PL unatt.)	1	θ-, -θ; SG 1 -m'-θ, 2 -f'-θ	3		3		2	1PL dāmox, 2PL dāmof (no SG)	3	
m11	ADP *abi + pron. 1/2 person	1	0	*abi + pron.1/2 person	1	SG 1 αβα/ο-μαγο, 2 αβ(α/ο)-φαγο, PL 1 αβ-αμαχο (2PL unattested)	1	f- ; SG 1 f-myk, 2 f-fyk (PL unattested)	3		3		2	1PL vāmox, 3PL vāmof (no SG)	3	
m12	ADP *upari + pron. 1/2 person	1	0	*upari + pron.1/2 person	2		1	pr- ; SG 1 pr-m(y)k, 2 pr-fyk (PL unattested)	2		2		2		2	
m13	affixed dem:*haca-DEM (+ suffix)	3	0	no	0	no	0	no	0	no	0	no	0	no	0	no
m14	affixed DEM: prod. of the pattern ADP + DEM-*antara-	1	0	no	0	no	0	no	0	no	0	no	0	no	0	no
m15	affixed DEM: prod. of the pattern ADP + DEM + suffix	1	0	no	0	no	0	no	0	no	0	no	0	no	0	no
m16	local adverbs with suffix: DEM + *arda-	3	0	no	0	no	0	no	0	no	0	no	0	no	0	no
m17	local adverbs with suffix: DEM-*θra-	1	0	no	1	μαρο, ταρο, οαρο	0	no	0	no	1	mara, ttara, vara	0	no	0	no
m18	local adverbs with suffix: DEM-*da-	1	0	no	1	μαλο, ταλο, οαλο	0	no	0	no	0	no	0	no	0	no
m19	local adverbs with suffix: DEM-*da-aida- (?)	2	0	no	1	μαληλο, ταληλο	0	no	0	no	0	no	0	no	0	no

Ossetic	Parachi	Pashto	Sarikoli	Shughni	Sogdian	Wakhi	Yaghnobi_W	Yaghnobi_E	Yazghulami	Yidgha	id
2 c	0 č	2 c	2 c	2 c	0 č	2 c	0 č	0 č	0 č	2 c	**p1**
0 b	0 b	1 β	1 β	1 β	1 β	1 β	1 β	1 β	1 β	1 β	**p2**
0 d	0 d	2 l ø	1 δ	1 δ	1 δ	1 δ	0 d	0 d	1 δ	2 l	**p3**
1 γ	0 g	1 γ	1 γ	1 γ	1 γ	1 γ	1 γ	1 γ	1 γ	1 γ	**p4**
2 γδ	4 t	6	3 yd	3 yd	1 γt	2 γδ	0 xt	2 γδ	2 γδ	2 γδ	**p5**
3 vd	4 ut	6 w	3 vd	3 vd	1 βt	5 b	0 ft	3 vd	3 vd	3 vd	**p6**
1 t	4 h	5 l	0 θ	0 θ	0 θ	? ?	1 t	2 s	0 θ	3 x̌	**p7**
2 ært	6 š	4 dr	10 ar	10 ar	3 tr, š	1 tr	1 tVr	8 sVr	5 c	7 x̌Vr	**p8**
2 rt	5 š	9 r	4 c	4 c	3 θr, rθ, š	1 tr	2 rVt	7 rVs	4 c	9 r	**p9**
? (DIR -tæ, OBL -t-)	7 PL -ān, -ā, -a; (GEN āna)	2 DIR -i, OBL -ān/-un/-o, etc.	10 (DIR -xεyl), OBL -ef	8 PL -ēn, -īf	1 DIR -V, GEN -ān (voc.pl. -β)	10 (DIR -išt) OBL -əv	? (DIR -t, OBL -ti)	? (DIR -t, OBL -ti)	5 (DIR -t), OBL -an	9 DIR -i, OBL -āf	**m1**
2 NOM -tæ, OBL -t-+SG	0 no	0 no	0 no	0 no	1 -t', -ty' (-t -F.SG-ending) (light stems)	0 no	2 DIR -t	2 DIR -t	3 DIR -t	0 no	**m2**
0 no	0 no	0 no	0 no	0 no	1 yes (heavy stems)	0 no	1 yes	1 yes	0 no	0 no	**m3**
0 No	0 No	0 no	2 DIR. -xεyl	0 No	1 -yšt-: DIR -yšt, -y'; OBL -yšty, -'n (anim. light stems)	1 -išt DIR.PL	0 No	0 No	0 No	0 No	**m4**
2 -uncæ / -ync	4 -an	4 -īn	4 -(y)in	4 -ēn, -an	3 -and	4 -ən	1 -or	1 -or	4 -an	5 -εt	**m5**
0 No	0 No	1 -i, -ī	0 No	0 No	0 No	0 No	0 No	0 No	0 No	0 No	**m6**
3 symax, sumax	4 wå	6 tāse	7 tamaš	7 tama	3 (ə)šmāx	9 sa(y)-iš(t)	3 šumox	3 šumox	7 təmox	2 maf/māf/ mof/mōf	**m7**
8 ie- (D)/a- u- (I)/uo- (D)	9 hē hō	4 ha-γa dā	7 yam, mi yad, di yы,yi, wi	6 yam, mi yid, di yu, yā/wi, wam	5 yw/'mw, 'mn šw/'tw (')xw/'w(w)	7 yəm yət yəw	12 iš/it ax/aw	12 iš/it ax/aw	10 du yu/way	7 mo- y- w-	**m8**
0 ʒy / ʒi	0 az	2 Poss.: SG 1 jmā, 2 stā; PL 1 zmung, 2 stāse	0 az	0 a-, as, az	1 SG 1 c’m’(k(H)), 2 c’f’k(H), (PL unattested)	0 cə	0 či	0 či	2 ACC: SG 1 ʒ-mon, 2 ʃ-tu; PL 1 ʒ-moχ, 2 ʃ-təmoχ	0 žə	**m9**
0 æd	3	0 la	3	3	1 SG 1 δ’m’(k), 2 δ’f’, (PL unattested)	0 də/tə	3	3	3	0 lo	**m10**
3	3	3	3	0 -avēn	0 β- (M), β-; 'βy- (B)	3	3	3	0 -be	0 va/vo	**m11**
2	2	2	0 par	0 par	1 SG 1 pr’m’k, 2 pr’β’k, (PL unattested)	0 pər	0 par	0 par	2	2	**m12**
0 no	0 no	0 no	0 no	1 azamand, azedand, azůdand; azam, azed, azůd	2 cyw'nt, cym'nt	0 no	0 no	0 no	0 no	0 no	**m13**
0 no	0 no	0 no	0 no	1 taramand, taredand, tarůdand	1 δyw(')nt, δm'nt; nyw'nt; prywynd, prymnd	0 no	0 no	0 no	0 no	0 no	**m14**
0 no	0 no	0 no	0 no	1 azamard, azedard, azůdard; taramard, taredard, tarůdard	1 cyw(')yδ, cytyδ, cym(')yδ; δyw(')yδ, δm'tδ; nyw'yδ, nytyδ, nymyδ; pr'yw'yδ, prytyδ, prymyδ	0 no	0 no	0 no	0 no	0 no	**m15**
1 ardæm, ūrdæm	0 no	0 no	0 no	1 yůdard; amard, adard, udard	1 (mrδ, trδ, ’wrδ); CSogd. wysprd	0 no	0 no	0 no	0 no	0 no	**m16**
0 no	0 no	0 no	0 no	0 no	1 (mrδ, trδ, ’wrδ); CSogd. kwrθ	0 no	0 no	0 no	0 no	0 no	**m17**
0 no	0 no	0 no	0 no	0 no	1 mδy, tδy, wδy	0 no	0 no	0 no	0 no	0 no	**m18**
0 no	0 no	0 no	0 no	0 no	1 mδ’yδ, tδ’yδ, wδyδ	0 no	0 no	0 no	0 no	0 no	**m19**

id	feature	0	1	2	3	4	5	6	7	8	9	10	11	12
p1	*č-	č	č/c	c										
p2	*b-	b	β											
p3	*d-	d	δ	l			Ø							
p4	*g-	g	γ											
p5	*xt	xt	γt	γδ	yd	t	k/	Ø						
p6	*ft	ft	βt	βδ	vd/ud	ut	b	w/u	Ø					
p7	*-θ-	θ	t	s	x̌	h	l	y						
p8	*θr-	θr	t(V)r	(V)rt	tr, š	dr	c	š	x̌(V)r	s(V)r	h(V)r	(V)r		
p9	*-θr-	θr	tr	r(V)t	θr, rθ, š	c	š	s(V)r	r(V)s	hr	r			
m1	nominal plural endings	NOM-ACC *-ayah, GEN *-ānām, *-īnām; DAT *-byah, INSTR *-biš, LOC *-su, *-šu, stem classes and ablaut.	NOM *-ayah vel sim., GEN *-ānām, OBL *-b-, stem classes and ablaut	DIR *-ayah; OBL *-ānām,*-īnam,*-b-; stem classes and ablaut	DIR *-ayah, OBL *-ānām/*-īnam	DIR *-ayah, OBL *-ānām	OBL *-ānām, (neo-NOM)	PL *ayah/-ānām/īnām; stem classes and ablaut	PL *-ānām	PL *-ānam, *-b-	DIR *-ayah, OBL *-b-	OBL *-b-, (neo-NOM)	PL *-b-	
m2	plurals < abstr. *-tā-	no	*-tā inflected as collective F.SG	-t- as agglutinative PL-suffix	-t as NOM.PL									
m3	PL DIR -t, OBL -ti	no	yes											
m4	Neo-Plural	no	*-išt-	-xεγl										
m5	inherited 3PL-endings	ACT *-nt(i) + MID *-r	3PL *-r (only)	3PL *-nti (only)	-nd	-n	-t							
m6	analogical 3PL-ending	inherited 3PL ending	3PL = 3SG											
m7	pronoun 2PL	Nom. *yūžam Gen. *šmāxam Dat. *yušmabya clitic *=wah	*yūžam *=wah	*yušmabya *=wah	*šmāxam *=wah	*=wah	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	2PL based on 2SG orthotone *tū/tawa	2SG orth. + 1PL	probably LW	unclear etymology			
m8	Deictic pronouns	D0 DIR *ha-, D0 OBL *ta- D1 DIR *aya-, D1 OBL *a-/ima- D2 DIR *aiša-, D2 OBL *aita- D3 DIR *haw-, D3 OBL *awa-	New stems D0 OBL *ta- D2 DIR *aiša-	3dx D0 OBL *ta- D1 DIR *aya- D1 OBL *a-/ima- D2 OBL *aita-	2dx D0 DIR *ha- D3 DIR *haw- D3 OBL *awa-	2dx D0 DIR *ha- D2 OBL *aita-	3dx D1 DIR *aya-, D1 OBL *a-/ima- D2 DIR *aiša-, D2 OBL *aita- D3 DIR *haw-, D3 OBL *awa-	3dx D1 OBL *a-/ima- D2 OBL *aita- D3 DIR *haw-, D3 OBL *awa-	3dx D1 OBL *a-/ima- D2 OBL *aita- D3 OBL *awa-	2dx D1 DIR *aya- D1 OBL *a-/ima- D3 DIR *haw- D3 OBL *awa-	2dx D1 OBL *a-/ima- D3 DIR *haw-	2dx D2 OBL *aita- D3 DIR *haw- D3 OBL *awa-	3dx D1 DIR *aya- D2 DIR *aiša- D2 OBL *aita- D3 DIR *haw-	2dx D2 DIR *aiša- D2 OBL *aita- D3 DIR *haw- D3 OBL *awa-
m9	*hača + pron.1/2 person	*hača + orth. pers. pron.	*hača + clitic pron., *hača-mā-ka-	*(ha)č(a) proclitic/prefix + *orth. pron.	*hača suffixlike postposition	Ø								
m10	*hada + pron.1/2 person	*hada + orth. pers. pron.	*hada + clitic pron., *hada-mā-ka-	*(ha)d(a) proclitic/prefix + *orth. pron.	Ø									
m11	*abi + pron.1/2 person	*abi + orth. pers. pron.	*abi + clitic pron., *abi-mā-ka-	*(a)b(i) proclitic/prefix + *orth. pron.	Ø									
m12	*upari + pron.1/2 person	*upari + orth. pers. pron.	*upari + clitic pron. *upari-mā-ka-	Ø										
m13	*hača-DEM	absent	*haca-DEM; *haca-DEM-antara	*haca-DEM-antara										
m14	productivity of the pattern adposition + DEM-*antara-	absent	*hada-, *upari-, *ana-, *tara- + *DEM-antara											
m15	productivity of the pattern adposition + DEM + suffix	absent	*hača, *hada, *upari, *ana, *tara-, + DEM + *-da-, *arda-											
m16	DEM + *arda-	absent	fused form: *ima-arda-, *ta-arda-, *awa-arda-, etc.											
m17	DEM-*θra-	absent	fused form: *ima-θra-, *ta-θra- *awa-θra-, etc.; *ku-θra-											
m18	DEM-*da-	absent	fused form: *ima-da-, *ta-da-, *awa-da-, etc.											
m19	DEM-*da-aida- (?)	absent	fused form: *ima-da-aida-, *ta-da-aida-, *awa-ta-aida-, etc.											

Appendix 4: Condensed list of allowed directionalities and costs of transitions

p1 0>1(c:1); 0>2(c:1); 1>0(c:1); 1>2(c:1); 2>0(c:2); 2>1(c:1)

p2 0>1(c:2); 1>0(c:1)

p3 0>1(c:2); 0>2(c:3); 1>0(c:1); 1>2(c:2); 2>0(c:3); 2>1(c:3)

p4 0>1(c:2); 1>0(c:1)

p5 0>1(c:2); 0>2(c:2); 0>3(c:3); 0>4(c:2); 0>5(c:4); 0>6(c:5); 1>0(c:1); 1>2(c:2); 1>3(c:2); 1>4(c:1); 1>5(c:5); 1>6(c:5); 2>0(c:3); 2>1(c:4); 2>3(c:2); 2>4(c:4); 2>5(c:5); 2>6(c:5); 3>0(c:4); 3>1(c:4); 3>2(c:4); 3>4(c:4); 3>6(c:5); 4>6(c:3); 4>3(c:5); 4>5(c:5); 5>6(c:3)

p6 0>1(c:2); 0>2(c:3); 0>3(c:2); 0>4(c:2); 0>5(c:4); 0>6(c:4); 0>7(c:5); 1>0(c:1); 1>2(c:2); 1>3(c:2); 1>4(c:1); 1>5(c:3); 1>6(c:3); 1>7(c:5); 2>0(c:4); 2>1(c:4); 2>3(c:1); 2>4(c:4); 2>5(c:4); 2>6(c:3); 2>7(c:5); 3>0(c:4); 3>1(c:4); 3>2(c:2); 3>4(c:3); 3>5(c:4); 3>6(c:2); 3>7(c:4); 4>0(c:3); 4>1(c:3); 4>2(c:3); 4>3(c:1); 4>5(c:4); 4>6(c:3); 4>7(c:5); 5>6(c:1); 5>7(c:4); 6>5(c:3); 6>7(c:1); 7>6(c:3)

p7 0>1(c:2); 0>2(c:2); 0>3(c:3); 0>4(c:2); 0>5(c:2); 0>6(c:3); 1>0(c:2); 1>2(c:4); 1>3(c:5); 1>4(c:4); 1>5(c:4); 1>6(c:4); 2>0(c:2); 2>1(c:4); 2>3(c:3); 2>4(c:1); 2>5(c:5); 2>6(c:4); 3>0(c:4); 3>1(c:5); 3>2(c:3); 3>4(c:1); 3>5(c:5); 3>6(c:3); 4>0(c:4); 4>1(c:5); 4>2(c:4); 4>3(c:2); 4>5(c:5); 4>6(c:2); 5>0(c:5); 5>1(c:5); 5>2(c:5); 5>3(c:5); 5>4(c:5); 5>6(c:3); 6>0(c:5); 6>1(c:5); 6>2(c:5); 6>3(c:4); 6>4(c:2); 6>5(c:5)

p8 0>1(c:2); 0>2(c:4); 0>3(c:3); 0>4(c:3); 0>5(c:3); 0>6(c:3); 0>7(c:3); 0>8(c:2); 0>9(c:2); 0>10(c:2); 1>0(c:2); 1>2(c:4); 1>3(c:3); 1>4(c:2); 1>5(c:3); 1>6(c:4); 1>7(c:4); 1>8(c:4); 1>9(c:4); 1>10(c:3); 2>0(c:4); 2>1(c:1); 2>3(c:4); 2>4(c:3); 2>5(c:4); 2>6(c:4); 2>7(c:5); 2>8(c:4); 2>9(c:5); 2>10(c:5); 4>0(c:3); 4>1(c:2); 4>2(c:4); 4>3(c:4); 4>5(c:4); 4>6(c:3); 4>7(c:4); 4>8(c:4); 4>9(c:4); 4>10(c:3); 5>6(c:3); 6>3(c:4); 6>5(c:4); 6>8(c:5); 6>10(c:4); 7>0(c:5); 7>1(c:5); 7>2(c:5); 7>4(c:4); 7>5(c:5); 7>6(c:4); 7>8(c:3); 7>9(c:1); 7>10(c:2); 8>0(c:3); 8>1(c:4); 8>2(c:5); 8>3(c:4); 8>4(c:5); 8>5(c:3); 8>6(c:2); 8>7(c:3); 8>9(c:1); 8>10(c:3); 9>0(c:4); 9>1(c:5); 9>2(c:5); 9>3(c:5); 9>4(c:4); 9>5(c:3); 9>6(c:3); 9>7(c:2); 9>8(c:4); 9>10(c:1); 10>4(c:5); 10>6(c:5); 10>7(c:5); 10>9(c:2)

p9 0>1(c:1); 0>2(c:2); 0>3(c:4); 0>4(c:3); 0>5(c:3); 0>6(c:2); 0>7(c:4); 0>8(c:1); 0>9(c:2); 1>0(c:2); 1>2(c:2); 1>3(c:4); 1>4(c:3); 1>5(c:4); 1>6(c:3); 1>7(c:4); 1>8(c:3); 1>9(c:2); 2>0(c:5); 2>1(c:3); 2>3(c:5); 2>4(c:5); 2>5(c:5); 2>6(c:5); 2>7(c:3); 2>8(c:5); 2>9(c:3); 4>3(c:5); 4>5(c:3); 5>1(c:5); 5>3(c:5); 5>4(c:4); 5>9(c:3); 6>0(c:3); 6>1(c:4); 6>2(c:4); 6>3(c:5); 6>4(c:4); 6>5(c:2); 6>7(c:3); 6>8(c:1); 6>9(c:1); 7>0(c:4); 7>1(c:4); 7>2(c:4); 7>3(c:5); 7>4(c:4); 7>5(c:3); 7>6(c:3); 7>8(c:4); 7>9(c:2); 8>0(c:4); 8>1(c:5); 8>2(c:5); 8>4(c:4); 8>5(c:3); 8>6(c:4); 8>7(c:4); 8>9(c:1); 9>5(c:4); 9>6(c:4); 9>7(c:4); 9>8(c:3) m1 0>1(c:1); 0>2(c:4); 0>3(c:4); 0>4(c:4); 0>5(c:5); 0>6(c:5); 0>7(c:5); 0>8(c:5); 0>9(c:5); 0>10(c:4); 0>11(c:5); 1>2(c:1); 1>3(c:3); 1>4(c:3); 1>5(c:5); 1>6(c:5); 1>7(c:5); 1>8(c:5); 1>9(c:4); 1>10(c:4); 1>11(c:5); 2>1(c:5); 2>3(c:3); 2>4(c:3); 2>5(c:4); 2>6(c:3); 2>7(c:5); 2>8(c:4); 2>9(c:3); 2>10(c:4); 2>11(c:5); 3>4(c:1); 3>5(c:3); 3>6(c:5); 3>7(c:4); 4>5(c:1); 4>6(c:5); 4>7(c:3); 5>7(c:3); 6>3(c:5); 6>4(c:5); 6>5(c:5); 6>7(c:2); 7>5(c:5); 8>5(c:5); 8>7(c:1); 8>10(c:5); 8>11(c:1); 9>10(c:0); 9>11(c:3); 10>11(c:3); 11>10(c:4)

m2 0>1(c:4); 1>0(c:1); 1>2(c:3); 1>3(c:3); 2>0(c:1); 2>3(c:3); 3>0(c:1); 3>2(c:3)

m3 0>1(c:3); 1>0(c:1)

m4 0>1(c:5); 0>2(c:5); 1>0(c:1); 1>2(c:5); 2>0(c:1); 2>1(c:5)

m5 0>1(c:3); 0>2(c:1); 0>3(c:5); 0>4(c:5); 0>5(c:5); 1>4(c:5); 1>5(c:5); 2>3(c:1); 2>4(c:3); 2>5(c:4); 3>2(c:5); 3>4(c:1); 3>5(c:3); 4>3(c:4); 4>5(c:4); 5>3(c:5); 5>4(c:4)

m6 0>1(c:3)

m7 0>1(c:3); 0>2(c:1); 0>3(c:1); 0>4(c:4); 0>5(c:5); 0>6(c:5); 0>7(c:5); 0>8(c:5); 0>9(c:5); 1>4(c:2); 1>5(c:4); 1>6(c:4); 1>7(c:4); 1>8(c:5); 1>9(c:5); 2>4(c:2); 2>6(c:4); 2>7(c:4); 2>8(c:5); 2>9(c:5); 3>4(c:2); 3>5(c:2); 3>6(c:3); 3>7(c:3); 3>8(c:5); 3>9(c:5); 4>6(c:4); 4>7(c:4); 4>8(c:5); 4>9(c:5); 5>6(c:1); 5>7(c:3); 5>8(c:5); 5>9(c:5); 6>7(c:2); 6>8(c:5); 6>9(c:5); 7>6(c:3); 7>8(c:5); 7>9(c:5); 8>6(c:3); 8>7(c:3); 8>9(c:5); 9>6(c:3); 9>7(c:3); 9>8(c:5)

m8 0>1(c:4); 0>2(c:3); 0>3(c:4); 0>4(c:4); 0>5(c:2); 0>6(c:4); 0>7(c:5); 0>8(c:4); 0>9(c:5); 0>10(c:5); 0>11(c:3); 0>12(c:4); 5>6(c:1); 5>7(c:2); 5>8(c:3); 5>9(c:4); 5>10(c:2); 5>11(c:1); 5>12(c:3); 6>7(c:1); 6>9(c:1); 6>10(c:1); 8>9(c:1); 12>10(c:1)

m9 0>1(c:3); 0>2(c:3); 0>3(c:1); 0>4(c:1); 1>0(c:2); 1>2(c:2); 1>3(c:2); 1>4(c:1); 2>0(c:2); 2>1(c:3); 2>3(c:2); 2>4(c:1); 3>0(c:4); 3>1(c:4); 3>2(c:4); 3>4(c:1)

m10 0>1(c:3); 0>2(c:3); 0>3(c:1); 1>0(c:2); 1>2(c:2); 1>3(c:1); 2>0(c:2); 2>1(c:3); 2>3(c:1)

m11 0>1(c:3); 0>2(c:3); 0>3(c:1); 1>2(c:2); 2>0(c:2); 2>1(c:3); 2>3(c:1)

m12 0>1(c:3); 0>2(c:1); 1>0(c:2); 1>2(c:1)

m13 0>1(c:3); 0>2(c:3); 1>0(c:1); 1>2(c:1); 2>0(c:1); 2>1(c:1)

m14 0>1(c:3); 1>0(c:1)

m15 0>1(c:3); 1>0(c:1)

m16 0>1(c:2); 1>0(c:1)

m17 0>1(c:2); 1>0(c:1)

m18 0>1(c:2); 1>0(c:1)

m19 0>1(c:3); 1>0(c:1)

#	id	isogloss	input state	input		output state	output	cost	
1	p1	*č-	0	č	=	0	č	=	
2	p1	*č-	0	č	>	1	č/c	1	0>1(c:1)
3	p1	*č-	0	č	>	2	c	1	0>2(c:1)
4	p1	*č-	1	č/c	>	0	č	1	1>0(c:1)
5	p1	*č-	1	č/c	=	1	č/c	=	
6	p1	*č-	1	č/c	>	2	c	1	1>2(c:1)
7	p1	*č-	2	c	>	0	č	2	2>0(c:2)
8	p1	*č-	2	c	>	1	č/c	1	2>1(c:1)
9	p1	*č-	2	c	=	2	c	=	

#	id	isogloss	input state	input		output state	output	cost	
1	p2	*b-	0	b	=	0	b	=	
2	p2	*b-	0	b	>	1	β	2	0>1(c:2)
3	p2	*b-	1	β	>	0	b	1	1>0(c:1)
4	p2	*b-	1	β	>	1	β	=	

#	id	isogloss	input state	input		output state	output	cost	
1	p3	*d-	0	d	=	0	d	=	
2	p3	*d-	0	d	>	1	δ	2	0>1(c:2)
3	p3	*d-	0	d	>	2	l	3	0>2(c:3)
4	p3	*d-	1	δ	>	0	d	1	1>0(c:1)
5	p3	*d-	1	δ	=	1	δ	=	
6	p3	*d-	1	δ	>	2	l	2	1>2(c:2)
7	p3	*d-	2	l	>	0	d	3	2>0(c:3)
8	p3	*d-	2	l	>	1	δ	3	2>1(c:3)
9	p3	*d-	2	l	=	2	l	=	

#	id	isogloss	input state	input		output state	output	cost	
1	p2	*g-	0	g	=	0	g	=	
2	p2	*g-	0	g	>	1	γ	2	0>1(c:2)
3	p2	*g-	1	γ	>	0	g	1	1>0(c:1)
4	p2	*g-	1	γ	>	1	γ	=	

#	id	isogloss	input state	input		output state	output	cost	
1	p5	*xt	0	xt	=	0	xt	=	
2	p5	*xt	0	xt	>	1	γt	2	0>1(c:2)
3	p5	*xt	0	xt	>	2	γδ	2	0>2(c:2)
4	p5	*xt	0	xt	>	3	γd	3	0>3(c:3)
5	p5	*xt	0	xt	>	4	t	2	0>4(c:2)
7	p5	*xt	0	xt	>	5	k, Ø	4	0>5(c:4)
8	p5	*xt	0	xt	>	6	Ø	5	0>6(c:5)
9	p5	*xt	1	γt	>	0	xt	1	1>0(c:1)
10	p5	*xt	1	γt	=	1	γt	=	

#	id	isogloss	input state	input		output state	output	cost	
11	p5	*xt	1	γt	>	2	γδ	2	1>2(c:2)
12	p5	*xt	1	γt	>	3	γd	2	1>3(c:2)

#	id	isogloss	input state	input		output state	output	cost	
13	p5	*xt	1	γt	>	4	t	1	1>4(c:1)
15	p5	*xt	1	γt	>	5	k, Ø	5	1>5(c:5)
16	p5	*xt	1	γt	>	6	Ø	5	1>6(c:5)
17	p5	*xt	2	γδ	>	0	xt	3	2>0(c:3)
18	p5	*xt	2	γδ	>	1	γt	4	2>1(c:4)
19	p5	*xt	2	γδ	=	2	γδ	=	
20	p5	*xt	2	γδ	>	3	γd	2	2>3(c:2)
21	p5	*xt	2	γδ	>	4	t	4	2>4(c:4)
23	p5	*xt	2	γδ	>	5	k, Ø	5	2>5(c:5)
24	p5	*xt	2	γδ	>	6	Ø	5	2>6(c:5)
25	p5	*xt	3	γd	>	0	xt	4	3>0(c:4)
26	p5	*xt	3	γd	>	1	γt	4	3>1(c:4)
27	p5	*xt	3	γd	>	2	γδ	4	3>2(c:4)
28	p5	*xt	3	γd	=	3	γd	=	
29	p5	*xt	3	γd	>	4	t	4	3>4(c:4)
31	p5	*xt	3	γd	>	5	k, Ø	n/a	
32	p5	*xt	3	γd	>	6	Ø	5	3>6(c:5)
33	p5	*xt	4	t	>	0	xt	n/a	
34	p5	*xt	4	t	>	1	γt	n/a	
35	p5	*xt	4	t	>	2	γδ	n/a	
36	p5	*xt	4	t	>	6	Ø	3	4>6(c:3)
36	p5	*xt	4	t	>	3	γd	5	4>3(c:5)
36	p5	*xt	4	t	>	5	k, Ø	5	4>5(c:5)
36	p5	*xt	4	t	=	4	t	=	
36	p5	*xt	5	k, Ø	>	6	Ø	3	5>6(c:3)
36	p5	*xt	5	k, Ø	=	5	k, Ø	=	
36	p5	*xt	5	k, Ø	>	0	xt	n/a	
36	p5	*xt	5	k, Ø	>	1	γt	n/a	
36	p5	*xt	5	k, Ø	>	2	γδ	n/a	
36	p5	*xt	5	k, Ø	>	3	γd	n/a	
36	p5	*xt	5	k, Ø	>	4	t	n/a	
36	p5	*xt	6	Ø	=	6	Ø	=	
36	p5	*xt	6	Ø	>	0	xt	n/a	
36	p5	*xt	6	Ø	>	1	γt	n/a	
36	p5	*xt	6	Ø	>	2	γδ	n/a	
36	p5	*xt	6	Ø	>	3	γd	n/a	
36	p5	*xt	6	Ø	>	4	t	n/a	
36	p5	*xt	6	Ø	>	5	k, Ø	n/a	

#	id	isogloss	input state	input		output state	output	cost	
1	p6	*-ft-	0	ft	>	1	βt	2	0>1(c:2)
2	p6	*-ft-	0	ft	>	2	βδ	3	0>2(c:3)
3	p6	*-ft-	0	ft	>	3	vd/ud	2	0>3(c:2)
4	p6	*-ft-	0	ft	>	4	ut	2	0>4(c:2)
5	p6	*-ft-	0	ft	>	5	b	4	0>5(c:4)
6	p6	*-ft-	0	ft	>	6	w/u	4	0>6(c:4)

#	id	isogloss	input state	input		output state	output	cost	
7	p6	*-ft-	0	ft	>	7		5	0>7(c:5)
8	p6	*-ft-	1	βt	>	0	ft	1	1>0(c:1)
9	p6	*-ft-	1	βt	>	2	βδ	2	1>2(c:2)
10	p6	*-ft-	1	βt	>	3	vd/ud	2	1>3(c:2)
11	p6	*-ft-	1	βt	>	4	ut	1	1>4(c:1)
12	p6	*-ft-	1	βt	>	5	b Ø	3	1>5(c:3)
13	p6	*-ft-	1	βt	>	6	w/u	3	1>6(c:3)
14	p6	*-ft-	1	βt	>	7		5	1>7(c:5)
15	p6	*-ft-	2	βδ	>	0	ft	4	2>0(c:4)
16	p6	*-ft-	2	βδ	>	1	βt	4	2>1(c:4)
17	p6	*-ft-	2	βδ	>	3	vd/ud	1	2>3(c:1)
18	p6	*-ft-	2	βδ	>	4	ut	4	2>4(c:4)
19	p6	*-ft-	2	βδ	>	5	b Ø	4	2>5(c:4)
20	p6	*-ft-	2	βδ	>	6	w/u	3	2>6(c:3)
21	p6	*-ft-	2	βδ	>	7		5	2>7(c:5)
22	p6	*-ft-	3	vd/ud	>	0	ft	4	3>0(c:4)
23	p6	*-ft-	3	vd/ud	>	1	βt	4	3>1(c:4)
24	p6	*-ft-	3	vd/ud	>	2	βδ	2	3>2(c:2)
25	p6	*-ft-	3	vd/ud	>	4	ut	3	3>4(c:3)
26	p6	*-ft-	3	vd/ud	>	5	b Ø	4	3>5(c:4)
27	p6	*-ft-	3	vd/ud	>	6	w/u	2	3>6(c:2)
28	p6	*-ft-	3	vd/ud	>	7		4	3>7(c:4)
29	p6	*-ft-	4	ut	>	0	ft	3	4>0(c:3)
30	p6	*-ft-	4	ut	>	1	βt	3	4>1(c:3)
31	p6	*-ft-	4	ut	>	2	βδ	3	4>2(c:3)
32	p6	*-ft-	4	ut	>	3	vd/ud	1	4>3(c:1)
33	p6	*-ft-	4	ut	>	5	b Ø	4	4>5(c:4)
34	p6	*-ft-	4	ut	>	6	w/u	3	4>6(c:3)
35	p6	*-ft-	4	ut	>	7		5	4>7(c:5)
36	p6	*-ft-	5	b	>	0	ft	n/a	
37	p6	*-ft-	5	b	>	1	βt	n/a	
38	p6	*-ft-	5	b	>	2	βδ	n/a	
39	p6	*-ft-	5	b	>	3	vd/ud	n/a	
40	p6	*-ft-	5	b	>	4	ut Ø	n/a	
41	p6	*-ft-	5	b	>	6	w/u	1	5>6(c:1)
42	p6	*-ft-	5	b	>	7		4	5>7(c:4)
43	p6	*-ft-	6	w/u	>	0	ft	n/a	
44	p6	*-ft-	6	w/u	>	1	βt	n/a	

#	id	isogloss	input state	input		output state	output	costs	
45	p6	*-ft-	6	w/u	>	2	βδ	n/a	
46	p6	*-ft-	6	w/u	>	3	vd/ud	n/a	
47	p6	*-ft-	6	w/u	>	4	ut Ø	n/a	
48	p6	*-ft-	6	w/u	>	5	b	3	6>5(c:3)
49	p6	*-ft-	6	w/u	>	7		1	6>7(c:1)

#	id	isogloss	input state	input		output state	output	costs	
50	p6	*-ft-	7	Ø	>	0	ft	n/a	
51	p6	*-ft-	7	Ø	>	1	βt	n/a	
52	p6	*-ft-	7	Ø	>	2	βδ	n/a	
53	p6	*-ft-	7	Ø	>	3	vd/ud	n/a	
54	p6	*-ft-	7	Ø	>	4	ut	n/a	
55	p6	*-ft-	7	Ø	>	5	b	n/a	
56	p6	*-ft-	7		>	6	w/u	3	7>6(c:3)

#	id	isogloss	input state	input		output state	output	costs	
1	p7	*-ϑ-	0	θ	>	1	t	2	0>1(c:2)
2	p7	*-ϑ-	0	θ	>	2	s	2	0>2(c:2)
3	p7	*-ϑ-	0	θ	>	3	x̌	3	0>3(c:3)
4	p7	*-ϑ-	0	θ	>	4	h	2	0>4(c:2)
5	p7	*-ϑ-	0	θ	>	5	l	2	0>5(c:2)
6	p7	*-ϑ-	0	θ	>	6	y	3	0>6(c:3)
7	p7	*-ϑ-	1	t	>	0	θ	2	1>0(c:2)
8	p7	*-ϑ-	1	t	>	2	s	4	1>2(c:4)
9	p7	*-ϑ-	1	t	>	3	x̌	5	1>3(c:5)
10	p7	*-ϑ-	1	t	>	4	h	4	1>4(c:4)
11	p7	*-ϑ-	1	t	>	5	l	4	1>5(c:4)
12	p7	*-ϑ-	1	t	>	6	y	4	1>6(c:4)
13	p7	*-ϑ-	2	s	>	0	θ	2	2>0(c:2)
14	p7	*-ϑ-	2	s	>	1	t	4	2>1(c:4)
15	p7	*-ϑ-	2	s	>	3	x̌	3	2>3(c:3)
16	p7	*-ϑ-	2	s	>	4	h	1	2>4(c:1)
17	p7	*-ϑ-	2	s	>	5	l	5	2>5(c:5)
18	p7	*-ϑ-	2	s	>	6	y	4	2>6(c:4)
19	p7	*-ϑ-	3	x̌	>	0	θ	4	3>0(c:4)
20	p7	*-ϑ-	3	x̌	>	1	t	5	3>1(c:5)
21	p7	*-ϑ-	3	x̌	>	2	s	3	3>2(c:3)
22	p7	*-ϑ-	3	x̌	>	4	h	1	3>4(c:1)
23	p7	*-ϑ-	3	x̌	>	5	l	5	3>5(c:5)
24	p7	*-ϑ-	3	x̌	>	6	y	3	3>6(c:3)
25	p7	*-ϑ-	4	h	>	0	θ	4	4>0(c:4)
26	p7	*-ϑ-	4	h	>	1	t	5	4>1(c:5)
27	p7	*-ϑ-	4	h	>	2	s	4	4>2(c:4)
28	p7	*-ϑ-	4	h	>	3	x̌	2	4>3(c:2)
29	p7	*-ϑ-	4	h	>	5	l	5	4>5(c:5)
30	p7	*-ϑ-	4	h	>	6	y	2	4>6(c:2)
31	p7	*-ϑ-	5	l	>	0	θ	5	5>0(c:5)
32	p7	*-ϑ-	5	l	>	1	t	5	5>1(c:5)
33	p7	*-ϑ-	5	l	>	2	s	5	5>2(c:5)
34	p7	*-ϑ-	5	l	>	3	x̌	5	5>3(c:5)
35	p7	*-ϑ-	5	l	>	4	h	5	5>4(c:5)
36	p7	*-ϑ-	5	l	>	6	y	3	5>6(c:3)
37	p7	*-ϑ-	6	y	>	0	θ	5	6>0(c:5)

#	id	isogloss	input state	input		output state	output	costs	
38	p7	*-ϑ-	6	y	>	1	t	5	6>1(c:5)
39	p7	*-ϑ-	6	y	>	2	s	5	6>2(c:5)
40	p7	*-ϑ-	6	y	>	3	x̌	4	6>3(c:4)
41	p7	*-ϑ-	6	y	>	4	h	2	6>4(c:2)
42	p7	*-ϑ-	6	y	>	5	l	5	6>5(c:5)

#	id	isogloss	input state	input		output state	output	costs	
1	p8	*ϑr-	0	θr	>	1	t(V)r	2	0>1(c:2)
2	p8	*ϑr-	0	θr	>	2	(V)rt	4	0>2(c:4)
3	p8	*ϑr-	0	θr	>	3	tr, š	3	0>3(c:3)
4	p8	*ϑr-	0	θr	>	4	dr	3	0>4(c:3)
5	p8	*ϑr-	0	θr	>	5	c	3	0>5(c:3)
6	p8	*ϑr-	0	θr	>	6	š	3	0>6(c:3)
7	p8	*ϑr-	0	θr	>	7	x̌(V)r	3	0>7(c:3)
8	p8	*ϑr-	0	θr	>	8	s(V)r	2	0>8(c:2)
9	p8	*ϑr-	0	θr	>	9	h(V)r	2	0>9(c:2)
10	p8	*ϑr-	0	θr	>	10	(V)r	2	0>10(c:2)
11	p8	*ϑr-	1	t(V)r	>	0	θr	2	1>0(c:2)
12	p8	*ϑr-	1	t(V)r	>	2	(V)rt	4	1>2(c:4)
13	p8	*ϑr-	1	t(V)r	>	3	tr, š	3	1>3(c:3)
14	p8	*ϑr-	1	t(V)r	>	4	dr	2	1>4(c:2)
15	p8	*ϑr-	1	t(V)r	>	5	c	3	1>5(c:3)
16	p8	*ϑr-	1	t(V)r	>	6	š	4	1>6(c:4)
17	p8	*ϑr-	1	t(V)r	>	7	x̌(V)r	4	1>7(c:4)
18	p8	*ϑr-	1	t(V)r	>	8	s(V)r	4	1>8(c:4)
19	p8	*ϑr-	1	t(V)r	>	9	h(V)r	4	1>9(c:4)
20	p8	*ϑr-	1	t(V)r	>	10	(V)r	3	1>10(c:3)
21	p8	*ϑr-	2	(V)rt	>	0	θr	4	2>0(c:4)
22	p8	*ϑr-	2	(V)rt	>	1	t(V)r	1	2>1(c:1)
23	p8	*ϑr-	2	(V)rt	>	3	tr, š	4	2>3(c:4)
24	p8	*ϑr-	2	(V)rt	>	4	dr	3	2>4(c:3)
25	p8	*ϑr-	2	(V)rt	>	5	c	4	2>5(c:4)
26	p8	*ϑr-	2	(V)rt	>	6	š	4	2>6(c:4)
27	p8	*ϑr-	2	(V)rt	>	7	x̌(V)r	5	2>7(c:5)
28	p8	*ϑr-	2	(V)rt	>	8	s(V)r	4	2>8(c:4)
29	p8	*ϑr-	2	(V)rt	>	9	h(V)r	5	2>9(c:5)
30	p8	*ϑr-	2	(V)rt	>	10	(V)r	5	2>10(c:5)
31	p8	*ϑr-	3	tr, š	>	0	θr	n/a	
32	p8	*ϑr-	3	tr, š	>	1	t(V)r	n/a	
33	p8	*ϑr-	3	tr, š	>	2	(V)rt	n/a	
34	p8	*ϑr-	3	tr, š	=	3	tr, š	=	
35	p8	*ϑr-	3	tr, š	>	4	dr	n/a	
36	p8	*ϑr-	3	tr, š	>	5	c	n/a	
37	p8	*ϑr-	3	tr, š	>	6	š	n/a	
38	p8	*ϑr-	3	tr, š	>	7	x̌(V)r	n/a	
39	p8	*ϑr-	3	tr, š	>	8	s(V)r	n/a	
40	p8	*ϑr-	3	tr, š	>	9	h(V)r	n/a	

#	id	isogloss	input state	input		output state	output	costs	
41	p8	*ϑr-	3	tr, š	>	10	(V)r	n/a	
42	p8	*ϑr-	4	dr	>	0	θr	3	4>0(c:3)
43	p8	*ϑr-	4	dr	>	1	t(V)r	2	4>1(c:2)
44	p8	*ϑr-	4	dr	>	2	(V)rt	4	4>2(c:4)
45	p8	*ϑr-	4	dr	>	3	tr, š	4	4>3(c:4)

#	id	isogloss	input state	input		output state	output	costs	
46	p8	*ϑr-	4	dr	>	5	c	4	4>5(c:4)
47	p8	*ϑr-	4	dr	>	6	š	3	4>6(c:3)
48	p8	*ϑr-	4	dr	>	7	x̌(V)r	4	4>7(c:4)
49	p8	*ϑr-	4	dr	>	8	s(V)r	4	4>8(c:4)
50	p8	*ϑr-	4	dr	>	9	h(V)r	4	4>9(c:4)
51	p8	*ϑr-	4	dr	>	10	(V)r	3	4>10(c:3)
52	p8	*ϑr-	5	c	>	0	θr	n/a	
53	p8	*ϑr-	5	c	>	1	t(V)r	n/a	
54	p8	*ϑr-	5	c	>	2	(V)rt	n/a	
55	p8	*ϑr-	5	c	>	3	tr, š	n/a	
56	p8	*ϑr-	5	c	>	4	dr	n/a	
57	p8	*ϑr-	5	c	>	6	š	3	5>6(c:3)
58	p8	*ϑr-	5	c	>	7	x̌(V)r	n/a	
59	p8	*ϑr-	5	c	>	8	s(V)r	n/a	
60	p8	*ϑr-	5	c	>	9	h(V)r	n/a	
61	p8	*ϑr-	5	c	>	10	(V)r	n/a	
62	p8	*ϑr-	6	š	>	0	θr	n/a	
63	p8	*ϑr-	6	š	>	1	t(V)r	n/a	
64	p8	*ϑr-	6	š	>	2	(V)rt	n/a	
65	p8	*ϑr-	6	š	>	3	tr, š	4	6>3(c:4)
66	p8	*ϑr-	6	š	>	4	dr	n/a	
67	p8	*ϑr-	6	š	>	5	c	4	6>5(c:4)
68	p8	*ϑr-	6	š	>	7	x̌(V)r	n/a	
69	p8	*ϑr-	6	š	>	8	s(V)r	5	6>8(c:5)
70	p8	*ϑr-	6	š	>	9	h(V)r	n/a	
71	p8	*ϑr-	6	š	>	10	(V)r	4	6>10(c:4)
72	p8	*ϑr-	7	x̌(V)r	>	0	θr	5	7>0(c:5)
73	p8	*ϑr-	7	x̌(V)r	>	1	t(V)r	5	7>1(c:5)
74	p8	*ϑr-	7	x̌(V)r	>	2	(V)rt	5	7>2(c:5)
75	p8	*ϑr-	7	x̌(V)r	>	3	tr, š	n/a	
76	p8	*ϑr-	7	x̌(V)r	>	4	dr	4	7>4(c:4)
77	p8	*ϑr-	7	x̌(V)r	>	5	c	5	7>5(c:5)
78	p8	*ϑr-	7	x̌(V)r	>	6	š	4	7>6(c:4)
79	p8	*ϑr-	7	x̌(V)r	>	8	s(V)r	3	7>8(c:3)
80	p8	*ϑr-	7	x̌(V)r	>	9	h(V)r	1	7>9(c:1)
81	p8	*ϑr-	7	x̌(V)r	>	10	(V)r	2	7>10(c:2)
82	p8	*ϑr-	8	s(V)r	>	0	θr	3	8>0(c:3)
83	p8	*ϑr-	8	s(V)r	>	1	t(V)r	4	8>1(c:4)
84	p8	*ϑr-	8	s(V)r	>	2	(V)rt	5	8>2(c:5)
85	p8	*ϑr-	8	s(V)r	>	3	tr, š	4	8>3(c:4)

#	id	isogloss	input state	input		output state	output	costs	
86	p8	*ϑr-	8	s(V)r	>	4	dr	5	8>4(c:5)
87	p8	*ϑr-	8	s(V)r	>	5	c	3	8>5(c:3)
88	p8	*ϑr-	8	s(V)r	>	6	š	2	8>6(c:2)
89	p8	*ϑr-	8	s(V)r	>	7	x̌(V)r	3	8>7(c:3)
90	p8	*ϑr-	8	s(V)r	>	9	h(V)r	1	8>9(c:1)
91	p8	*ϑr-	8	s(V)r	>	10	(V)r	3	8>10(c:3)
92	p8	*ϑr-	9	h(V)r	>	0	θr	4	9>0(c:4)
93	p8	*ϑr-	9	h(V)r	>	1	t(V)r	5	9>1(c:5)
94	p8	*ϑr-	9	h(V)r	>	2	(V)rt	5	9>2(c:5)
95	p8	*ϑr-	9	h(V)r	>	3	tr, š	5	9>3(c:5)
96	p8	*ϑr-	9	h(V)r	>	4	dr	4	9>4(c:4)
97	p8	*ϑr-	9	h(V)r	>	5	c	3	9>5(c:3)
98	p8	*ϑr-	9	h(V)r	>	6	š	3	9>6(c:3)
99	p8	*ϑr-	9	h(V)r	>	7	x̌(V)r	2	9>7(c:2)
100	p8	*ϑr-	9	h(V)r	>	8	s(V)r	4	9>8(c:4)
101	p8	*ϑr-	9	h(V)r	>	10	(V)r	1	9>10(c:1)
102	p8	*ϑr-	10	(V)r	>	0	θr	n/a	
103	p8	*ϑr-	10	(V)r	>	1	t(V)r	n/a	
104	p8	*ϑr-	10	(V)r	>	2	(V)rt	n/a	
105	p8	*ϑr-	10	(V)r	>	3	tr, š	n/a	
106	p8	*ϑr-	10	(V)r	>	4	dr	5	10>4(c:5)
107	p8	*ϑr-	10	(V)r	>	5	c	n/a	
108	p8	*ϑr-	10	(V)r	>	6	š	5	10>6(c:5)
109	p8	*ϑr-	10	(V)r	>	7	x̌(V)r	5	10>7(c:5)
110	p8	*ϑr-	10	(V)r	>	8	s(V)r	n/a	
111	p8	*ϑr-	10	(V)r	>	9	h(V)r	2	10>9(c:2)

#	id	isogloss	input state	input		output state	output	costs	
1	p9	*-ϑr-	0	θr	>	1	tr	1	0>1(c:1)
2	p9	*-ϑr-	0	θr	>	2	r(V)t	2	0>2(c:2)
3	p9	*-ϑr-	0	θr	>	3	θr, rθ, š	4	0>3(c:4)
4	p9	*-ϑr-	0	θr	>	4	c	3	0>4(c:3)
5	p9	*-ϑr-	0	θr	>	5	š	3	0>5(c:3)
6	p9	*-ϑr-	0	θr	>	6	s(V)r	2	0>6(c:2)
7	p9	*-ϑr-	0	θr	>	7	r(V)s	4	0>7(c:4)
8	p9	*-ϑr-	0	θr	>	8	hr	1	0>8(c:1)
9	p9	*-ϑr-	0	θr	>	9	r	2	0>9(c:2)
10	p9	*-ϑr-	1	tr	>	0	θr	2	1>0(c:2)
11	p9	*-ϑr-	1	tr	>	2	r(V)t	2	1>2(c:2)
12	p9	*-ϑr-	1	tr	>	3	θr, rθ, š	4	1>3(c:4)
13	p9	*-ϑr-	1	tr	>	4	c	3	1>4(c:3)
14	p9	*-ϑr-	1	tr	>	5	š	4	1>5(c:4)
15	p9	*-ϑr-	1	tr	>	6	s(V)r	3	1>6(c:3)
16	p9	*-ϑr-	1	tr	>	7	r(V)s	4	1>7(c:4)
17	p9	*-ϑr-	1	tr	>	8	hr	3	1>8(c:3)
18	p9	*-ϑr-	1	tr	>	9	r	2	1>9(c:2)
19	p9	*-ϑr-	2	r(V)t	>	0	θr	5	2>0(c:5)

#	id	isogloss	input state	input		output state	output	costs	
20	p9	*-ϑr-	2	r(V)t	>	1	tr	3	2>1(c:3)
21	p9	*-ϑr-	2	r(V)t	>	3	θr, rθ, š	5	2>3(c:5)
22	p9	*-ϑr-	2	r(V)t	>	4	c	5	2>4(c:5)
23	p9	*-ϑr-	2	r(V)t	>	5	š	5	2>5(c:5)
24	p9	*-ϑr-	2	r(V)t	>	6	s(V)r	5	2>6(c:5)
25	p9	*-ϑr-	2	r(V)t	>	7	r(V)s	3	2>7(c:3)
26	p9	*-ϑr-	2	r(V)t	>	8	hr	5	2>8(c:5)
27	p9	*-ϑr-	2	r(V)t	>	9	r	3	2>9(c:3)
28	p9	*-ϑr-	3	θr, rθ, š	>	0	θr	n/a	
29	p9	*-ϑr-	3	θr, rθ, š	>	1	tr	n/a	
30	p9	*-ϑr-	3	θr, rθ, š	>	2	r(V)t	n/a	
31	p9	*-ϑr-	3	θr, rθ, š	=	3	θr, rθ, š	=	
32	p9	*-ϑr-	3	θr, rθ, š	>	4	c	n/a	
33	p9	*-ϑr-	3	θr, rθ, š	>	5	š	n/a	
34	p9	*-ϑr-	3	θr, rθ, š	>	6	s(V)r	n/a	
35	p9	*-ϑr-	3	θr, rθ, š	>	7	r(V)s	n/a	
36	p9	*-ϑr-	3	θr, rθ, š	>	8	hr	n/a	
37	p9	*-ϑr-	3	θr, rθ, š	>	9	r	n/a	
38	p9	*-ϑr-	4	c	>	0	θr	n/a	
39	p9	*-ϑr-	4	c	>	1	tr	n/a	
40	p9	*-ϑr-	4	c	>	2	r(V)t	n/a	
41	p9	*-ϑr-	4	c	>	3	θr, rθ, š	5	4>3(c:5)
42	p9	*-ϑr-	4	c	>	5	š	3	4>5(c:3)
43	p9	*-ϑr-	4	c	>	6	s(V)r	n/a	
44	p9	*-ϑr-	4	c	>	7	r(V)s	n/a	
45	p9	*-ϑr-	4	c	>	8	hr	n/a	

#	id	isogloss	input state	input		output state	output	costs	
46	p9	*-ϑr-	4	c	>	9	r	n/a	
47	p9	*-ϑr-	5	š	>	0	θr	n/a	
48	p9	*-ϑr-	5	š	>	1	tr	5	5>1(c:5)
49	p9	*-ϑr-	5	š	>	2	r(V)t	n/a	
50	p9	*-ϑr-	5	š	>	3	θr, rθ, š	5	5>3(c:5)
51	p9	*-ϑr-	5	š	>	4	c	4	5>4(c:4)
52	p9	*-ϑr-	5	š	>	6	s(V)r	n/a	
53	p9	*-ϑr-	5	š	>	7	r(V)s	n/a	
54	p9	*-ϑr-	5	š	>	8	hr	n/a	
55	p9	*-ϑr-	5	š	>	9	r	3	5>9(c:3)
56	p9	*-ϑr-	6	s(V)r	>	0	θr	3	6>0(c:3)
57	p9	*-ϑr-	6	s(V)r	>	1	tr	4	6>1(c:4)
58	p9	*-ϑr-	6	s(V)r	>	2	r(V)t	4	6>2(c:4)
59	p9	*-ϑr-	6	s(V)r	>	3	θr, rθ, š	5	6>3(c:5)
60	p9	*-ϑr-	6	s(V)r	>	4	c	4	6>4(c:4)
61	p9	*-ϑr-	6	s(V)r	>	5	š	2	6>5(c:2)
62	p9	*-ϑr-	6	s(V)r	>	7	r(V)s	3	6>7(c:3)
63	p9	*-ϑr-	6	s(V)r	>	8	hr	1	6>8(c:1)
64	p9	*-ϑr-	6	s(V)r	>	9	r	1	6>9(c:1)

#	id	isogloss	input state	input		output state	output	costs	
65	p9	*-ϑr-	7	r(V)s	>	0	θr	4	7>0(c:4)
66	p9	*-ϑr-	7	r(V)s	>	1	tr	4	7>1(c:4)
67	p9	*-ϑr-	7	r(V)s	>	2	r(V)t	4	7>2(c:4)
68	p9	*-ϑr-	7	r(V)s	>	3	θr, rθ, š	5	7>3(c:5)
69	p9	*-ϑr-	7	r(V)s	>	4	c	4	7>4(c:4)
70	p9	*-ϑr-	7	r(V)s	>	5	š	3	7>5(c:3)
71	p9	*-ϑr-	7	r(V)s	>	6	s(V)r	3	7>6(c:3)
72	p9	*-ϑr-	7	r(V)s	>	8	hr	4	7>8(c:4)
73	p9	*-ϑr-	7	r(V)s	>	9	r	2	7>9(c:2)
74	p9	*-ϑr-	8	hr	>	0	θr	4	8>0(c:4)
75	p9	*-ϑr-	8	hr	>	1	tr	5	8>1(c:5)
76	p9	*-ϑr-	8	hr	>	2	r(V)t	5	8>2(c:5)
77	p9	*-ϑr-	8	hr	>	3	θr, rθ, š	n/a	
78	p9	*-ϑr-	8	hr	>	4	c	4	8>4(c:4)
79	p9	*-ϑr-	8	hr	>	5	š	3	8>5(c:3)
80	p9	*-ϑr-	8	hr	>	6	s(V)r	4	8>6(c:4)
81	p9	*-ϑr-	8	hr	>	7	r(V)s	4	8>7(c:4)
82	p9	*-ϑr-	8	hr	>	9	r	1	8>9(c:1)
83	p9	*-ϑr-	9	r	>	0	θr	n/a	
84	p9	*-ϑr-	9	r	>	1	tr	n/a	
85	p9	*-ϑr-	9	r	>	2	r(V)t	n/a	
86	p9	*-ϑr-	9	r	>	3	θr, rθ, š	n/a	
87	p9	*-ϑr-	9	r	>	4	c	n/a	
88	p9	*-ϑr-	9	r	>	5	š	4	9>5(c:4)
89	p9	*-ϑr-	9	r	>	6	s(V)r	4	9>6(c:4)
90	p9	*-ϑr-	9	r	>	7	r(V)s	4	9>7(c:4)
91	p9	*-ϑr-	9	r	>	8	hr	3	9>8(c:3)

#	id	isogloss	input state	input		output state	output	cost	
1	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	=	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	=	
2	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	1	0>1(c:1)
3	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	4	0>2(c:4)
4	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	3	DIR *-ayah, OBL *-ānām/*-īnam	4	0>3(c:4)
5	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	4	DIR *-ayah, OBL *-ānām	4	0>4(c:4)
6	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	5	OBL *-ānām (neo-NOM)	5	0>5(c:5)
7	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	5	0>6(c:5)
8	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	7	PL *-ānām	5	0>7(c:5)
9	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	8	PL *-ānam, *-b-	5	0>8(c:5)
10	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	9	DIR *-ayah, OBL *-b-	5	0>9(c:5)
11	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	10	OBL *-b- (neo-NOM)	4	0>10(c:4)
12	m1	nominal plural endings	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	>	11	PL *-b-	5	0>11(c:5)

13	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
14	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	=	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	=	
15	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	1	1>2(c:1)
16	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	3	DIR *-ayah, OBL *-ānām/*-īnam	3	1>3(c:3)
17	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	4	DIR *-ayah, OBL *-ānām	3	1>4(c:3)
18	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	5	OBL *-ānām (neo-NOM)	5	1>5(c:5)
19	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	5	1>6(c:5)
20	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	7	PL *-ānām	5	1>7(c:5)
21	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	8	PL *-ānam, *-b-	5	1>8(c:5)
22	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	9	DIR *-ayah, OBL *-b-	4	1>9(c:4)

23	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	10	OBL *-b- (neo-NOM)	4	1>10(c:4)
24	m1	nominal plural endings	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	>	11	PL *-b-	5	1>11(c:5)
25	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
26	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	5	2>1(c:5)
27	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	=	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	=	
28	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	3	DIR *-ayah, OBL *-ānām/*-īnam	3	2>3(c:3)
29	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	4	DIR *-ayah, OBL *-ānām	3	2>4(c:3)
30	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	5	OBL *-ānām (neo-NOM)	4	2>5(c:4)
31	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	3	2>6(c:3)
32	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	7	PL *-ānām	5	2>7(c:5)
33	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	8	PL *-ānam, *-b-	4	2>8(c:4)
34	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	9	DIR *-ayah, OBL *-b-	3	2>9(c:3)
35	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	10	OBL *-b- (neo-NOM)	4	2>10(c:4)
36	m1	nominal plural endings	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	>	11	PL *-b-	5	2>11(c:5)

37	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
38	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
39	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
40	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	=	3	DIR *-ayah, OBL *-ānām/*-īnam	=	
41	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	4	DIR *-ayah, OBL *-ānām	1	3>4(c:1)
42	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	5	OBL *-ānām (neo-NOM)	3	3>5(c:3)
43	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	5	3>6(c:5)
44	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	7	PL *-ānām	4	3>7(c:4)
45	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	8	PL *-ānam, *-b-	n/a	
46	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	9	DIR *-ayah, OBL *-b-	n/a	
47	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	10	OBL *-b- (neo-NOM)	n/a	
48	m1	nominal plural endings	3	DIR *-ayah, OBL *-ānām/*-īnam	>	11	PL *-b-	n/a	
49	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
50	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
51	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
52	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a	
53	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	=	4	DIR *-ayah, OBL *-ānām	=	
54	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	5	OBL *-ānām (neo-NOM)	1	4>5(c:1)
55	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	5	4>6(c:5)
56	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	7	PL *-ānām	3	4>7(c:3)
57	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	8	PL *-ānam, *-b-	n/a	

58	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	9	DIR *-ayah, OBL *-b-	n/a	
59	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	10	OBL *-b- (neo-NOM)	n/a	
60	m1	nominal plural endings	4	DIR *-ayah, OBL *-ānām	>	11	PL *-b-	n/a	
61	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
62	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
63	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
64	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a	
65	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	4	DIR *-ayah, OBL *-ānām	n/a	
66	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	=	5	OBL *-ānām (neo-NOM)	=	
67	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	n/a	
68	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	7	PL *-ānām	3	5>7(c:3)
69	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	8	PL *-ānam, *-b-	n/a	
70	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	9	DIR *-ayah, OBL *-b-	n/a	
71	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	10	OBL *-b- (neo-NOM)	n/a	
72	m1	nominal plural endings	5	OBL *-ānām (neo-NOM)	>	11	PL *-b-	n/a	
73	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
74	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
75	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
76	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	3	DIR *-ayah, OBL *-ānām/*-īnam	5	6>3(c:5)
77	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	4	DIR *-ayah, OBL *-ānām	5	6>4(c:5)

78	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	5	OBL *-ānām (neo-NOM)	5	6>5(c:5)
79	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	=	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	=	
80	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	7	PL *-ānām	2	6>7(c:2)
81	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	8	PL *-ānam, *-b-	n/a	
82	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	9	DIR *-ayah, OBL *-b-	n/a	
83	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	10	OBL *-b- (neo-NOM)	n/a	
84	m1	nominal plural endings	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	>	11	PL *-b-	n/a	
85	m1	nominal plural endings	7	PL *-ānām	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
86	m1	nominal plural endings	7	PL *-ānām	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
87	m1	nominal plural endings	7	PL *-ānām	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
88	m1	nominal plural endings	7	PL *-ānām	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a	
89	m1	nominal plural endings	7	PL *-ānām	>	4	DIR *-ayah, OBL *-ānām	n/a	
90	m1	nominal plural endings	7	PL *-ānām	>	5	OBL *-ānām (neo-NOM)	5	7>5(c:5)
91	m1	nominal plural endings	7	PL *-ānām	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	n/a	
92	m1	nominal plural endings	7	PL *-ānām	=	7	PL *-ānām	=	
93	m1	nominal plural endings	7	PL *-ānām	>	8	PL *-ānam, *-b-	n/a	
94	m1	nominal plural endings	7	PL *-ānām	>	9	DIR *-ayah, OBL *-b-	n/a	
95	m1	nominal plural endings	7	PL *-ānām	>	10	OBL *-b- (neo-NOM)	n/a	
96	m1	nominal plural endings	7	PL *-ānām	>	11	PL *-b-	n/a	

97	m1	nominal plural endings	8	PL *-ānam, *-b-	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a		
98	m1	nominal plural endings	8	PL *-ānam, *-b-	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a		
99	m1	nominal plural endings	8	PL *-ānam, *-b-	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a		
100	m1	nominal plural endings	8	PL *-ānam, *-b-	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a		
101	m1	nominal plural endings	8	PL *-ānam, *-b-	>	4	DIR *-ayah, OBL *-ānām	n/a		
102	m1	nominal plural endings	8	PL *-ānam, *-b-	>	5	OBL *-ānām (neo-NOM)		5	8>5(c:5)
103	m1	nominal plural endings	8	PL *-ānam, *-b-	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	n/a		
104	m1	nominal plural endings	8	PL *-ānam, *-b-	>	7	PL *-ānām		1	8>7(c:1)
105	m1	nominal plural endings	8	PL *-ānam, *-b-	=	8	PL *-ānam, *-b-	=		
106	m1	nominal plural endings	8	PL *-ānam, *-b-	>	9	DIR *-ayah, OBL *-b-	n/a		
107	m1	nominal plural endings	8	PL *-ānam, *-b-	>	10	OBL *-b- (neo-NOM)		5	8>10(c:5)
108	m1	nominal plural endings	8	PL *-ānam, *-b-	>	11	PL *-b-		1	8>11(c:1)
109	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a		
110	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a		
111	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a		
112	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a		
113	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	4	DIR *-ayah, OBL *-ānām	n/a		
114	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	5	OBL *-ānām (neo-NOM)	n/a		
115	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	n/a		
116	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	7	PL *-ānām	n/a		
117	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	8	PL *-ānam, *-b-	n/a		

118	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	=	9	DIR *-ayah, OBL *-b-	=	
119	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	10	OBL *-b- (neo-NOM)	0	9>10(c:0)
120	m1	nominal plural endings	9	DIR *-ayah, OBL *-b-	>	11	PL *-b-	3	9>11(c:3)
121	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
122	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
123	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
124	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a	
125	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	4	DIR *-ayah, OBL *-ānām	n/a	
126	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	5	OBL *-ānām (neo-NOM)	n/a	
127	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	n/a	
128	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	7	PL *-ānām	n/a	
129	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	8	PL *-ānam, *-b-	n/a	
130	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	9	DIR *-ayah, OBL *-b-	n/a	
131	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	=	10	OBL *-b- (neo-NOM)	=	
132	m1	nominal plural endings	10	OBL *-b- (neo-NOM)	>	11	PL *-b-	3	10>11(c:3)
133	m1	nominal plural endings	11	PL *-b-	>	0	nom *-ayah; gen *-ānām, *-īnām; dat *-byah instr *-biš. Stem classes and ablaut	n/a	
134	m1	nominal plural endings	11	PL *-b-	>	1	NOM *-ayah vel sim., GEN *-ānām/*-īnām, "DAT" *-b- stem classes and ablaut	n/a	
135	m1	nominal plural endings	11	PL *-b-	>	2	DIR *-ayah, OBL *-ānām/*-īnam/*-b- stem classes and ablaut	n/a	
136	m1	nominal plural endings	11	PL *-b-	>	3	DIR *-ayah, OBL *-ānām/*-īnam	n/a	
137	m1	nominal plural endings	11	PL *-b-	>	4	DIR *-ayah, OBL *-ānām	n/a	
138	m1	nominal plural endings	11	PL *-b-	>	5	OBL *-ānām (neo-NOM)	n/a	

139	m1	nominal plural endings	11	PL *-b-	>	6	PL *-ayah, *-ānām/īnām; stem classes and ablaut	n/a	
140	m1	nominal plural endings	11	PL *-b-	>	7	PL *-ānām	n/a	
141	m1	nominal plural endings	11	PL *-b-	>	8	PL *-ānam, *-b-	n/a	
142	m1	nominal plural endings	11	PL *-b-	>	9	DIR *-ayah, OBL *-b-	n/a	
143	m1	nominal plural endings	11	PL *-b-	>	10	OBL *-b- (neo-NOM)	4	11>10(c:4)
144	m1	nominal plural endings	11	PL *-b-	=	11	PL *-b-	=	

#	**id**	**isogloss**	**input state**	**input**		**output state**	**output**	**cost**	
1	m2	plurals < abstr. *-tā-	0	no	=	0	no	=	
2	m2	plurals < abstr. *-tā-	0	no	>	1	*-tā as coll. in pl. meaning	4	0>1(c:4)
3	m2	plurals < abstr. *-tā-	0	no	>	2	-t- as agg. pl.-suffix	n/a	
4	m2	plurals < abstr. *-tā-	0	no	>	3	-t as nom.pl.	n/a	
5	m2	plurals < abstr. *-tā-	1	*-tā as coll. in pl. meaning	>	0	no	1	1>0(c:1)
6	m2	plurals < abstr. *-tā-	1	*-tā as coll. in pl. meaning	=	1	*-tā as coll. in pl. meaning	=	
7	m2	plurals < abstr. *-tā-	1	*-tā as coll. in pl. meaning	>	2	-t- as agg. pl.-suffix	3	1>2(c:3)
8	m2	plurals < abstr. *-tā-	1	*-tā as coll. in pl. meaning	>	3	-t as nom.pl.	3	1>3(c:3)
9	m2	plurals < abstr. *-tā-	2	-t- as agg. pl.-suffix	>	0	no	1	2>0(c:1)
10	m2	plurals < abstr. *-tā-	2	-t- as agg. pl.-suffix	>	1	*-tā as coll. in pl. meaning	n/a	
11	m2	plurals < abstr. *-tā-	2	-t- as agg. pl.-suffix	=	2	-t- as agg. pl.-suffix	=	
12	m2	plurals < abstr. *-tā-	2	-t- as agg. pl.-suffix	>	3	-t as nom.pl.	3	2>3(c:3)
13	m2	plurals < abstr. *-tā-	3	-t as nom.pl.	>	0	no	1	3>0(c:1)
14	m2	plurals < abstr. *-tā-	3	-t as nom.pl.	>	1	*-tā as coll. in pl. meaning	n/a	
15	m2	plurals < abstr. *-tā-	3	-t as nom.pl.	>	2	-t- as agg. pl.-suffix	3	3>2(c:3)
16	m2	plurals < abstr. *-tā-	3	-t as nom.pl.	=	3	-t as nom.pl.	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m3	plurals in DIR -t, OBL -ti	0	no	=	0	no	=	
2	m3	plurals in DIR -t, OBL -ti	0	no	>	1	DIR -t, OBL -ti	3	0>1(c:3)
3	m3	plurals in DIR -t, OBL -ti	1	DIR -t, OBL -ti	>	0	no	1	1>0(c:1)
4	m3	plurals in DIR -t, OBL -ti	1	DIR -t, OBL -ti	=	1	DIR -t, OBL -ti	n/a	

#	id	isogloss	input state	input		output state	output	cost	
1	m4	Neo-Plural	0	no	=	0	no	=	
2	m4	Neo-Plural	0	no	>	1	*-išt-	5	0>1(c:5)
3	m4	Neo-Plural	0	no	>	2	-xεyl	5	0>2(c:5)
4	m4	Neo-Plural	1	*-išt-	>	0	no	1	1>0(c:1)
5	m4	Neo-Plural	1	*-išt-	=	1	*-išt-	=	
6	m4	Neo-Plural	1	*-išt-	>	2	-xεyl	5	1>2(c:5)
7	m4	Neo-Plural	2	-xεyl	>	0	no	1	2>0(c:1)
8	m4	Neo-Plural	2	-xεyl	>	1	*-išt-	5	2>1(c:5)
9	m4	Neo-Plural	2	-xεyl	=	2	-xεyl	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m5	inherited 3PL-endings	0	ACT *-nt(i) + MID *-r	=	0	ACT *-nt(i) + MID *-r	=	
2	m5	inherited 3PL-endings	0	ACT *-nt(i) + MID *-r	>	1	3PL *-r (only)	3	0>1(c:3)
3	m5	inherited 3PL-endings	0	ACT *-nt(i) + MID *-r	>	2	3PL *-nti (only)	1	0>2(c:1)
4	m5	inherited 3PL-endings	0	ACT *-nt(i) + MID *-r	>	3	-nd	5	0>3(c:5)

5	m5	inherited 3PL-endings	0 ACT *-nt(i) + MID *-r	>	4 -n	5	0>4(c:5)
6	m5	inherited 3PL-endings	0 ACT *-nt(i) + MID *-r	>	5 -t	5	0>5(c:5)
7	m5	inherited 3PL-endings	1 3PL *-r (only)	>	0 ACT *-nt(i) + MID *-r	n/a	
8	m5	inherited 3PL-endings	1 3PL *-r (only)	=	1 3PL *-r (only)	=	
9	m5	inherited 3PL-endings	1 3PL *-r (only)	>	2 3PL *-nti (only)	n/a	
10	m5	inherited 3PL-endings	1 3PL *-r (only)	>	3 -nd	n/a	
11	m5	inherited 3PL-endings	1 3PL *-r (only)	>	4 -n	5	1>4(c:5)
12	m5	inherited 3PL-endings	1 3PL *-r (only)	>	5 -t	5	1>5(c:5)
13	m5	inherited 3PL-endings	2 3PL *-nti (only)	>	0 ACT *-nt(i) + MID *-r	n/a	
14	m5	inherited 3PL-endings	2 3PL *-nti (only)	>	1 3PL *-r (only)	n/a	
15	m5	inherited 3PL-endings	2 3PL *-nti (only)	=	2 3PL *-nti (only)	=	
16	m5	inherited 3PL-endings	2 3PL *-nti (only)	>	3 -nd	1	2>3(c:1)
17	m5	inherited 3PL-endings	2 3PL *-nti (only)	>	4 -n	3	2>4(c:3)
18	m5	inherited 3PL-endings	2 3PL *-nti (only)	>	5 -t	4	2>5(c:4)
19	m5	inherited 3PL-endings	3 -nd	>	0 ACT *-nt(i) + MID *-r	n/a	
20	m5	inherited 3PL-endings	3 -nd	>	1 3PL *-r (only)	n/a	
21	m5	inherited 3PL-endings	3 -nd	>	2 3PL *-nti (only)	5	3>2(c:5)
22	m5	inherited 3PL-endings	3 -nd	=	3 -nd	=	
23	m5	inherited 3PL-endings	3 -nd	>	4 -n	1	3>4(c:1)
24	m5	inherited 3PL-endings	3 -nd	>	5 -t	3	3>5(c:3)
25	m5	inherited 3PL-endings	4 -n	>	0 ACT *-nt(i) + MID *-r	n/a	
26	m5	inherited 3PL-endings	4 -n	>	1 3PL *-r (only)	n/a	
27	m5	inherited 3PL-endings	4 -n	>	2 3PL *-nti (only)	n/a	
28	m5	inherited 3PL-endings	4 -n	>	3 -nd	4	4>3(c:4)
29	m5	inherited 3PL-endings	4 -n	=	4 -n	=	
30	m5	inherited 3PL-endings	4 -n	>	5 -t	4	4>5(c:4)
31	m5	inherited 3PL-endings	5 -t	>	0 ACT *-nt(i) + MID *-r	n/a	
32	m5	inherited 3PL-endings	5 -t	>	1 3PL *-r (only)	n/a	
33	m5	inherited 3PL-endings	5 -t	>	2 3PL *-nti (only)	n/a	
34	m5	inherited 3PL-endings	5 -t	>	3 -nd	5	5>3(c:5)
35	m5	inherited 3PL-endings	5 -t	>	4 -n	4	5>4(c:4)

36	m5	inherited 3PL-endings	5	-t	=	5	-t	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m6	analogical 3PL-ending	0	inherited 3PL ending	=	0	inherited 3PL ending	=	
2	m6	analogical 3PL-ending	0	inherited 3PL ending	>	1	3PL = 3SG	3	0>1(c:3)
3	m6	analogical 3PL-ending	1	3PL = 3SG	>	0	inherited 3PL ending	n/a	
4	m6	analogical 3PL-ending	1	3PL = 3SG	=	1	3PL = 3SG	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m8	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	=	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	=	
2	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	1	*yūžam, *=wah	3	0>1(c:3)
3	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	2	*yušmabya, *=wah	1	0>2(c:1)
4	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	3	*šmāxam, *=wah	1	0>3(c:1)
5	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	4	*=wah	4	0>4(c:4)
6	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	5	0>5(c:5)
7	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	6	2PL based on 2SG *tū/tawa	5	0>6(c:5)
8	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	7	2SG orthotone + 1PL	5	0>7(c:5)

9	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	8	probably LW	5	0>8(c:5)
10	m7	pronoun 2PL	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	>	9	unclear etymology	5	0>9(c:5)
11	m7	pronoun 2PL	1	*yūžam, *=wah	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
12	m7	pronoun 2PL	1	*yūžam, *=wah	=	1	*yūžam, *=wah	=	
13	m7	pronoun 2PL	1	*yūžam, *=wah	>	2	*yušmabya, *=wah	n/a	
14	m7	pronoun 2PL	1	*yūžam, *=wah	>	3	*šmāxam, *=wah	n/a	
15	m7	pronoun 2PL	1	*yūžam, *=wah	>	4	*=wah	2	1>4(c:2)
16	m7	pronoun 2PL	1	*yūžam, *=wah	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	4	1>5(c:4)
17	m7	pronoun 2PL	1	*yūžam, *=wah	>	6	2PL based on 2SG *tū/tawa	4	1>6(c:4)
18	m7	pronoun 2PL	1	*yūžam, *=wah	>	7	2SG orthotone + 1PL	4	1>7(c:4)
19	m7	pronoun 2PL	1	*yūžam, *=wah	>	8	probably LW	5	1>8(c:5)
20	m7	pronoun 2PL	1	*yūžam, *=wah	>	9	unclear etymology	5	1>9(c:5)
21	m7	pronoun 2PL	2	*yušmabya, *=wah	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
22	m7	pronoun 2PL	2	*yušmabya, *=wah	>	1	*yūžam, *=wah	n/a	
23	m7	pronoun 2PL	2	*yušmabya, *=wah	=	2	*yušmabya, *=wah	=	
24	m7	pronoun 2PL	2	*yušmabya, *=wah	>	3	*šmāxam, *=wah	n/a	
25	m7	pronoun 2PL	2	*yušmabya, *=wah	>	4	*=wah	2	2>4(c:2)
26	m7	pronoun 2PL	2	*yušmabya, *=wah	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	n/a	
27	m7	pronoun 2PL	2	*yušmabya, *=wah	>	6	2PL based on 2SG *tū/tawa	4	2>6(c:4)
28	m7	pronoun 2PL	2	*yušmabya, *=wah	>	7	2SG orthotone + 1PL	4	2>7(c:4)
29	m7	pronoun 2PL	2	*yušmabya, *=wah	>	8	probably LW	5	2>8(c:5)
30	m7	pronoun 2PL	2	*yušmabya, *=wah	>	9	unclear etymology	5	2>9(c:5)
31	m7	pronoun 2PL	3	*šmāxam, *=wah	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
32	m7	pronoun 2PL	3	*šmāxam, *=wah	>	1	*yūžam, *=wah	n/a	

33	m7	pronoun 2PL	3	*šmāxam, *=wah	>	2	*yušmabya, *=wah	n/a	
34	m7	pronoun 2PL	3	*šmāxam, *=wah	=	3	*šmāxam, *=wah	=	
35	m7	pronoun 2PL	3	*šmāxam, *=wah	>	4	*=wah	2	3>4(c:2)
36	m7	pronoun 2PL	3	*šmāxam, *=wah	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	2	3>5(c:2)
37	m7	pronoun 2PL	3	*šmāxam, *=wah	>	6	2PL based on 2SG *tū/tawa	3	3>6(c:3)
38	m7	pronoun 2PL	3	*šmāxam, *=wah	>	7	2SG orthotone + 1PL	3	3>7(c:3)
39	m7	pronoun 2PL	3	*šmāxam, *=wah	>	8	probably LW	5	3>8(c:5)
40	m7	pronoun 2PL	3	*šmāxam, *=wah	>	9	unclear etymology	5	3>9(c:5)
41	m7	pronoun 2PL	4	*=wah	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
42	m7	pronoun 2PL	4	*=wah	>	1	*yūžam, *=wah	n/a	
43	m7	pronoun 2PL	4	*=wah	>	2	*yušmabya, *=wah	n/a	
44	m7	pronoun 2PL	4	*=wah	>	3	*šmāxam, *=wah	n/a	
45	m7	pronoun 2PL	4	*=wah	=	4	*=wah	=	
46	m7	pronoun 2PL	4	*=wah	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	n/a	
47	m7	pronoun 2PL	4	*=wah	>	6	2PL based on 2SG *tū/tawa	4	4>6(c:4)
48	m7	pronoun 2PL	4	*=wah	>	7	2SG orthotone + 1PL	4	4>7(c:4)
49	m7	pronoun 2PL	4	*=wah	>	8	probably LW	5	4>8(c:5)
50	m7	pronoun 2PL	4	*=wah	>	9	unclear etymology	5	4>9(c:5)
51	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
52	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	1	*yūžam, *=wah	n/a	
53	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	2	*yušmabya, *=wah	n/a	
54	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	3	*šmāxam, *=wah	n/a	
55	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	4	*=wah	n/a	

56	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	=	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	=	
57	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	6	2PL based on 2SG *tū/tawa	1	5>6(c:1)
58	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	7	2SG orthotone + 1PL	3	5>7(c:3)
59	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	8	probably LW	5	5>8(c:5)
60	m7	pronoun 2PL	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	>	9	unclear etymology	5	5>9(c:5)
61	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
62	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	1	*yūžam, *=wah	n/a	
63	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	2	*yušmabya, *=wah	n/a	
64	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	3	*šmāxam, *=wah	n/a	
65	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	4	*=wah	n/a	
66	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	n/a	
67	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	=	6	2PL based on 2SG *tū/tawa	=	
68	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	7	2SG orthotone + 1PL	2	6>7(c:2)
69	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	8	probably LW	5	6>8(c:5)
70	m7	pronoun 2PL	6	2PL based on 2SG *tû/tawa	>	9	unclear etymology	5	6>9(c:5)
71	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a	
72	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	1	*yūžam, *=wah	n/a	
73	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	2	*yušmabya, *=wah	n/a	
74	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	3	*šmāxam, *=wah	n/a	
75	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	4	*=wah	n/a	
76	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	n/a	
77	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	6	2PL based on 2SG *tū/tawa	3	7>6(c:3)

78	m7	pronoun 2PL	7	2SG orthotone + 1PL	=	7	2SG orthotone + 1PL	=		
79	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	8	probably LW		5	7>8(c:5)
80	m7	pronoun 2PL	7	2SG orthotone + 1PL	>	9	unclear etymology		5	7>9(c:5)
81	m7	pronoun 2PL	8	probably LW	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a		
82	m7	pronoun 2PL	8	probably LW	>	1	*yūžam, *=wah	n/a		
83	m7	pronoun 2PL	8	probably LW	>	2	*yušmabya, *=wah	n/a		
84	m7	pronoun 2PL	8	probably LW	>	3	*šmāxam, *=wah	n/a		
85	m7	pronoun 2PL	8	probably LW	>	4	*=wah	n/a		
86	m7	pronoun 2PL	8	probably LW	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	n/a		
87	m7	pronoun 2PL	8	probably LW	>	6	2PL based on 2SG *tū/tawa		3	8>6(c:3)
88	m7	pronoun 2PL	8	probably LW	>	7	2SG orthotone + 1PL		3	8>7(c:3)
89	m7	pronoun 2PL	8	probably LW	=	8	probably LW	=		
90	m7	pronoun 2PL	8	probably LW	>	9	unclear etymology		5	8>9(c:5)
91	m7	pronoun 2PL	9	unclear etymology	>	0	Nom. *yūžam, Gen. *šmāxam Dat. *yušmabya, clitic *=wah	n/a		
92	m7	pronoun 2PL	9	unclear etymology	>	1	*yūžam, *=wah	n/a		
93	m7	pronoun 2PL	9	unclear etymology	>	2	*yušmabya, *=wah	n/a		
94	m7	pronoun 2PL	9	unclear etymology	>	3	*šmāxam, *=wah	n/a		
95	m7	pronoun 2PL	9	unclear etymology	>	4	*=wah	n/a		
96	m7	pronoun 2PL	9	unclear etymology	>	5	2PL *šmāxam > = 1PL *māxam replaced by form based on 2SG	n/a		
97	m7	pronoun 2PL	9	unclear etymology	>	6	2PL based on 2SG *tū/tawa		3	9>6(c:3)
98	m7	pronoun 2PL	9	unclear etymology	>	7	2SG orthotone + 1PL		3	9>7(c:3)
99	m7	pronoun 2PL	9	unclear etymology	>	8	probably LW		5	9>8(c:5)
100	m7	pronoun 2PL	9	unclear etymology	=	9	unclear etymology	=		

#	id	isogloss	input state	input		output state	output	cost
1	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	=	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	=
2	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	0>1(c:4) 4
3	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	0>2(c:3) 3
4	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	0>3(c:4) 4
5	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita-	0>4(c:4) 4
6	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	0>5(c:2) 2
7	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	0>6(c:4) 4

8	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	5	0>7(c:5)
9	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	4	0>8(c:4)
10	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	5	0>9(c:5)
11	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	5	0>10(c:5)
12	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	3	0>11(c:3)
13	m8	Deictic pronouns	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	4	0>12(c:4)
14	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
15	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	=	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	=	
16	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a	
17	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a	
18	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a	

19	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
20	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
21	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	n/a
22	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	n/a
23	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	n/a
24	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
25	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	n/a
26	m8	Deictic pronouns	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
27	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
28	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a
29	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	=	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	=
30	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a
31	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a

32	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
33	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
34	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	n/a
35	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	n/a
36	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	n/a
37	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
38	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	n/a
39	m8	Deictic pronouns	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
40	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
41	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a
42	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a
43	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	=	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	=

44	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	n/a
45	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-// D2 OBL *aita-//	n/a
46	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	6	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
47	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	7	D2 OBL *aita-//D3 OBL *awa- 2dx: D1 DIR *aya-//D1 OBL *a-/ima-//	n/a
48	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	8	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
49	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	9	D3 DIR *haw-// 2dx: D2 OBL *aita-//	n/a
50	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	10	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	n/a
51	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	11	D3 DIR *haw-// D2 DIR *aiša-//D2 OBL *aita-//	n/a
52	m8	Deictic pronouns	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	>	12	D3 DIR *haw-//D3 OBL *awa- D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	n/a
53	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	0	D3 DIR *haw-//D3 OBL *awa-	n/a
54	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša- uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-//	n/a
55	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	2	D2 OBL *aita-// 2dx: D0 DIR *ha-//	n/a
56	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	3	D3 DIR *haw-//D3 OBL *awa-	n/a
57	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	=	4	D0 DIR *ha-//D2 OBL *aita- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	=
58	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	5	D3 DIR *haw-//D3 OBL *awa-	n/a

							3dx: D1 OBL *a-/ima-// D2 OBL *aita-//	
59	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	6	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
60	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	7	D2 OBL *aita-//D3 OBL *awa- 2dx: D1 DIR *aya-//D1 OBL *a-/ima-//	n/a
61	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	8	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
62	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	9	D3 DIR *haw-// 2dx: D2 OBL *aita-//	n/a
63	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	10	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	n/a
64	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	11	D3 DIR *haw-// D2 DIR *aiša-//D2 OBL *aita-//	n/a
65	m8	Deictic pronouns	4	D0 DIR *ha-//D2 OBL *aita-	>	12	D3 DIR *haw-//D3 OBL *awa- D0 DIR *ha-//D0 OBL *ta-//	n/a
				3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//			D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	
66	m8	Deictic pronouns	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	>	0	D3 DIR *haw-//D3 OBL *awa-	n/a
67	m8	Deictic pronouns	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša- uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-//	n/a
68	m8	Deictic pronouns	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	>	2	D2 OBL *aita-// 2dx: D0 DIR *ha-//	n/a
69	m8	Deictic pronouns	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	>	3	D3 DIR *haw-//D3 OBL *awa-	n/a
70	m8	Deictic pronouns	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	>	4	D0 DIR *ha-//D2 OBL *aita- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	n/a
71	m8	Deictic pronouns	5	D3 DIR *haw-//D3 OBL *awa-	=	5	D3 DIR *haw-//D3 OBL *awa-	=

72	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	1	5>6(c:1)
73	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	2	5>7(c:2)
74	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	3	5>8(c:3)
75	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	4	5>9(c:4)
76	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	2	5>10(c:2)
77	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	1	5>11(c:1)
78	m8	Deictic pronouns	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	3	5>12(c:3)
79	m8	Deictic pronouns	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
80	m8	Deictic pronouns	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a	
81	m8	Deictic pronouns	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a	
82	m8	Deictic pronouns	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a	

				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//					
83	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a	
				3dx: D1 OBL *a-/ima-//			3dx: D1 DIR *aya-//D1 OBL *a-/ima-		
				D2 OBL *aita-//			D2 DIR *aiša-//D2 OBL *aita-//		
84	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	5	D3 DIR *haw-//D3 OBL *awa-	n/a	
				3dx: D1 OBL *a-/ima-//			3dx: D1 OBL *a-/ima-//		
				D2 OBL *aita-//			D2 OBL *aita-//		
85	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	=	6	D3 DIR *haw-//D3 OBL *awa-	=	
				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//			3dx: D1 OBL *a-/ima-//		6>7(c:1)
86	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	7	D2 OBL *aita-//D3 OBL *awa-	1	
				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//			2dx: D1 DIR *aya-//D1 OBL *a-/ima-//		
87	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	8	D3 DIR *haw-//D3 OBL *awa-	n/a	
				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//			3dx: D1 OBL *a-/ima-//		6>9(c:1)
88	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	9	D3 DIR *haw-//	1	
				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//			2dx: D2 OBL *aita-//		6>10(c:1)
89	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	10	D3 DIR *haw-//D3 OBL *awa-	1	
				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//			3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//		
90	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	11	D3 DIR *haw-//	n/a	
				3dx: D1 OBL *a-/ima-//					
				D2 OBL *aita-//			D2 DIR *aiša-//D2 OBL *aita-//		
91	m8	Deictic pronouns	6	D3 DIR *haw-//D3 OBL *awa-	>	12	D3 DIR *haw-//D3 OBL *awa-	n/a	
							D0 DIR *ha-//D0 OBL *ta-//		
							D1 DIR *aya-//D1 OBL *a-/ima-		
				3dx: D1 OBL *a-/ima-//			D2 DIR *aiša-//D2 OBL *aita-//		
92	m8	Deictic pronouns	7	D2 OBL *aita-//D3 OBL *awa-	>	0	D3 DIR *haw-//D3 OBL *awa-	n/a	
				3dx: D1 OBL *a-/ima-//					
93	m8	Deictic pronouns	7	D2 OBL *aita-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a	
							uncl.: D0 OBL *ta-//		
				3dx: D1 OBL *a-/ima-//			D1 DIR *aya-//D1 OBL *a-/ima-//		
94	m8	Deictic pronouns	7	D2 OBL *aita-//D3 OBL *awa-	>	2	D2 OBL *aita-//	n/a	

95	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a
96	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a
97	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
98	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
99	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	=	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	=
100	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	n/a
101	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	n/a
102	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
103	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	n/a
104	m8	Deictic pronouns	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
105	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
106	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a
107	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a
108	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a
109	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a

110	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
111	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
112	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	n/a	
113	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	=	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	=	
114	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	1	8>9(c:1)
115	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
116	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	n/a	
117	m8	Deictic pronouns	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
118	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
119	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a	
120	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a	
121	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a	
122	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a	
123	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	

124	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
125	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	n/a
126	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	n/a
127	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	=	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	=
128	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
129	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	n/a
130	m8	Deictic pronouns	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
131	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
132	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a
133	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a
134	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a
135	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a
136	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
137	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a
138	m8	Deictic pronouns	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	n/a

				2dx: D2 OBL *aita-//			2dx: D1 DIR *aya-//D1 OBL *a-/ima-//	
139	m8	Deictic pronouns	10	D3 DIR *haw-//D3 OBL *awa- 2dx: D2 OBL *aita-//	>	8	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
140	m8	Deictic pronouns	10	D3 DIR *haw-//D3 OBL *awa- 2dx: D2 OBL *aita-//	>	9	D3 DIR *haw-// 2dx: D2 OBL *aita-//	n/a
141	m8	Deictic pronouns	10	D3 DIR *haw-//D3 OBL *awa- 2dx: D2 OBL *aita-//	=	10	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	=
142	m8	Deictic pronouns	10	D3 DIR *haw-//D3 OBL *awa- 2dx: D2 OBL *aita-//	>	11	D3 DIR *haw-// D2 DIR *aiša-//D2 OBL *aita-//	n/a
143	m8	Deictic pronouns	10	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	12	D3 DIR *haw-//D3 OBL *awa- D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	n/a
144	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	0	D3 DIR *haw-//D3 OBL *awa-	n/a
145	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša- uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-//	n/a
146	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	2	D2 OBL *aita-// 2dx: D0 DIR *ha-//	n/a
147	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	3	D3 DIR *haw-//D3 OBL *awa-	n/a
148	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	4	D0 DIR *ha-//D2 OBL *aita- 3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-//	n/a
149	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	5	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-// D2 OBL *aita-//	n/a
150	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	6	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
151	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	7	D2 OBL *aita-//D3 OBL *awa- 2dx: D1 DIR *aya-//D1 OBL *a-/ima-//	n/a
152	m8	Deictic pronouns	11	D3 DIR *haw-// 3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-//	>	8	D3 DIR *haw-//D3 OBL *awa- 3dx: D1 OBL *a-/ima-//	n/a
153	m8	Deictic pronouns	11	D3 DIR *haw-//	>	9	D3 DIR *haw-//	n/a

154	m8	Deictic pronouns	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
155	m8	Deictic pronouns	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	=	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	=	
156	m8	Deictic pronouns	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	>	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
157	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	0	D0 DIR *ha-//D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
158	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	1	New stems: D0 OBL *ta-//D2 DIR *aiša-	n/a	
159	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	2	uncl.: D0 OBL *ta-// D1 DIR *aya-//D1 OBL *a-/ima-// D2 OBL *aita-//	n/a	
160	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	3	2dx: D0 DIR *ha-// D3 DIR *haw-//D3 OBL *awa-	n/a	
161	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	4	D0 DIR *ha-//D2 OBL *aita-	n/a	
162	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	5	3dx: D1 DIR *aya-//D1 OBL *a-/ima- D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
163	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	6	3dx: D1 OBL *a-/ima-// D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	n/a	
164	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	7	3dx: D1 OBL *a-/ima-// D2 OBL *aita-//D3 OBL *awa-	n/a	
165	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	8	2dx: D1 DIR *aya-//D1 OBL *a-/ima-// D3 DIR *haw-//D3 OBL *awa-	n/a	
166	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	9	3dx: D1 OBL *a-/ima-// D3 DIR *haw-//	n/a	
167	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	10	2dx: D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	1	12>10(c:1)
168	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	>	11	3dx: D1 DIR *aya-//D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//	n/a	

169	m8	Deictic pronouns	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	=	12	D2 DIR *aiša-//D2 OBL *aita-// D3 DIR *haw-//D3 OBL *awa-	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m9	*hača + pron.1/2 person	0	*hača + orth. pers. pron.	=	0	*hača + orth. pers. pron.	=	
2	m9	*hača + pron.1/2 person	0	*hača + orth. pers. pron.	>	1	*hača + clitic pron.; *hača-mā-ka-	3	0>1(c:3)
3	m9	*hača + pron.1/2 person	0	*hača + orth. pers. pron.	>	2	*(ha)č(a) proclitic/prefix + *orth. pron.	3	0>2(c:3)
4	m9	*hača + pron.1/2 person	0	*hača + orth. pers. pron.	>	3	*hača suffixlike postposition	1	0>3(c:1)
5	m9	*hača + pron.1/2 person	0	*hača + orth. pers. pron.	>	4	Ø	1	0>4(c:1)
6	m9	*hača + pron.1/2 person	1	*hača + clitic pron.; *hača-mā-ka-	>	0	*hača + orth. pers. pron.	2	1>0(c:2)
7	m9	*hača + pron.1/2 person	1	*hača + clitic pron.; *hača-mā-ka-	=	1	*hača + clitic pron.; *hača-mā-ka-	=	
8	m9	*hača + pron.1/2 person	1	*hača + clitic pron.; *hača-mā-ka-	>	2	*(ha)č(a) proclitic/prefix + *orth. pron.	2	1>2(c:2)
9	m9	*hača + pron.1/2 person	1	*hača + clitic pron.; *hača-mā-ka-	>	3	*hača suffixlike postposition	2	1>3(c:2)
10	m9	*hača + pron.1/2 person	1	*hača + clitic pron.; *hača-mā-ka-	>	4	Ø	1	1>4(c:1)
11	m9	*hača + pron.1/2 person	2	*(ha)č(a) proclitic/prefix + *orth. pron.	>	0	*hača + orth. pers. pron.	2	2>0(c:2)
12	m9	*hača + pron.1/2 person	2	*(ha)č(a) proclitic/prefix + *orth. pron.	>	1	*hača + clitic pron.; *hača-mā-ka-	3	2>1(c:3)
13	m9	*hača + pron.1/2 person	2	*(ha)č(a) proclitic/prefix + *orth. pron.	=	2	*(ha)č(a) proclitic/prefix + *orth. pron.	=	
14	m9	*hača + pron.1/2 person	2	*(ha)č(a) proclitic/prefix + *orth. pron.	>	3	*hača suffixlike postposition	2	2>3(c:2)
15	m9	*hača + pron.1/2 person	2	*(ha)č(a) proclitic/prefix + *orth. pron.	>	4	Ø	1	2>4(c:1)
16	m9	*hača + pron.1/2 person	3	*hača suffixlike postposition	>	0	*hača + orth. pers. pron.	4	3>0(c:4)
17	m9	*hača + pron.1/2 person	3	*hača suffixlike postposition	>	1	*hača + clitic pron.; *hača-mā-ka-	4	3>1(c:4)
18	m9	*hača + pron.1/2 person	3	*hača suffixlike postposition	>	2	*(ha)č(a) proclitic/prefix + *orth. pron.	4	3>2(c:4)
19	m9	*hača + pron.1/2 person	3	*hača suffixlike postposition	=	3	*hača suffixlike postposition	=	
20	m9	*hača + pron.1/2 person	3	*hača suffixlike postposition	>	4	Ø	1	3>4(c:1)
21	m9	*hača + pron.1/2 person	4	Ø	>	0	*hača + orth. pers. pron.	n/a	
22	m9	*hača + pron.1/2 person	4	Ø	>	1	*hača + clitic pron.; *hača-mā-ka-	n/a	
23	m9	*hača + pron.1/2 person	4	Ø	>	2	*(ha)č(a) proclitic/prefix + *orth. pron.	n/a	
24	m9	*hača + pron.1/2 person	4	Ø	>	3	*hača suffixlike postposition	n/a	
25	m9	*hača + pron.1/2 person	4	Ø	=	4	Ø	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m10	*hada + pron.1/2 person	0	*hada + orth. pers. pron.	=	0	*hada + orth. pers. pron.	=	
2	m10	*hada + pron.1/2 person	0	*hada + orth. pers. pron.	>	1	*hada + clitic pron.; *hada-mā-ka-	3	0>1(c:3)
3	m10	*hada + pron.1/2 person	0	*hada + orth. pers. pron.	>	2	*(ha)d(a) as procl./pref. +*orth. pron.	3	0>2(c:3)
4	m10	*hada + pron.1/2 person	0	*hada + orth. pers. pron.	>	3	Ø	1	0>3(c:1)
5	m10	*hada + pron.1/2 person	1	*hada + clitic pron.; *hada-mā-ka-	>	0	*hada + orth. pers. pron.	2	1>0(c:2)
6	m10	*hada + pron.1/2 person	1	*hada + clitic pron.; *hada-mā-ka-	=	1	*hada + clitic pron.; *hada-mā-ka-	=	
7	m10	*hada + pron.1/2 person	1	*hada + clitic pron.; *hada-mā-ka-	>	2	*(ha)d(a) as procl./pref. +*orth. pron.	2	1>2(c:2)
8	m10	*hada + pron.1/2 person	1	*hada + clitic pron.; *hada-mā-ka-	>	3	Ø	1	1>3(c:1)
9	m10	*hada + pron.1/2 person	2	*(ha)d(a) as procl./pref. +*orth. pron.	>	0	*hada + orth. pers. pron.	2	2>0(c:2)
10	m10	*hada + pron.1/2 person	2	*(ha)d(a) as procl./pref. +*orth. pron.	>	1	*hada + clitic pron.; *hada-mā-ka-	3	2>1(c:3)
11	m10	*hada + pron.1/2 person	2	*(ha)d(a) as procl./pref. +*orth. pron.	=	2	*(ha)d(a) as procl./pref. +*orth. pron.	=	
12	m10	*hada + pron.1/2 person	2	*(ha)d(a) as procl./pref. +*orth. pron.	>	3	Ø	1	2>3(c:1)
13	m10	*hada + pron.1/2 person	3	Ø	>	0	*hada + orth. pers. pron.	n/a	
14	m10	*hada + pron.1/2 person	3	Ø	>	1	*hada + clitic pron.; *hada-mā-ka-	n/a	
15	m10	*hada + pron.1/2 person	3	Ø	>	2	*(ha)d(a) as procl./pref. +*orth. pron.	n/a	
16	m10	*hada + pron.1/2 person	3	Ø	=	3	Ø	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m11	*abi + pron.1/2 person	0	*abi + orth. pers. pron.	=	0	*abi + orth. pers. pron.	=	
2	m11	*abi + pron.1/2 person	0	*abi + orth. pers. pron.	>	1	*abi + clitic pron.; *abi-mā-ka-	3	0>1(c:3)
3	m11	*abi + pron.1/2 person	0	*abi + orth. pers. pron.	>	2	*(a)b(i) as procl./pref. +*orth. pron.	3	0>2(c:3)
4	m11	*abi + pron.1/2 person	0	*abi + orth. pers. pron.	>	3	Ø	1	0>3(c:1)
5	m11	*abi + pron.1/2 person	1	*abi + clitic pron.; *abi-mā-ka-	>	0	*abi + orth. pers. pron.	=	

6	m11	*abi + pron.1/2 person	1	*abi + clitic pron.; *abi-mā-ka-	=	1	*abi + clitic pron.; *abi-mā-ka-	=	
7	m11	*abi + pron.1/2 person	1	*abi + clitic pron.; *abi-mā-ka-	>	2	*(a)b(i) as procl./pref. +*orth. pron.	2	1>2(c:2)
8	m11	*abi + pron.1/2 person	1	*abi + clitic pron.; *abi-mā-ka-	>	3	Ø	n/a	
9	m11	*abi + pron.1/2 person	2	*(a)b(i) as procl./pref. +*orth. pron.	>	0	*abi + orth. pers. pron.	2	2>0(c:2)
10	m11	*abi + pron.1/2 person	2	*(a)b(i) as procl./pref. +*orth. pron.	>	1	*abi + clitic pron.; *abi-mā-ka-	3	2>1(c:3)
11	m11	*abi + pron.1/2 person	2	*(a)b(i) as procl./pref. +*orth. pron.	=	2	*(a)b(i) as procl./pref. +*orth. pron.	=	
12	m11	*abi + pron.1/2 person	2	*(a)b(i) as procl./pref. +*orth. pron.	>	3	Ø	1	2>3(c:1)
13	m11	*abi + pron.1/2 person	3	Ø	>	0	*abi + orth. pers. pron.	n/a	
14	m11	*abi + pron.1/2 person	3	Ø	>	1	*abi + clitic pron.; *abi-mā-ka-	n/a	
15	m11	*abi + pron.1/2 person	3	Ø	>	2	*(a)b(i) as procl./pref. +*orth. pron.	n/a	
16	m11	*abi + pron.1/2 person	3	Ø	=	3	Ø	=	

#	**id**	**isogloss**	**input state**	**input**		**output state**	**output**	**cost**	
1	m12	*upari + pron.1/2 person	0	*upari + orth. pers. pron.	=	0	*upari + orth. pers. pron.	=	
2	m12	*upari + pron.1/2 person	0	*upari + orth. pers. pron.	>	1	*upari + clitic pron.; *upari-mā-ka-	3	0>1(c:3)
3	m12	*upari + pron.1/2 person	0	*upari + orth. pers. pron.	>	2	Ø	1	0>2(c:1)
4	m12	*upari + pron.1/2 person	1	*upari + clitic pron.; *upari-mā-ka-	>	0	*upari + orth. pers. pron.	2	1>0(c:2)
5	m12	*upari + pron.1/2 person	1	*upari + clitic pron.; *upari-mā-ka-	=	1	*upari + clitic pron.; *upari-mā-ka-	=	
6	m12	*upari + pron.1/2 person	1	*upari + clitic pron.; *upari-mā-ka-	>	2	Ø	1	1>2(c:1)
7	m12	*upari + pron.1/2 person	2	Ø	>	0	*upari + orth. pers. pron.	n/a	
8	m12	*upari + pron.1/2 person	2	Ø	>	1	*upari + clitic pron.; *upari-mā-ka-	n/a	
9	m12	*upari + pron.1/2 person	2	Ø	=	2	Ø	=	

#	**id**	**isogloss**	**input state**	**input**		**output state**	**output**	**cost**	
1	m13	*hača-DEM	0	absence of fused form	=	0	absence of fused form	=	

2	m13	*hača-DEM	0	absence of fused form	>	1	*haca-DEM & *haca-DEM-antara	3	0>1(c:3)
3	m13	*hača-DEM	0	absence of fused form	>	2	*haca-DEM-antara	3	0>2(c:3)
4	m13	*hača-DEM	1	*haca-DEM & *haca-DEM-antara	>	0	absence of fused form	1	1>0(c:1)
5	m13	*hača-DEM	1	*haca-DEM & *haca-DEM-antara	=	1	*haca-DEM & *haca-DEM-antara	=	
6	m13	*hača-DEM	1	*haca-DEM & *haca-DEM-antara	>	2	*haca-DEM-antara	1	1>2(c:1)
7	m13	*hača-DEM	2	*haca-DEM-antara	>	0	absence of fused form	1	2>0(c:1)
8	m13	*hača-DEM	2	*haca-DEM-antara	>	1	*haca-DEM & *haca-DEM-antara	1	2>1(c:1)
9	m13	*hača-DEM	2	*haca-DEM-antara	=	2	*haca-DEM-antara	=	

#	**id**	**isogloss**	**input state**	**input**		**output state**	**output**	**cost**	
1	m14	prod. of ADP+dem+antara-	0	absence of fused form	=	0	absence of fused form	=	
2	m14	prod. of ADP+dem+antara-	0	absence of fused form	>	1	*hada-, *upari-, *ana-, *tara- + *DEM-antara	3	0>1(c:3)
3	m14	prod. of ADP+dem+antara-	1	*hada-, *upari-, *ana-, *tara- + *DEM-antara	>	0	absence of fused form	1	1>0(c:1)
4	m14	prod. of ADP+dem+antara-	1	*hada-, *upari-, *ana-, *tara- + *DEM-antara	=	1	*hada-, *upari-, *ana-, *tara- + *DEM-antara	=	

#	**id**	**isogloss**	**input state**	**input**		**output state**	**output**	**cost**	
1	m15	prod. of ADP+DEM+suff.	0	absence of fused form	=	0	*hača, *hada, *upari, *ana, *tara- + DEM + *-da-, *arda-	=	
2	m15	prod. of ADP+DEM+suff.	0	absence of fused form	>	1	*hača, *hada, *upari, *ana, *tara- + DEM + *-da-, *arda-	3	0>1(c:3)
3	m15	prod. of ADP+DEM+suff.	1	*hača, *hada, *upari, *ana, *tara- + DEM + *-da-, *arda-	>	0	absence of fused form	1	1>0(c:1)
4	m15	prod. of ADP+DEM+suff.	1	*hača, *hada, *upari, *ana, *tara- + DEM + *-da-, *arda-	=	1	*hača, *hada, *upari, *ana, *tara- + DEM + *-da-, *arda-	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m16	DEM + *arda-	0	absent	=	0	absent	=	
2	m16	DEM + *arda-	0	absent	>	1	fused form: *ima-arda-, *ta-arda-, *awa-arda-, etc.	2	0>1(c:2)
3	m16	DEM + *arda-	1	fused form: *ima-arda-, *ta-arda-, *awa-arda-, etc.	>	0	absent	1	1>0(c:1)
4	m16	DEM + *arda-	1	fused form: *ima-arda-, *ta-arda-, *awa-arda-, etc.	=	1	fused form: *ima-arda-, *ta-arda-, *awa-arda-, etc.	=	

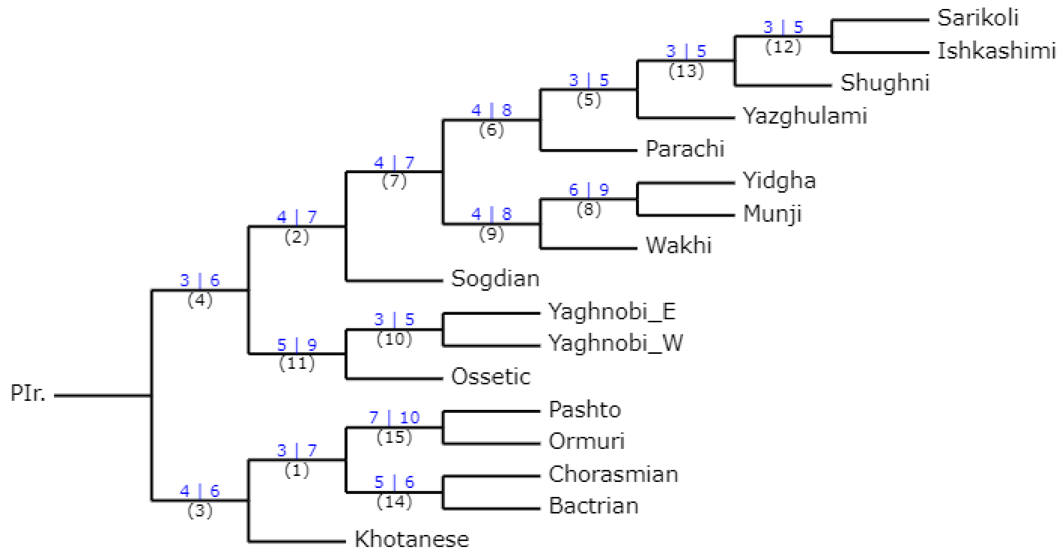
#	id	isogloss	input state	input		output state	output	cost	
1	m17	DEM-*θra-	0	absent	=	0	absent	=	
2	m17	DEM-*θra-	0	absent	>	1	fused form: *ima-θra-, *awa-θra-, etc.; *ku-θra-	2	0>1(c:2)
3	m17	DEM-*θra-	1	fused form: *ima-θra-, *awa-θra-, etc.; *ku-θra-	>	0	absent	1	1>0(c:1)
4	m17	DEM-*θra-	1	fused form: *ima-θra-, *awa-θra-, etc.; *ku-θra-	=	1	fused form: *ima-θra-, *awa-θra-, etc.; *ku-θra-	=	

#	id	isogloss	input state	input		output state	output	cost	
1	m18	DEM-*da-	0	absent	=	0	absent	=	
2	m18	DEM-*da-	0	absent	>	1	fused form: *ima-da-, *ta-da-, *awa-da-, etc.	2	0>1(c:2)
3	m18	DEM-*da-	1	fused form: *ima-da-, *ta-da-, *awa-da-, etc.	>	0	absent	1	1>0(c:1)
4	m18	DEM-*da-	1	fused form: *ima-da-, *ta-da-, *awa-da-, etc.	=	1	fused form: *ima-da-, *ta-da-, *awa-da-, etc.	=	

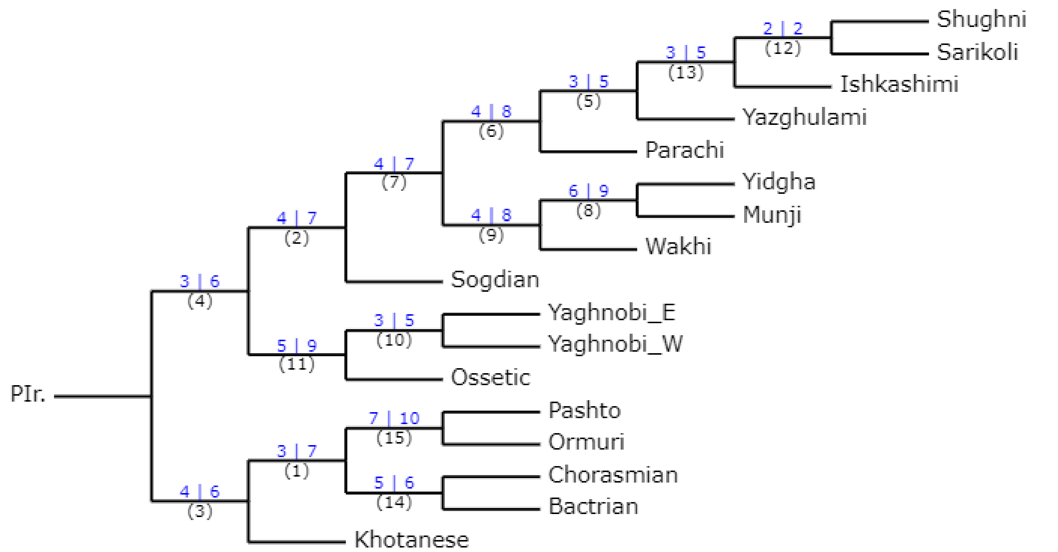
#	id	isogloss	input state	input		output state	output	cost	
1	m19	DEM-*da-aida- (?)	0	absent	=	0	absent	=	
2	m19	DEM-*da-aida- (?)	0	absent	>	1	fused form: *ima-da-aida-, *ta-da-aida-, *awa-da-aida-, etc.	3	0>1(c:3)
3	m19	DEM-*da-aida- (?)	1	fused form: *ima-da-aida-, *ta-da-aida-, *awa-da-aida-, etc.	>	0	absent	1	1>0(c:1)
4	m19	DEM-*da-aida- (?)	1	fused form: *ima-da-aida-, *ta-da-aida-, *awa-da-aida-, etc.	=	1	fused form: *ima-da-aida-, *ta-da-aida-, *awa-da-aida-, etc.	=	

7.7. Appendix 7: A Sample of the best trees of the final run

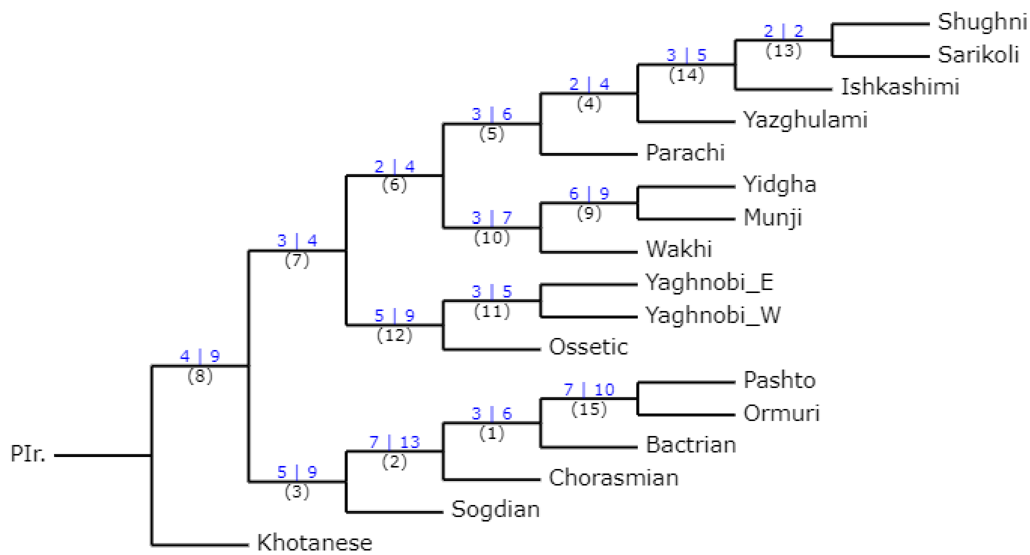
"Tree 1" Score 655 (a) (paup 6)



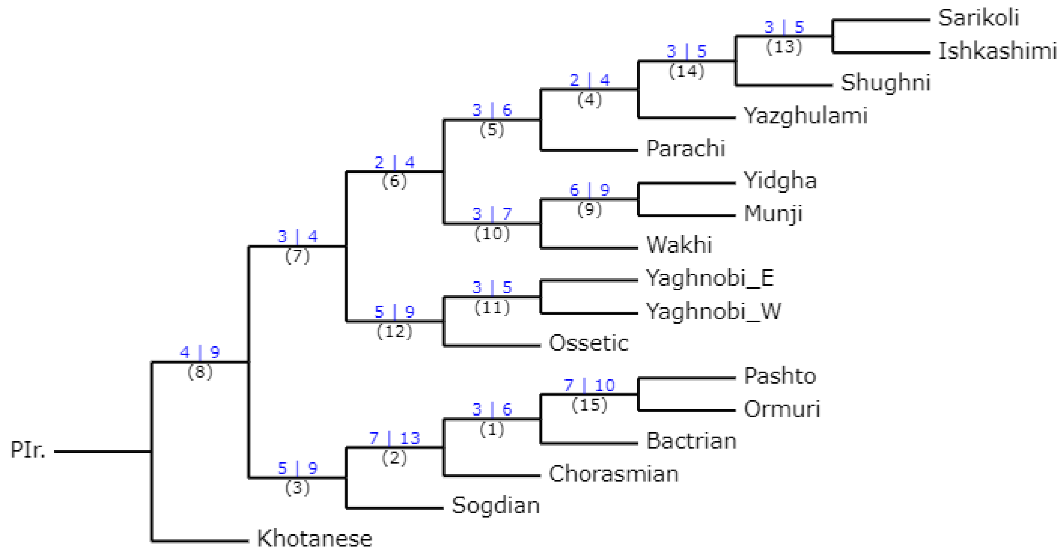
Score 655 (b) (paup 7)



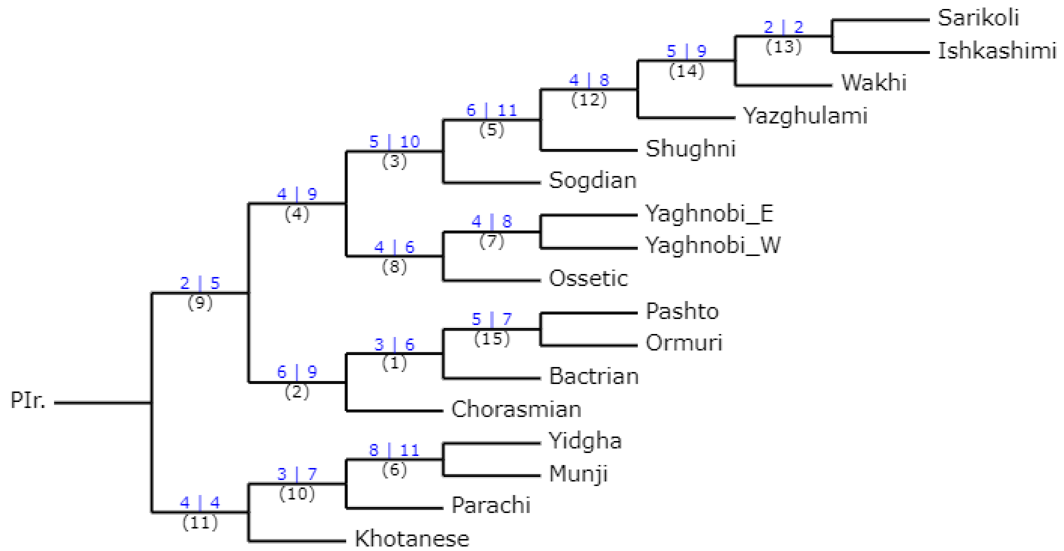
"Tree 2" Score 656 (a) (paup 2)



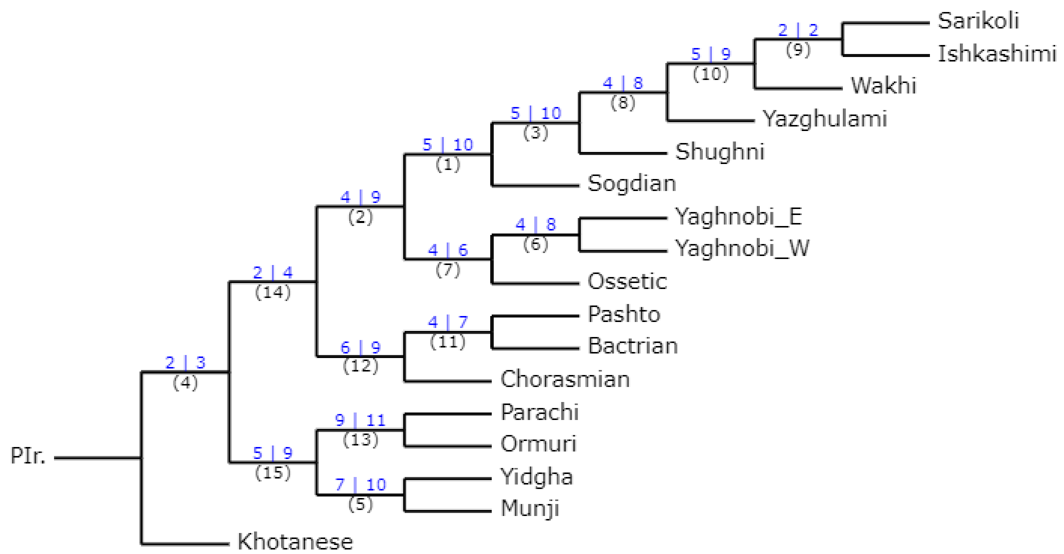
Score 656 (b) (paup 4)



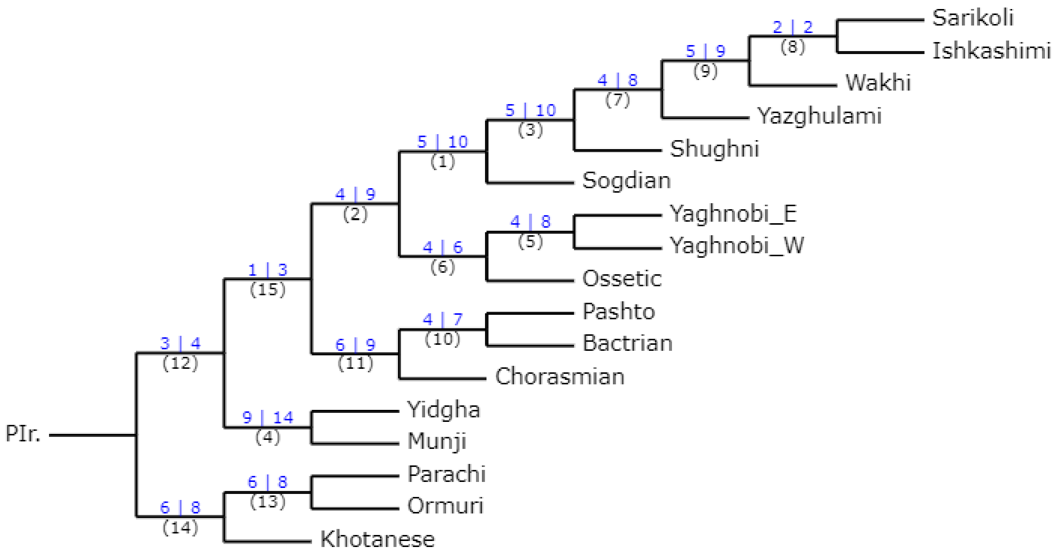
“Tree 3” **Score 656 (c) (paup 20)**



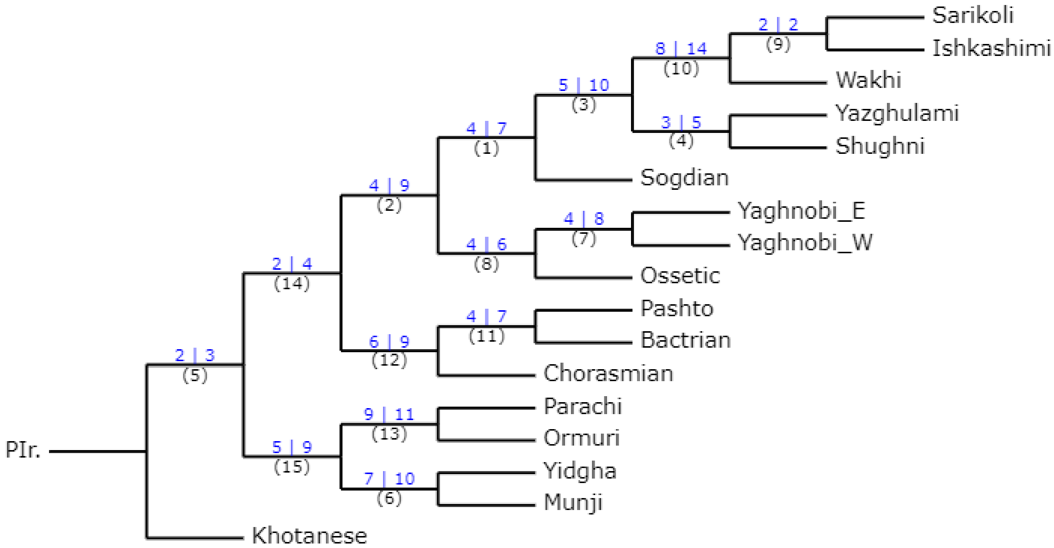
“Tree 4” **Score 657_(16)**



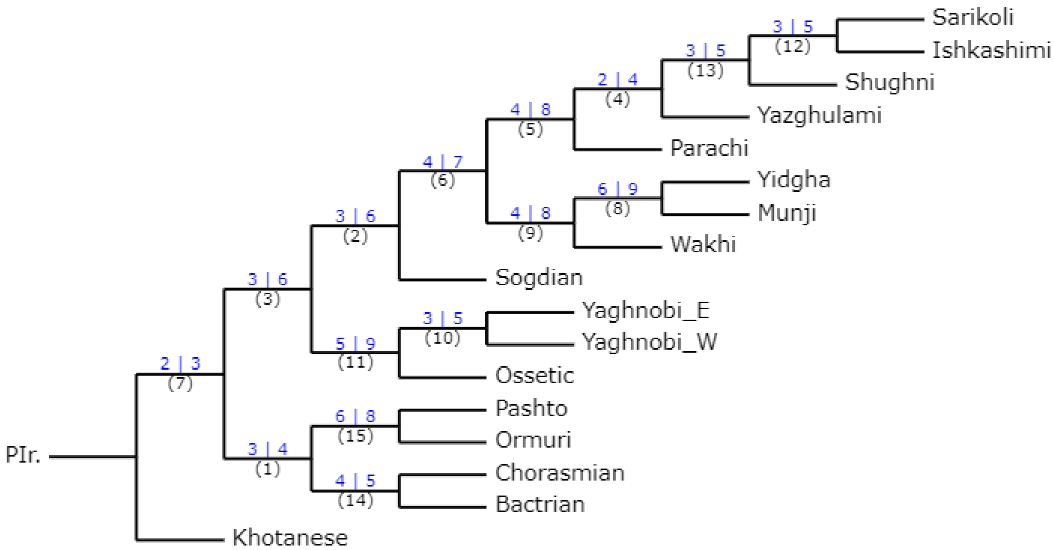
“Tree 5” Score 658 (a) (paup 14)



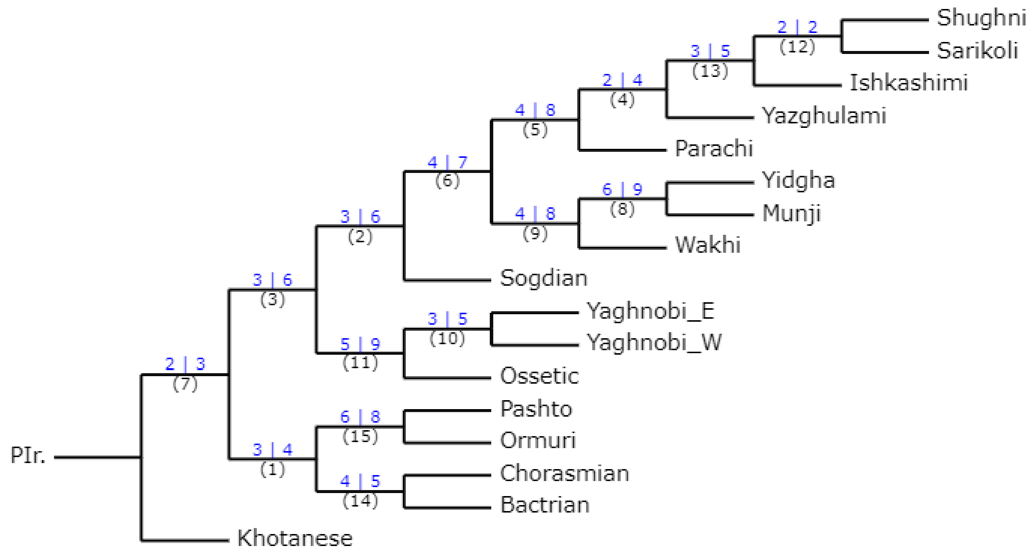
Score 658 (b) (paup 28)



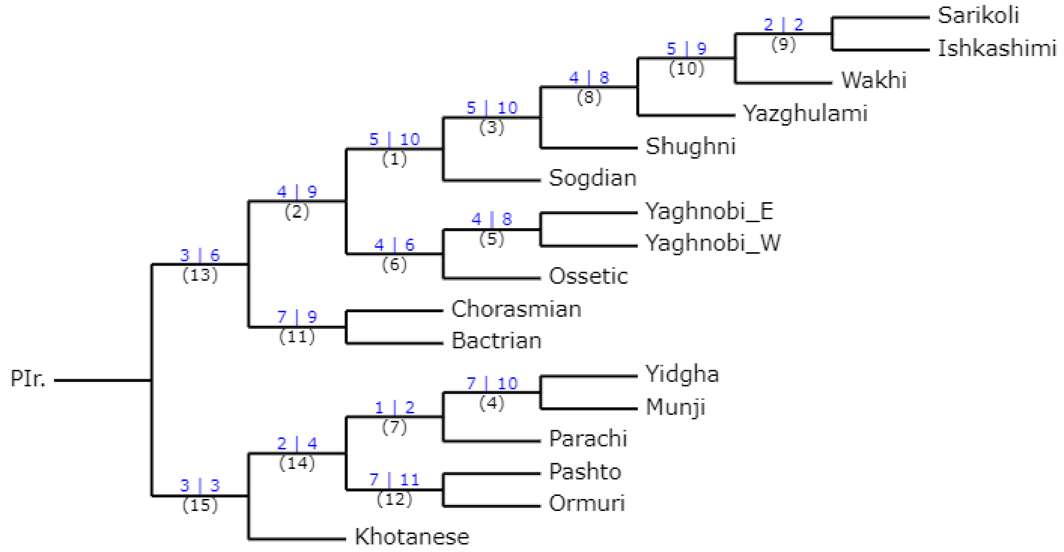
Score 659 (a) (paup 5)



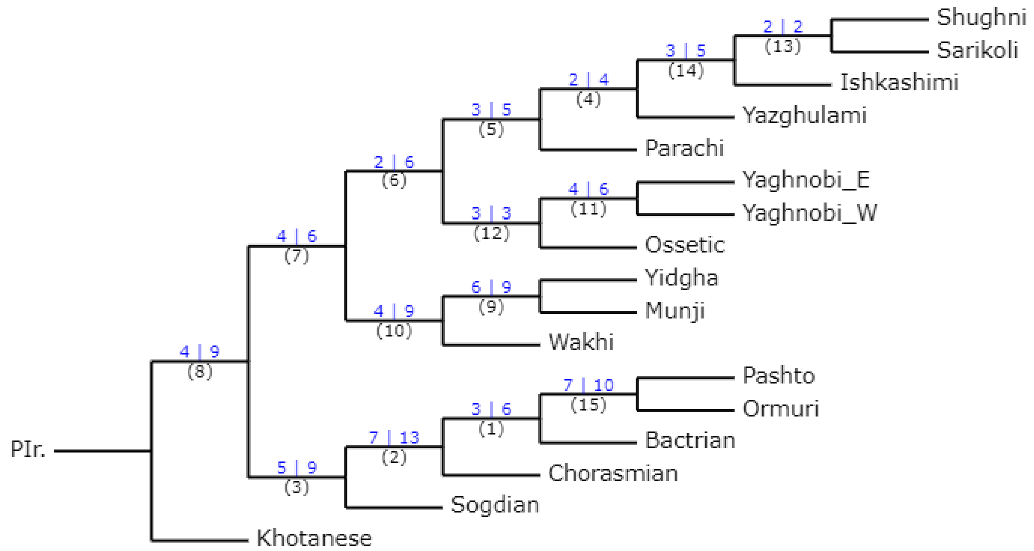
Score 659 (b) (paup 8)



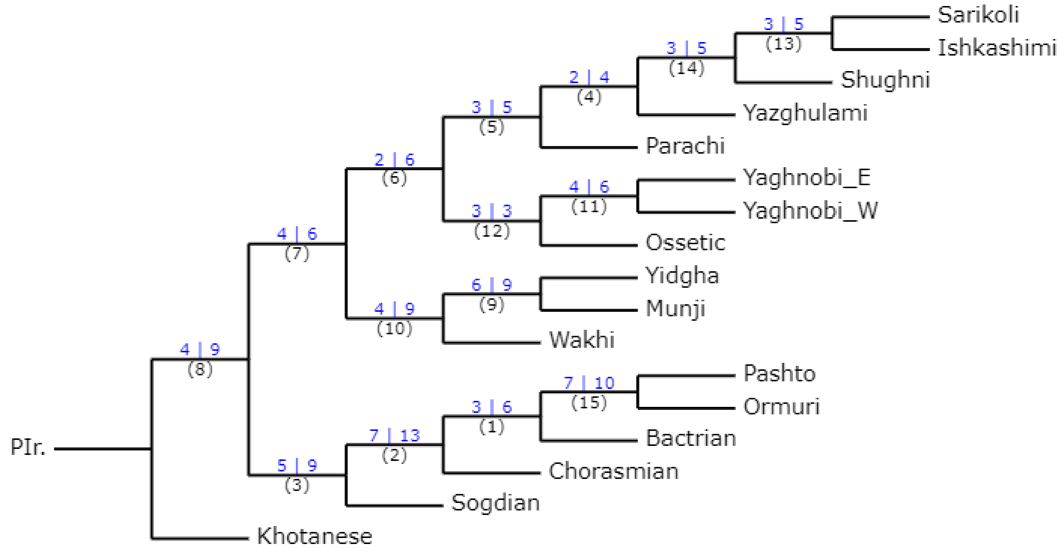
Score 659 (c) (paup 31)



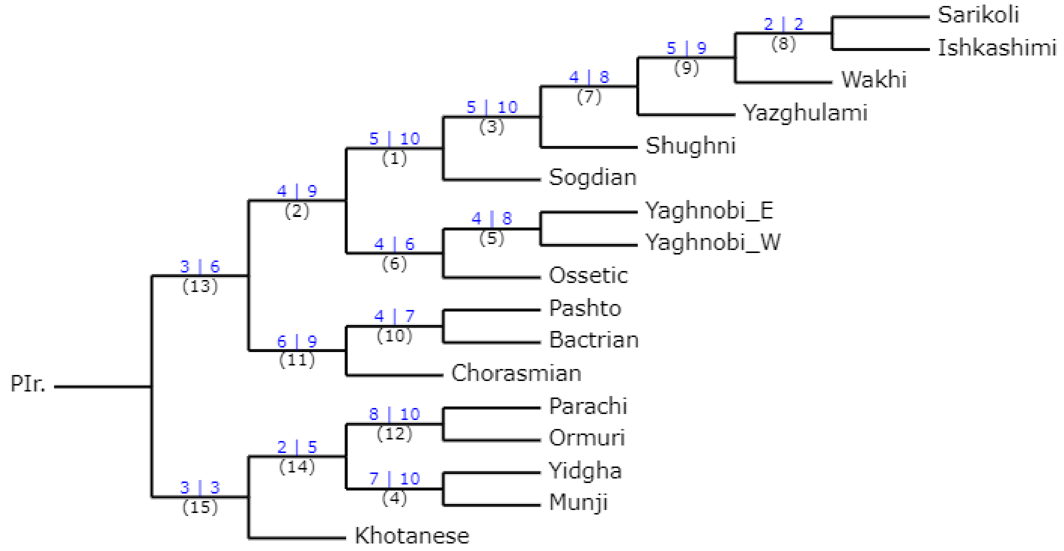
Score 660 (a) (paup 1)



Score 660 (b) (paup 3)



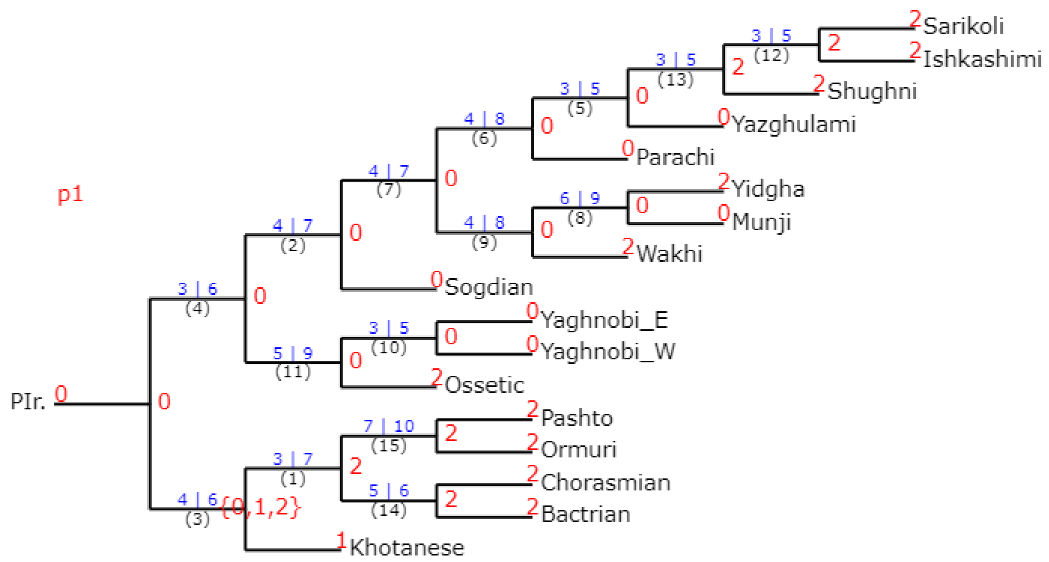
Score 660 (c) (paup 11)



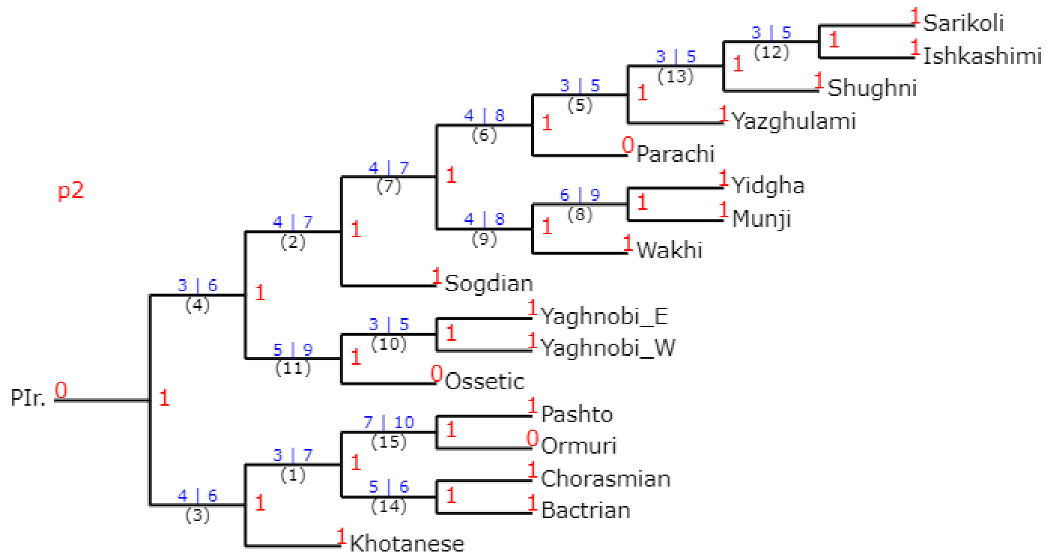
7.8 The inferred development of all characters on Tree 1 and Tree 2

7.8.1. Tree 1

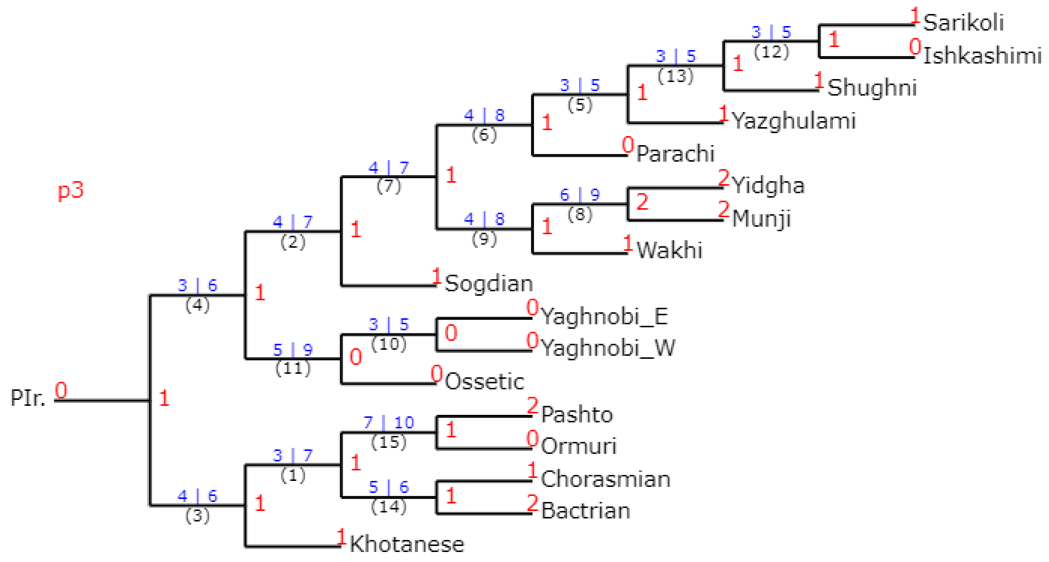
p1



p2

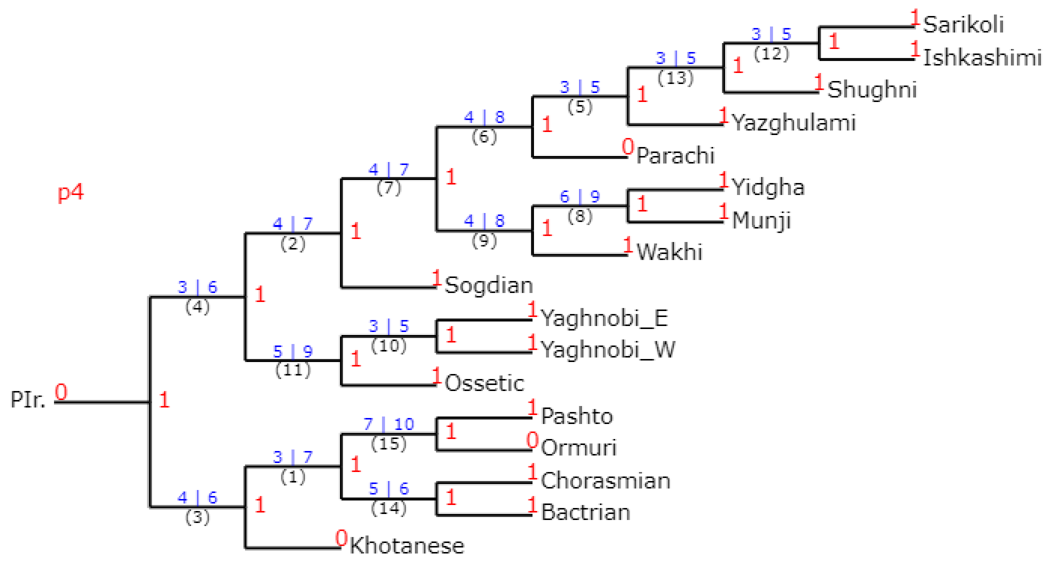


p3



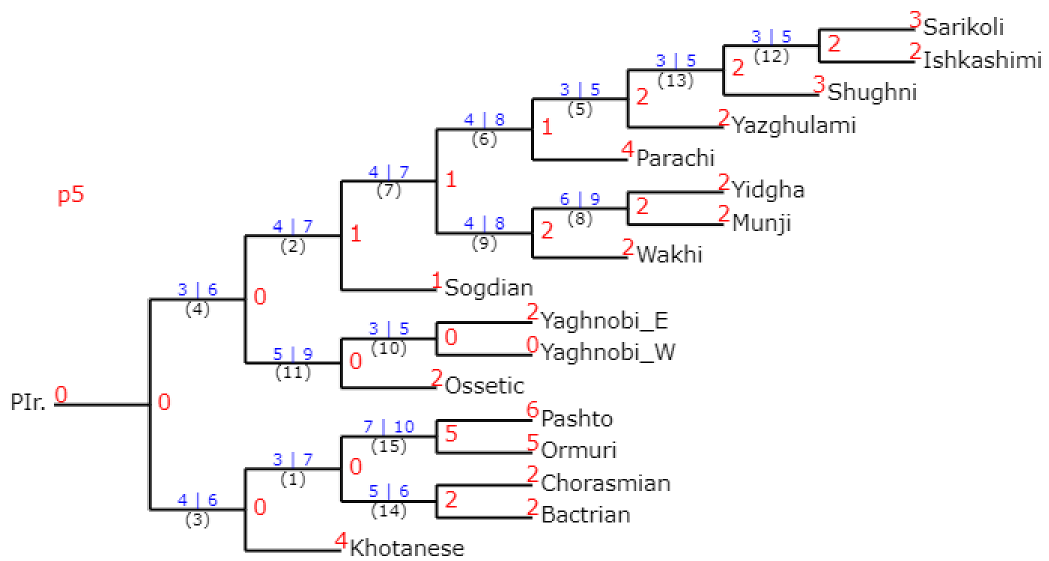
p3
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

p4



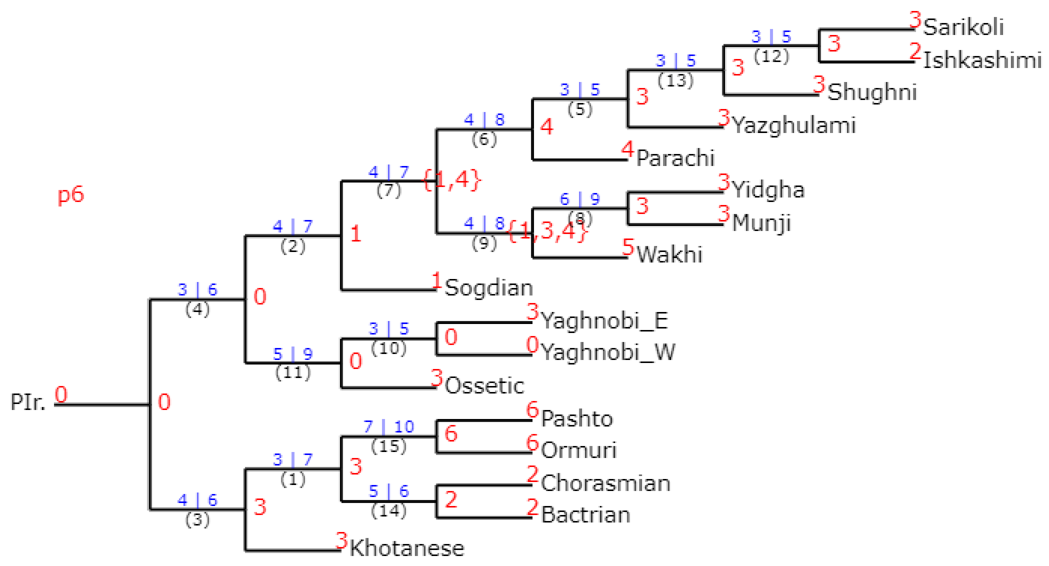
p4
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

p5

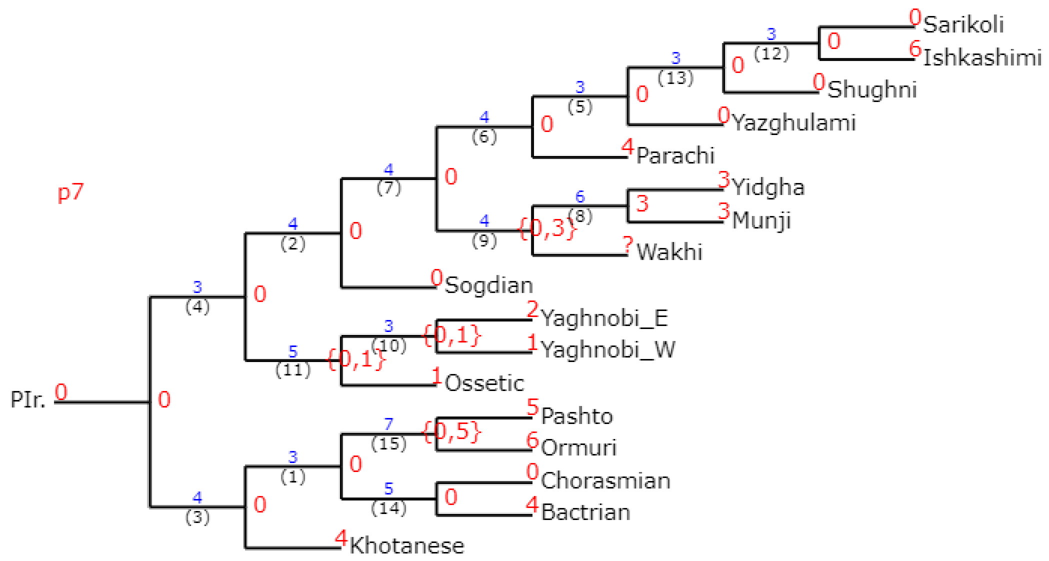


p5
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

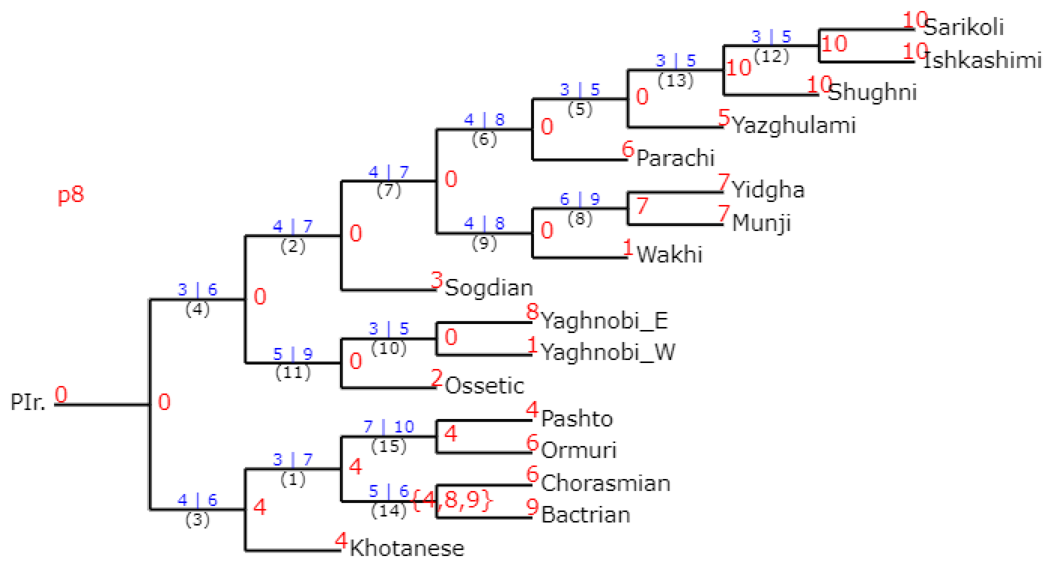
p6


p6
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

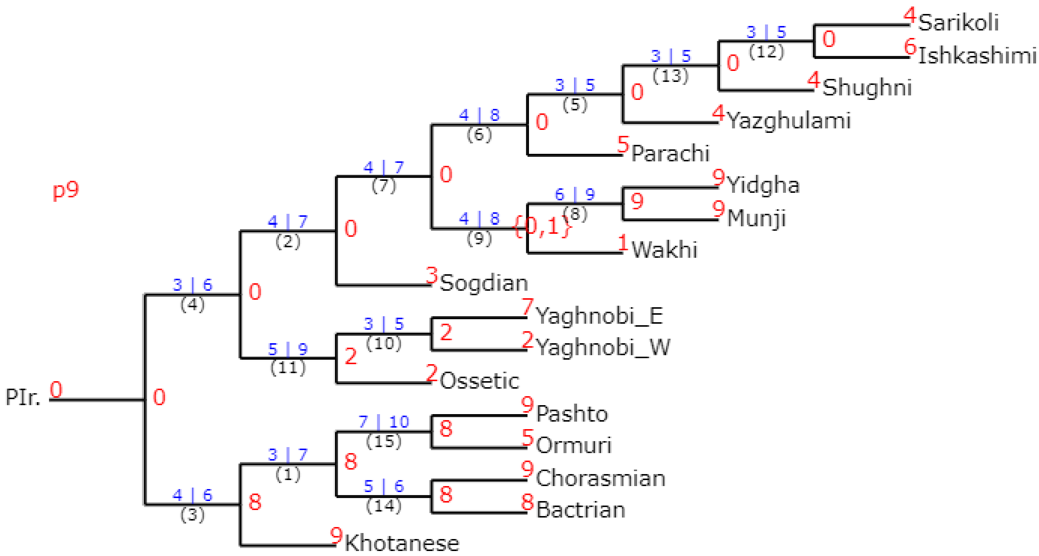
p7


p7
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

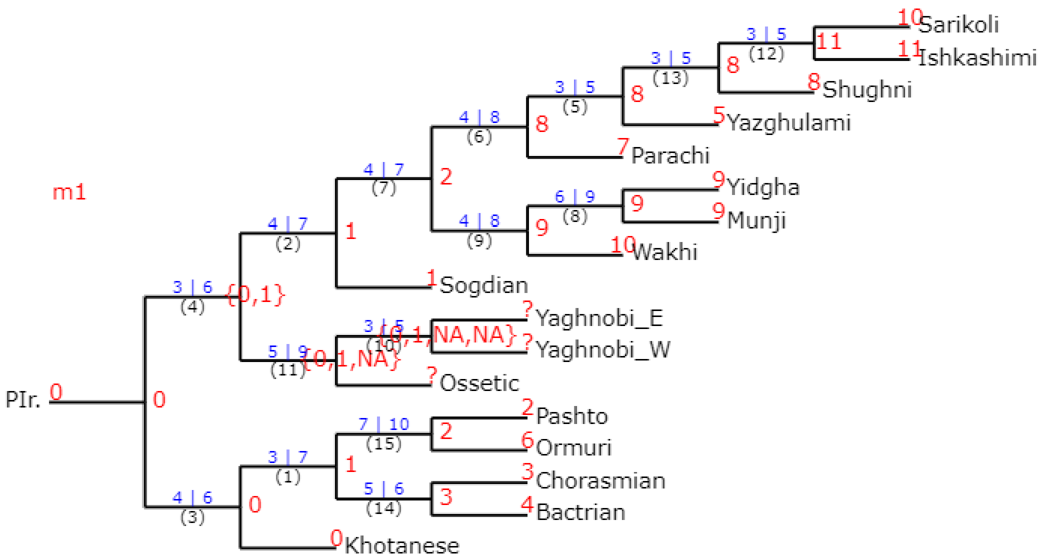
p8


p8
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

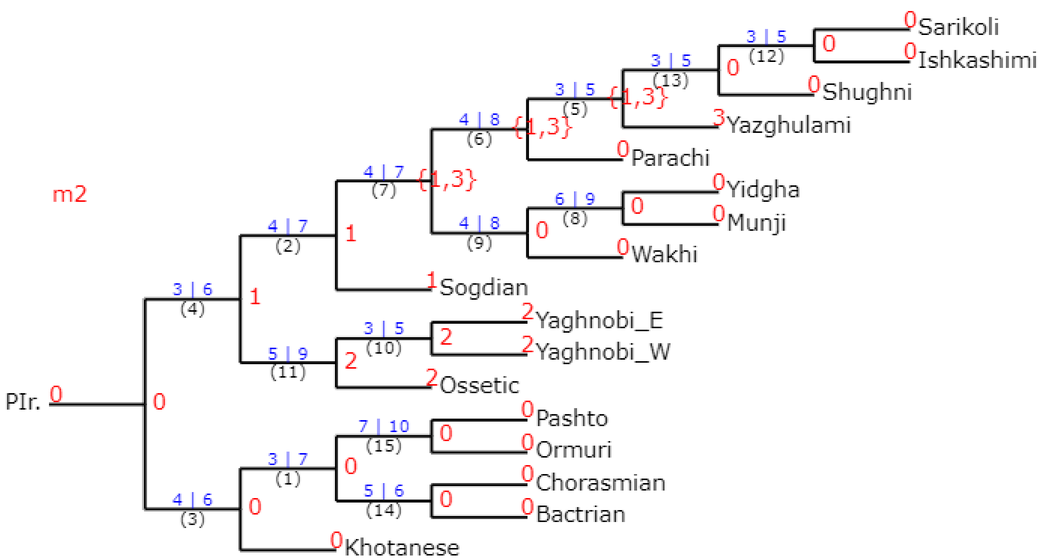
p9


p9
PIr. 0
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

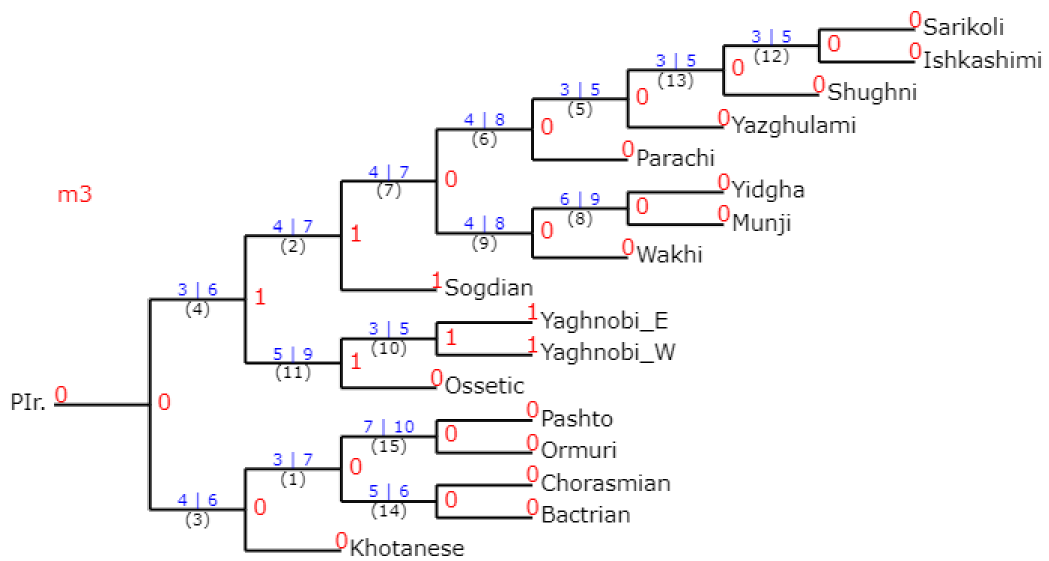
m1


m1
PIr. 0
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

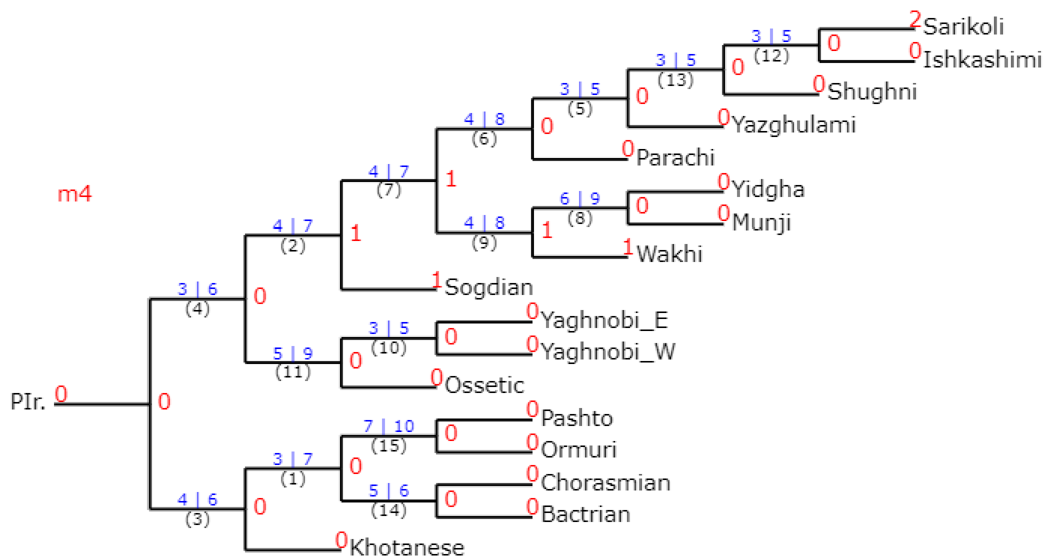
m2


m2
PIr. 0
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

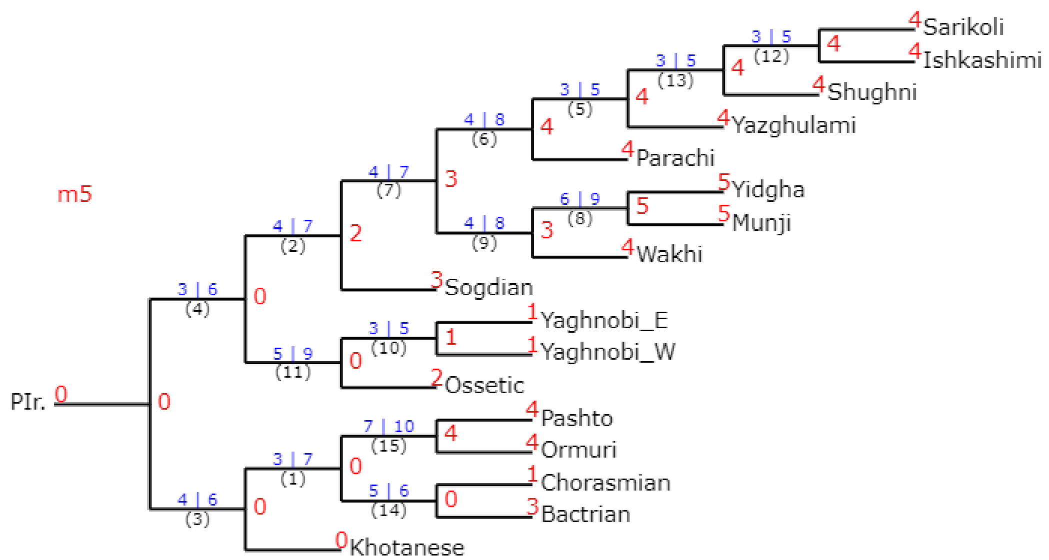
m3


m3
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

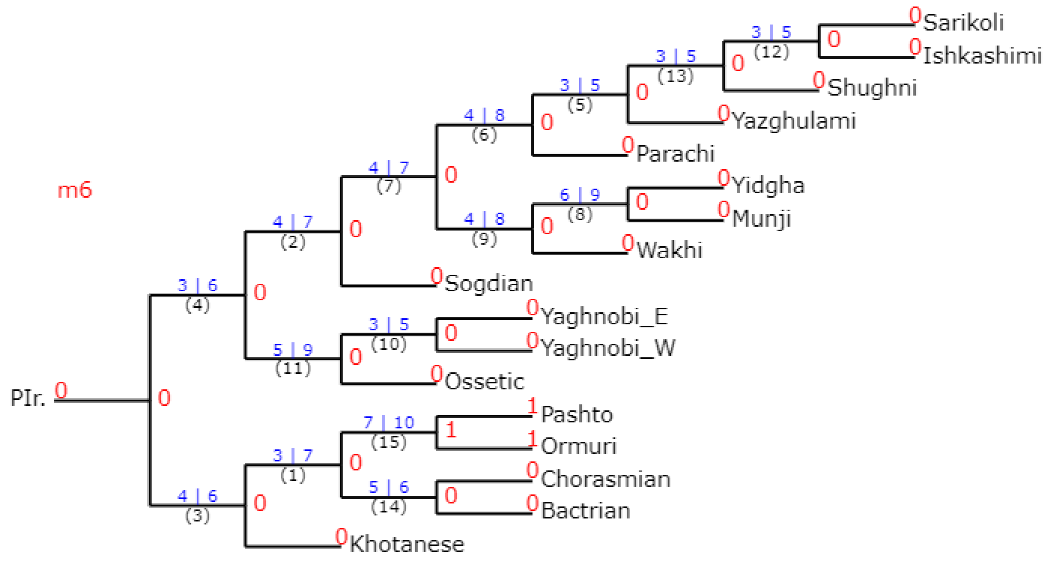
m4


m4
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

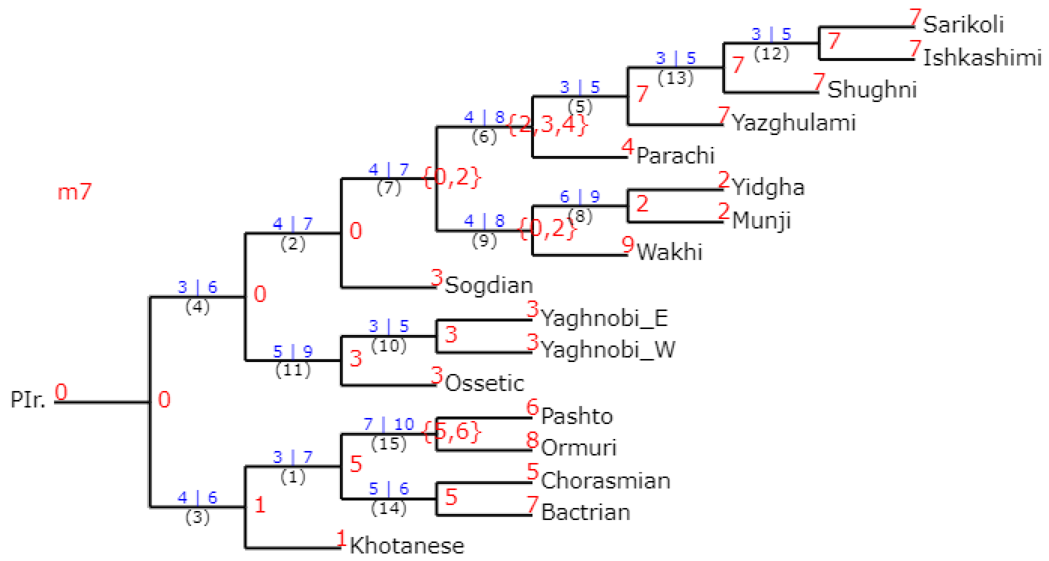
m5


m5
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

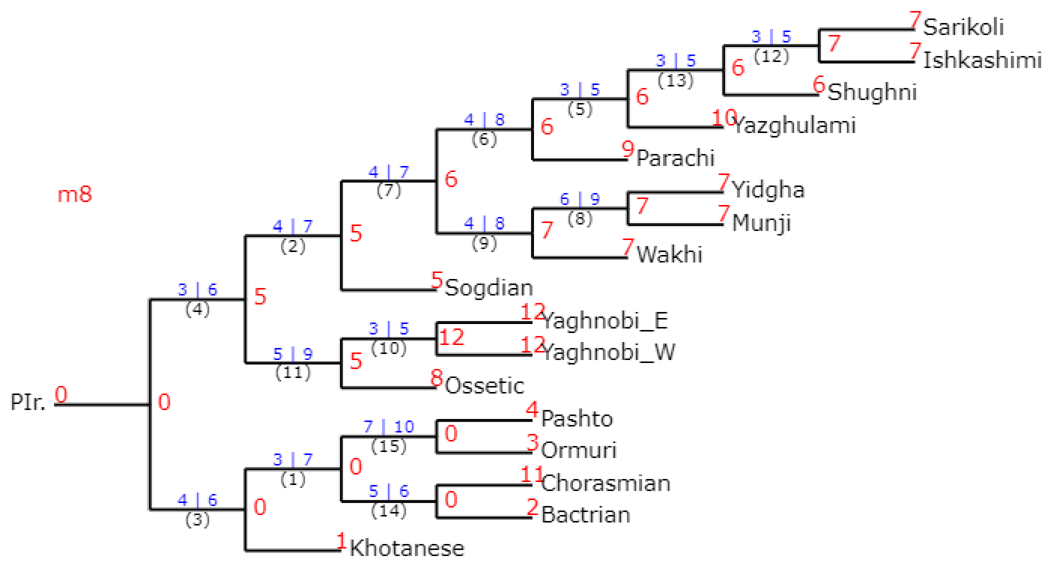
m6


m6
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

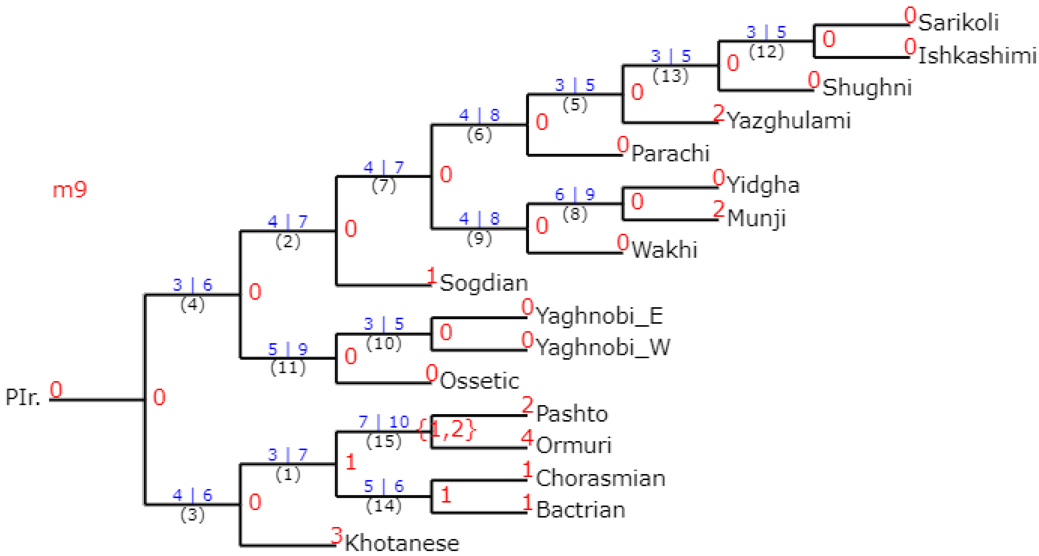
m7


m7
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

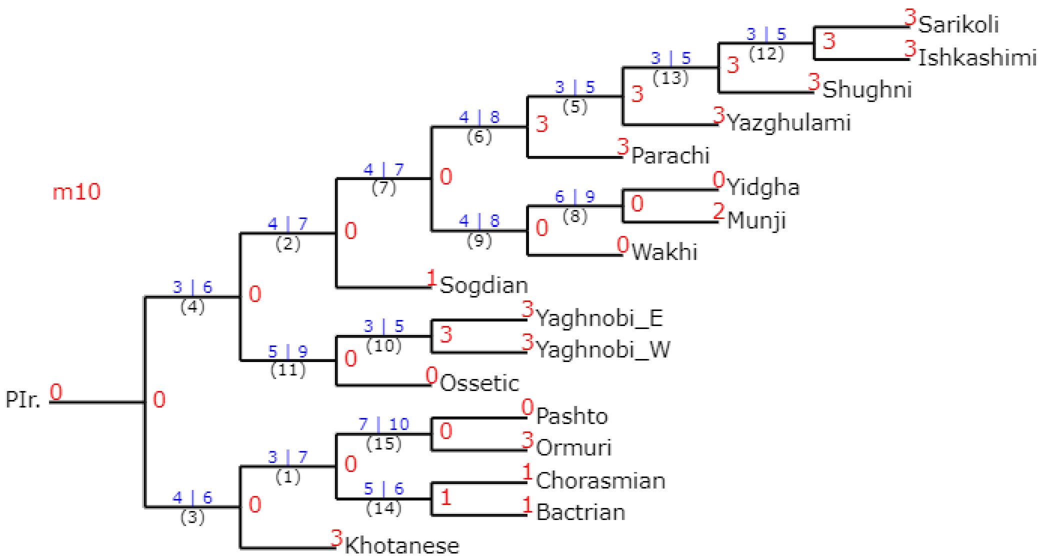
m8


m8
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

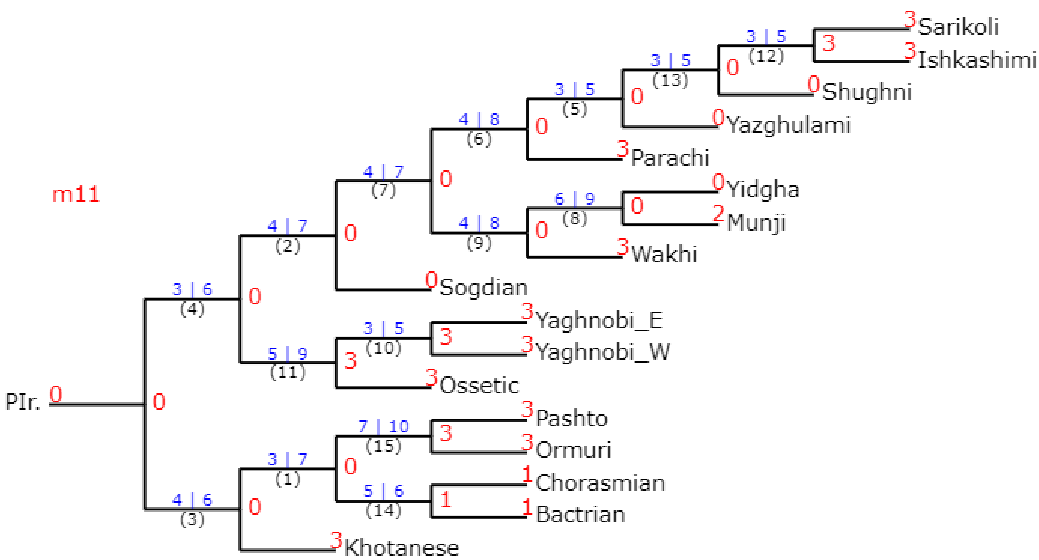
m9


m9
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

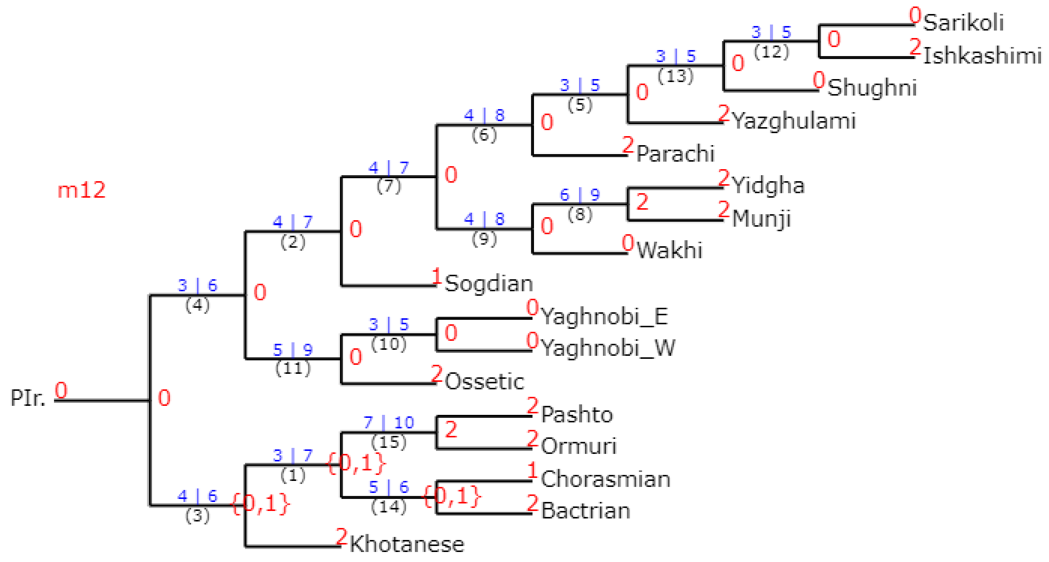
m10


m10
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

m11

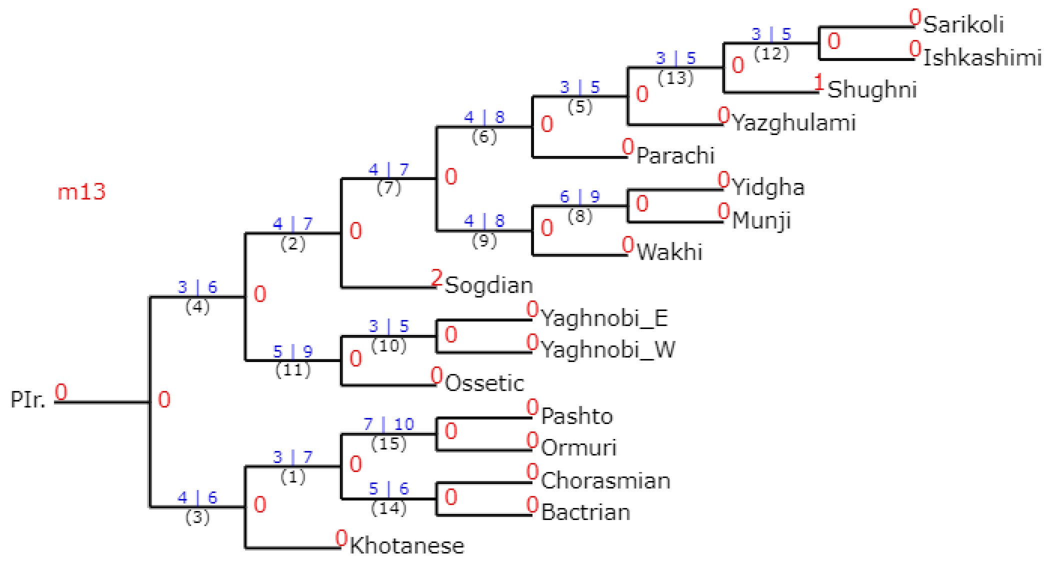

m11
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

m12



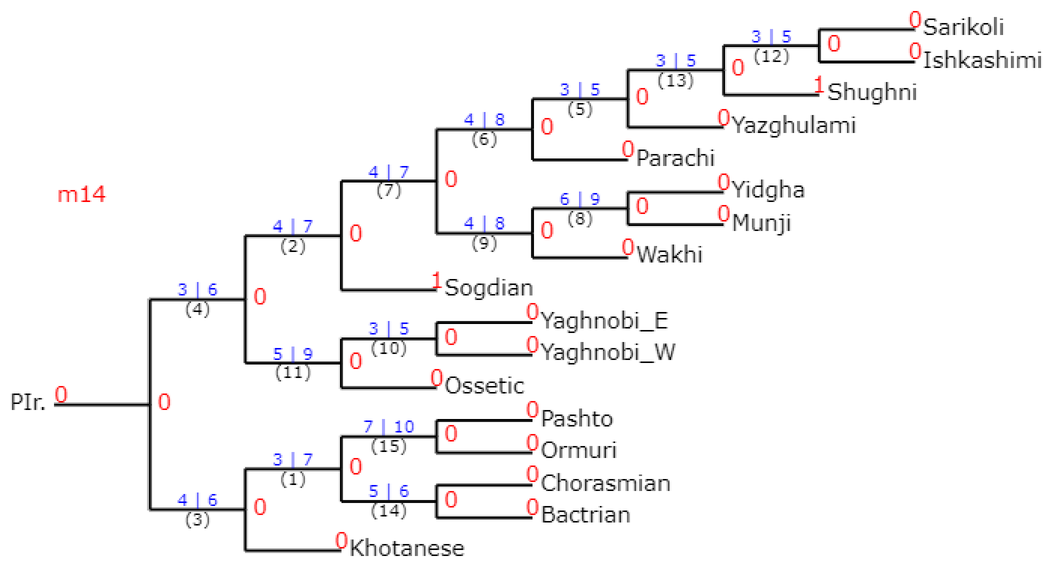
m12
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

m13



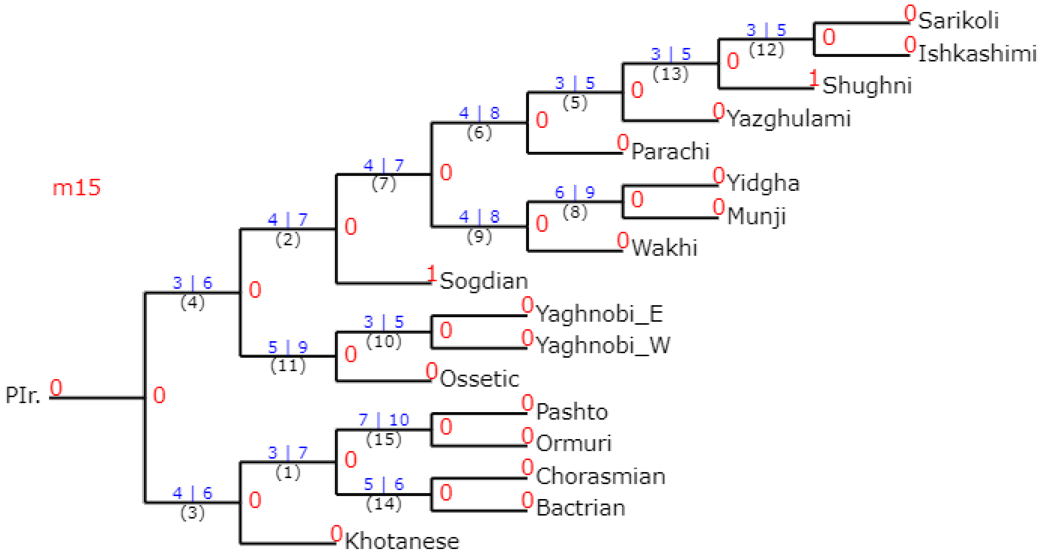
m13
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

m14

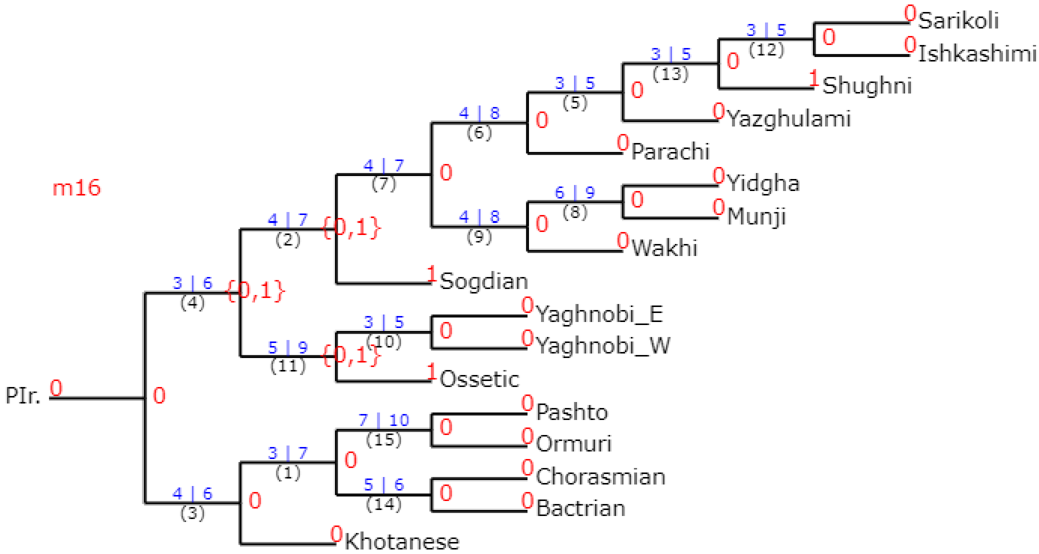


m14
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

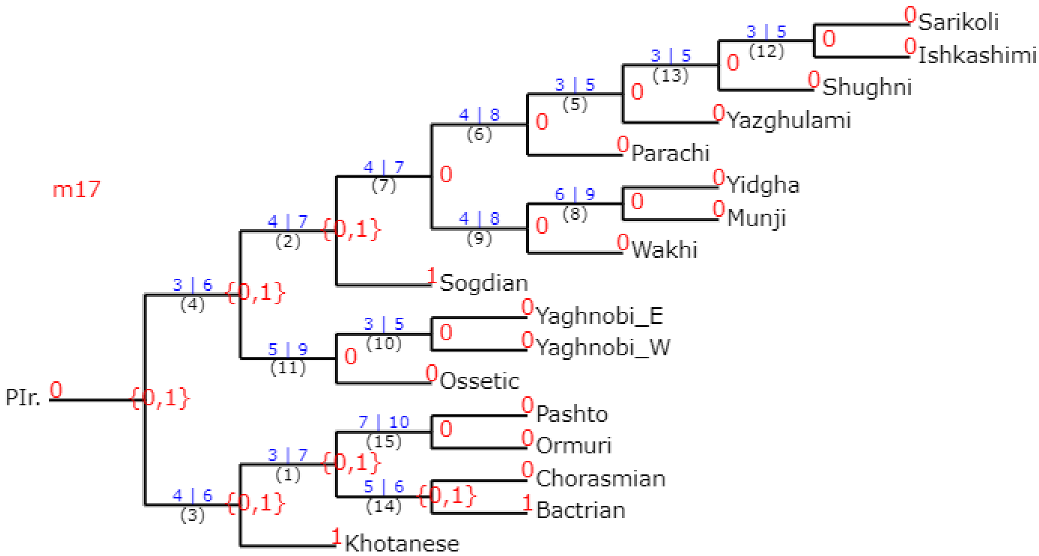
m15


m15
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

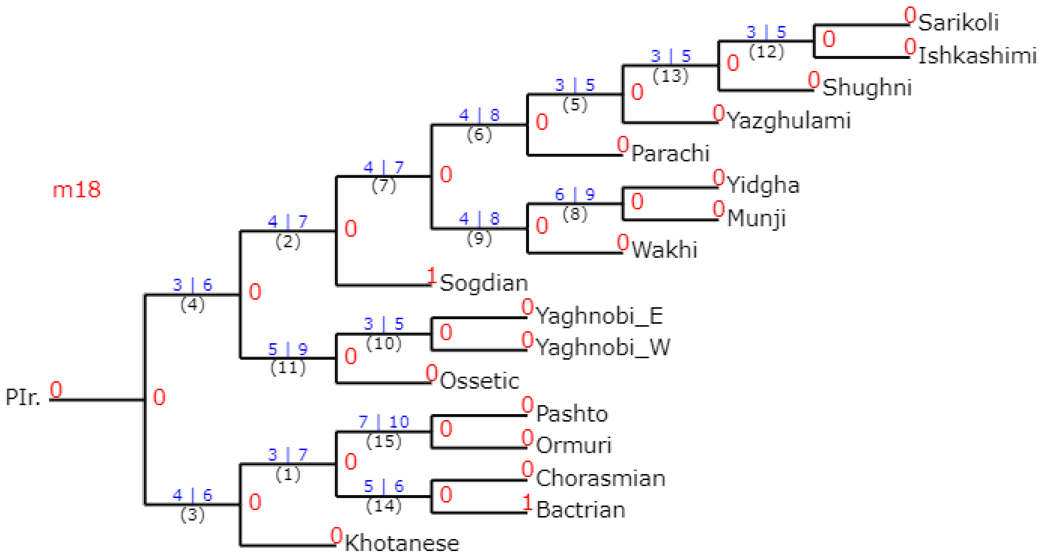
m16


m16
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

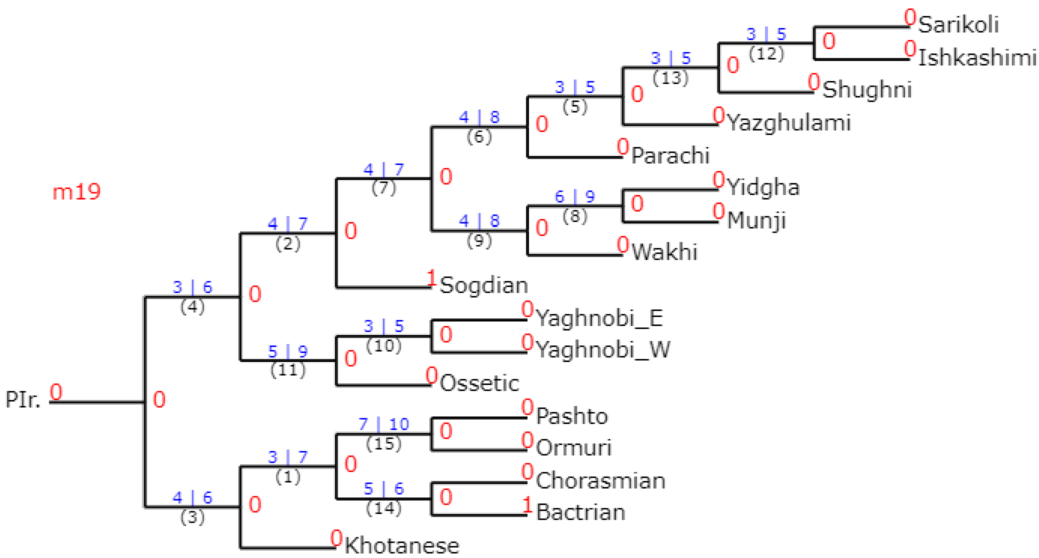
m17


m17
PIr.
Sarikoli
Ishkashimi
Shughni
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Sogdian
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Chorasmian
Bactrian
Khotanese

m18

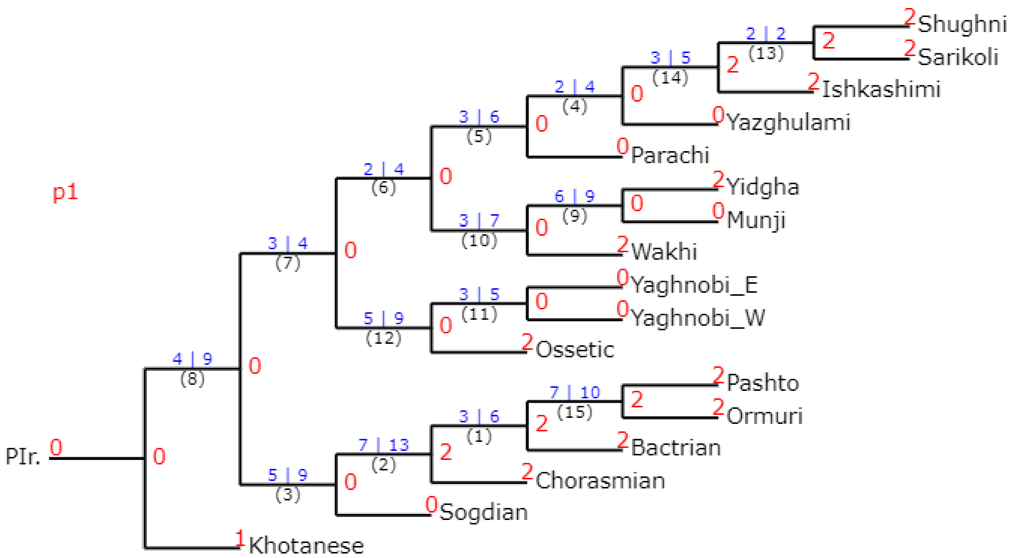


m19

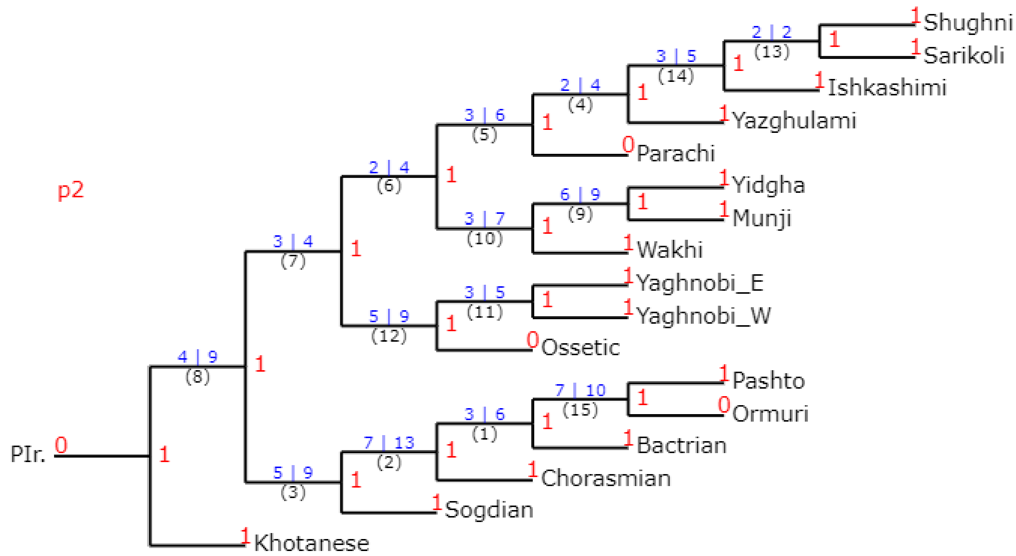


7.8.2. Tree 2

p1

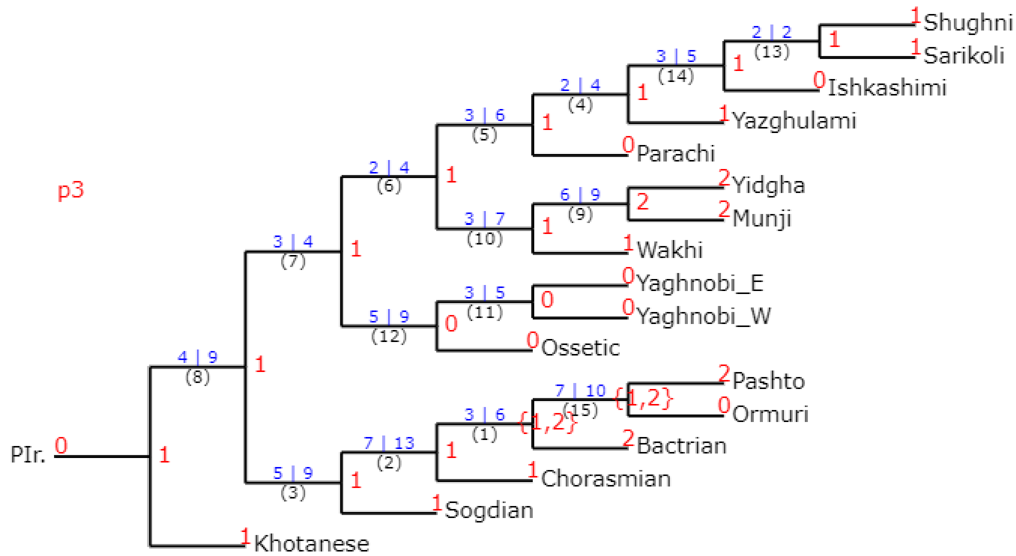


p2



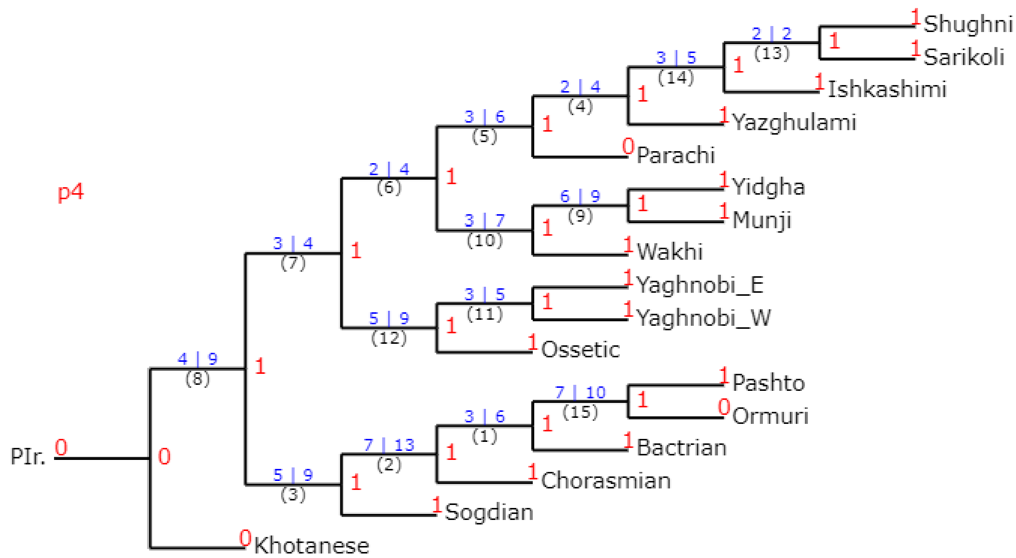
p2
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese
2 | 2
(13)
3 | 5
(14)
2 | 4
(4)
3 | 6
(5)
2 | 4
(6)
6 | 9
(9)
3 | 7
(10)
3 | 4
(7)
3 | 5
(11)
5 | 9
(12)
4 | 9
(8)
7 | 10
(15)
3 | 6
(1)
7 | 13
(2)
5 | 9
(3)

p3



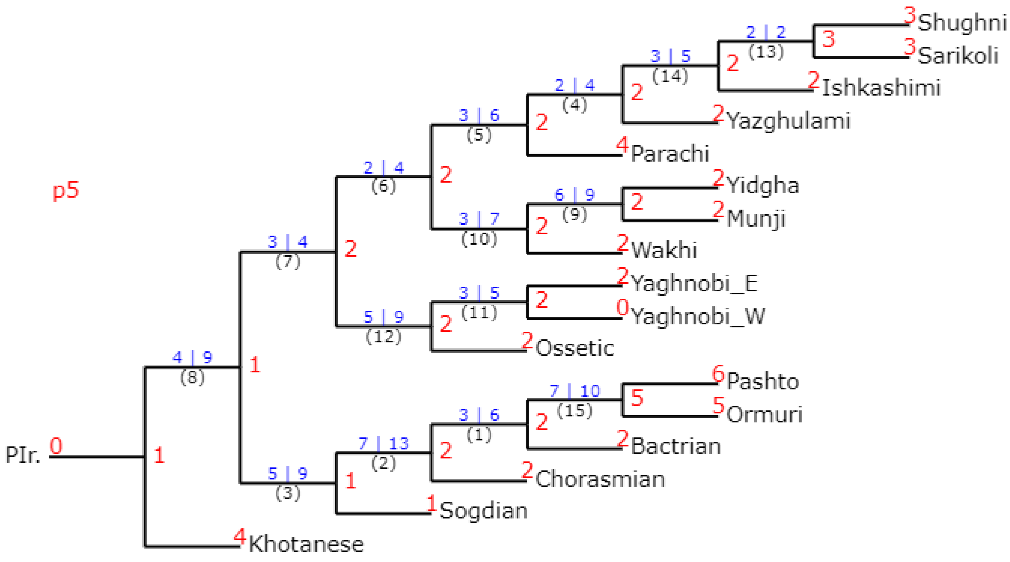
p3
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese
{1,2}
{1,2}

p4

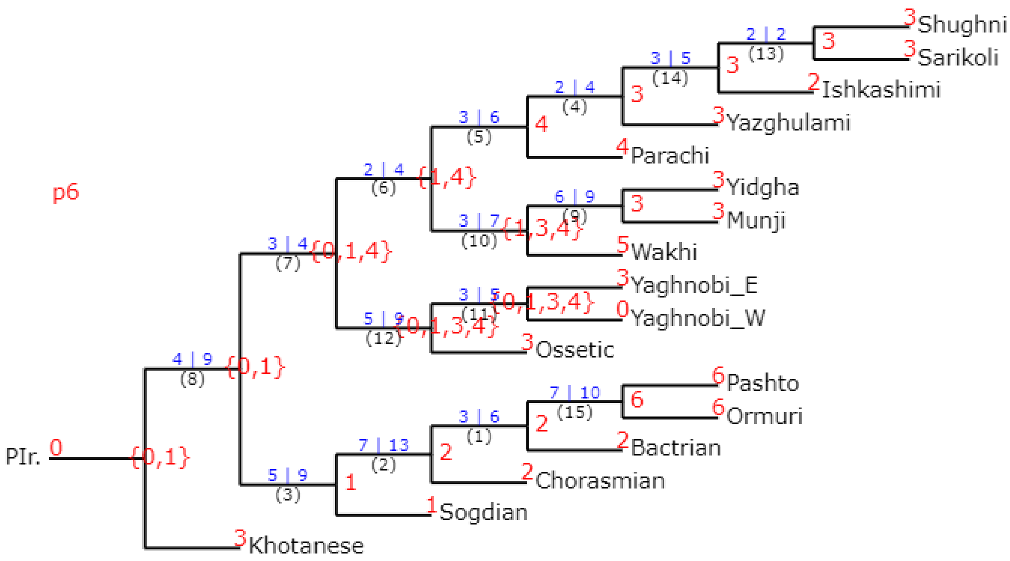


p4
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

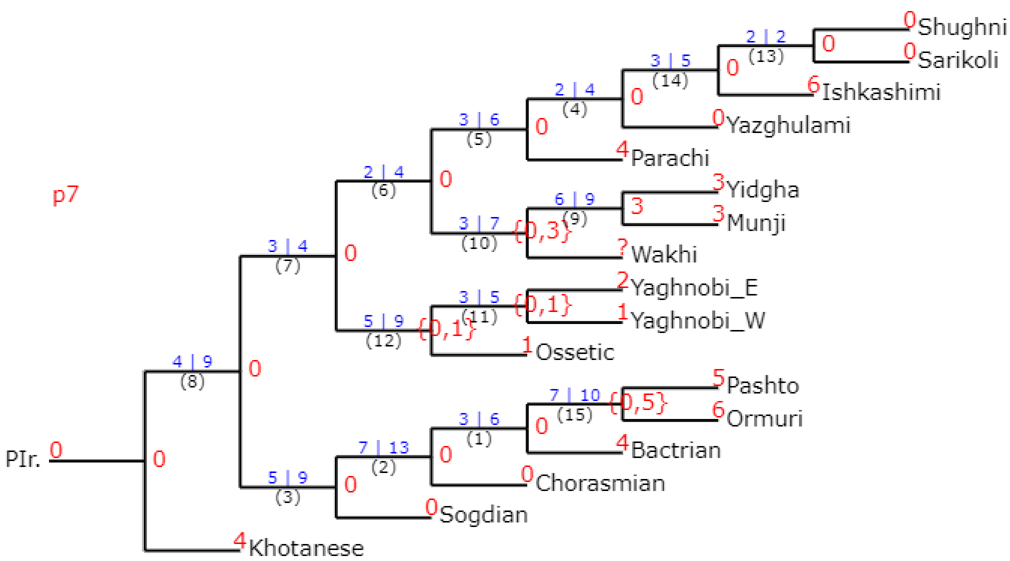
p5


p5
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

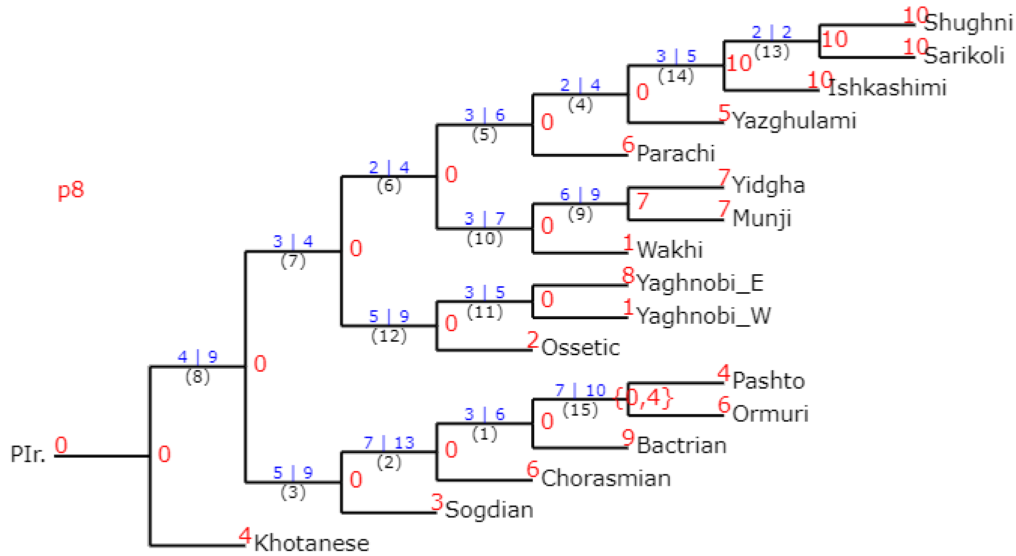
p6


p6
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

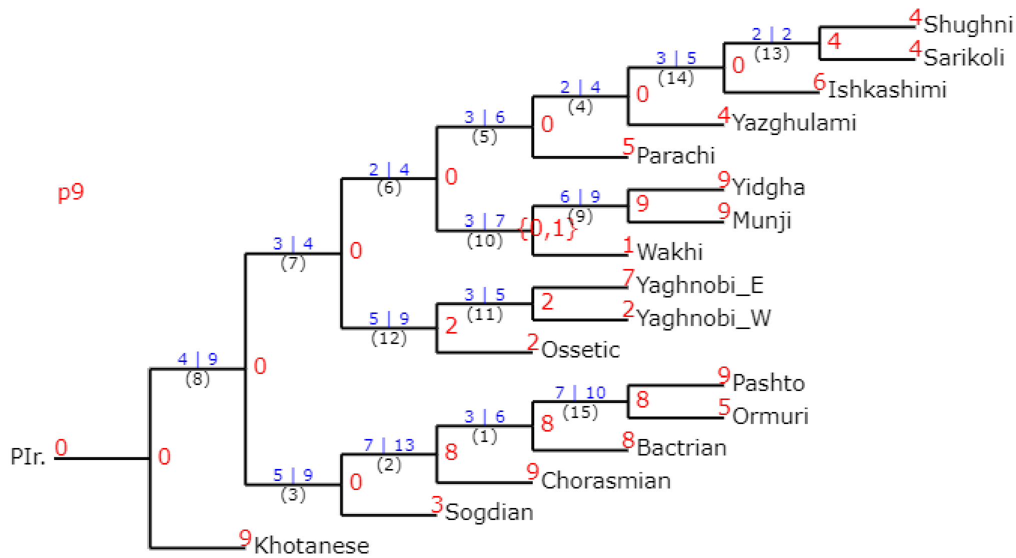
p7


p7
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

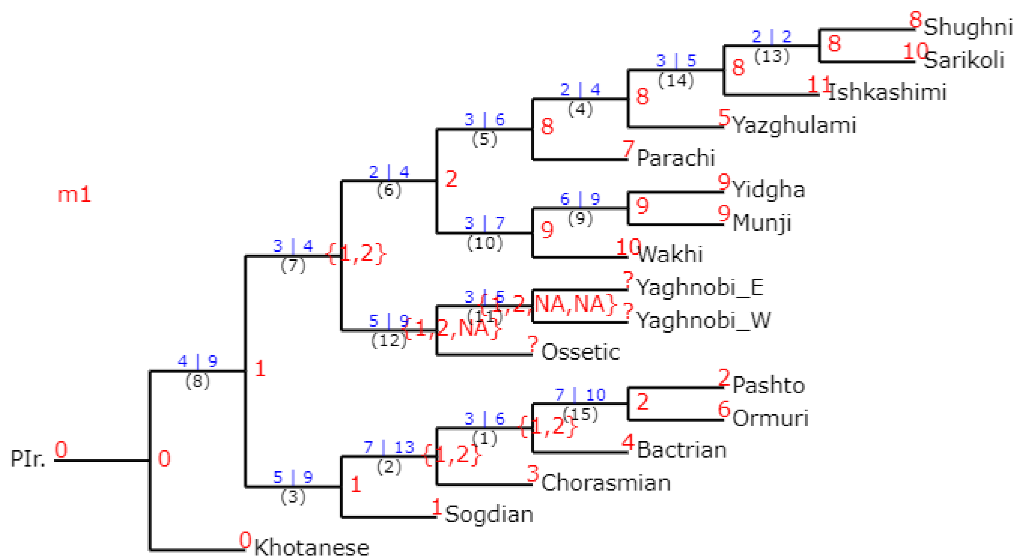
p8


p8
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

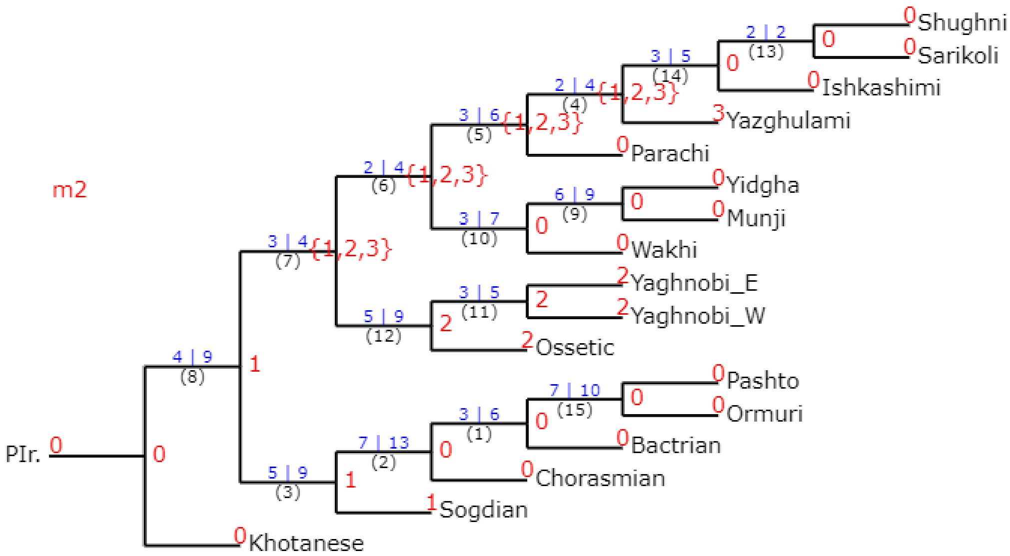
p9


p9
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

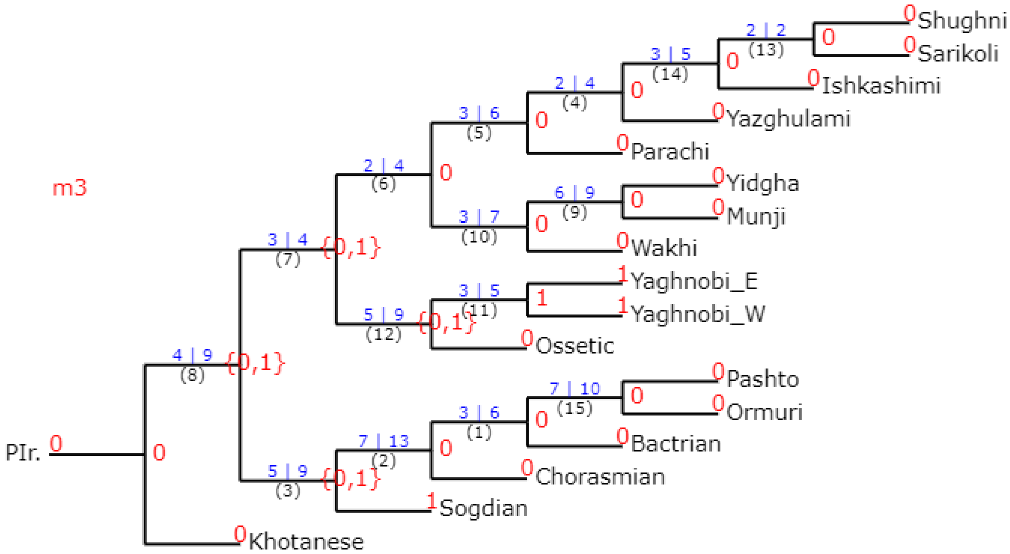
m1


m1
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

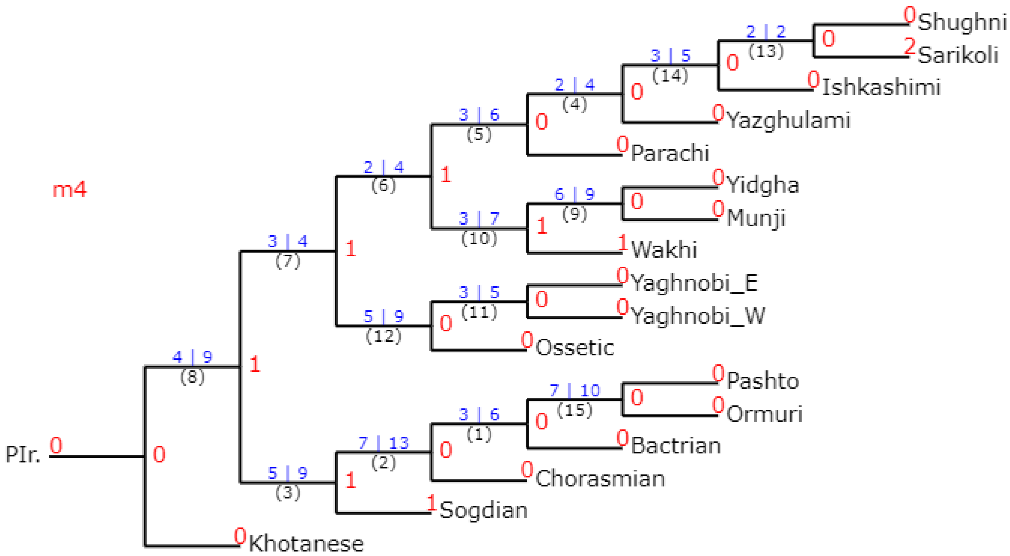
m2


m2
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

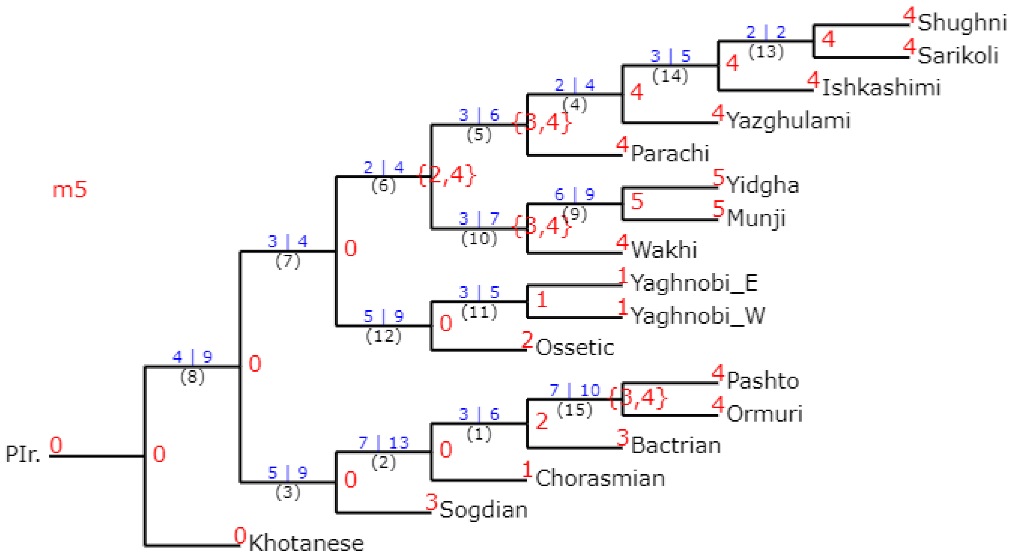
m3


m3
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

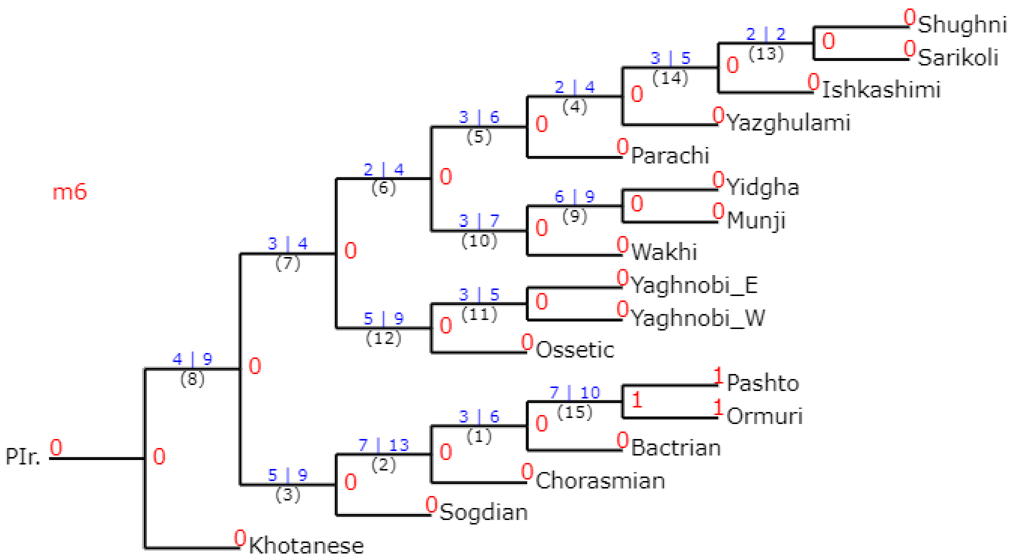
m4


m4
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

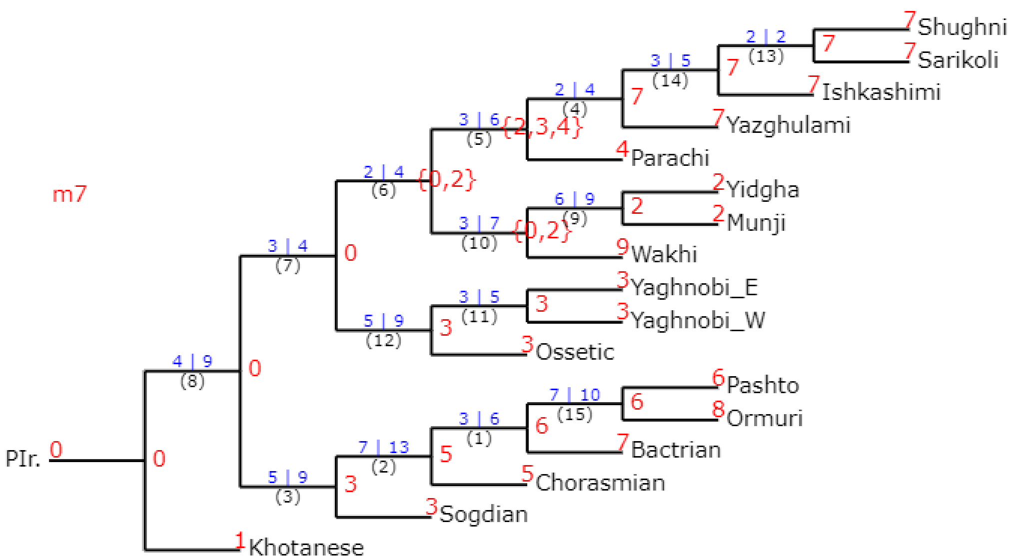
m5


m5
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

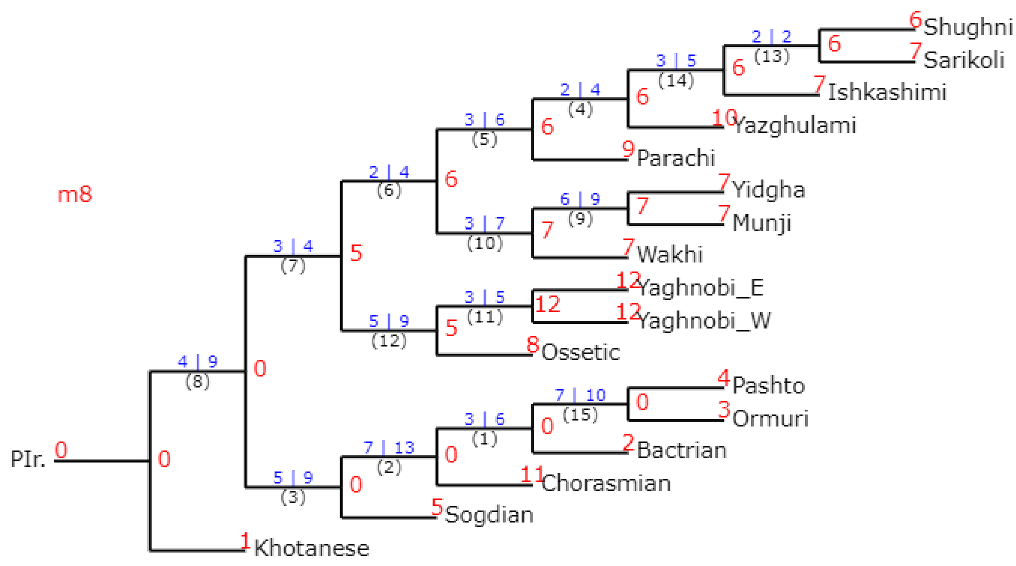
m6


m6
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

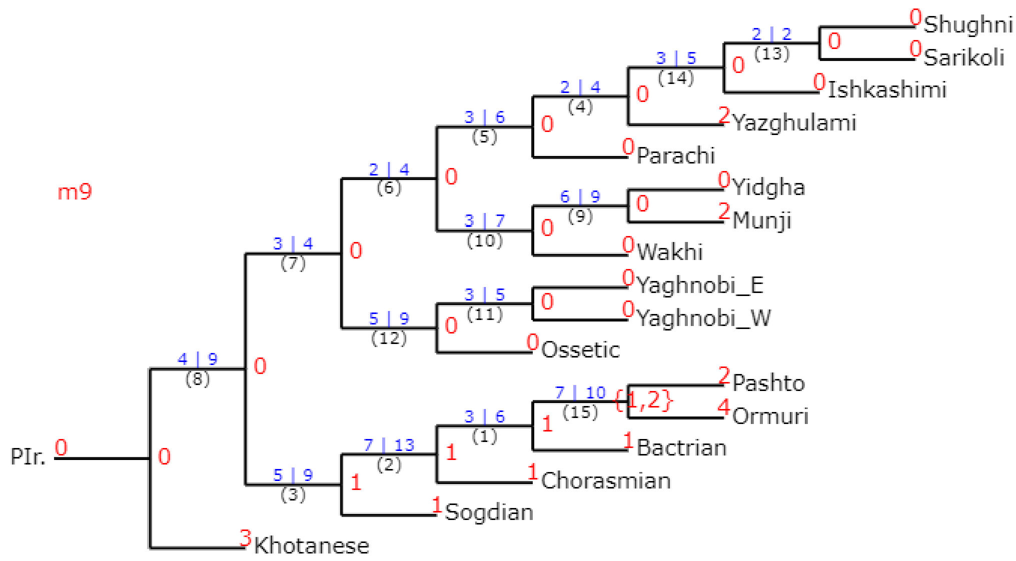
m7


m7
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

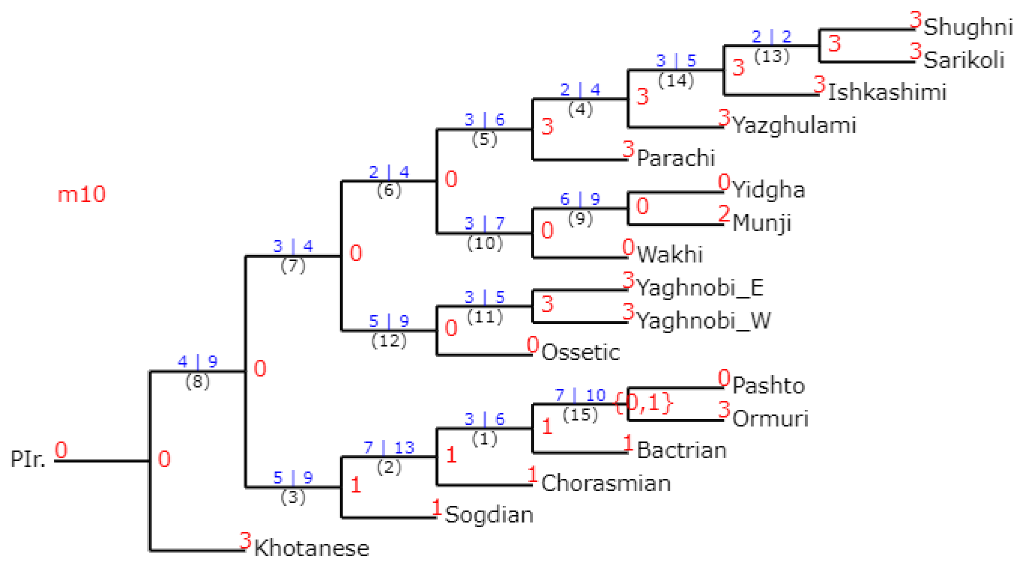
m8


m8
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

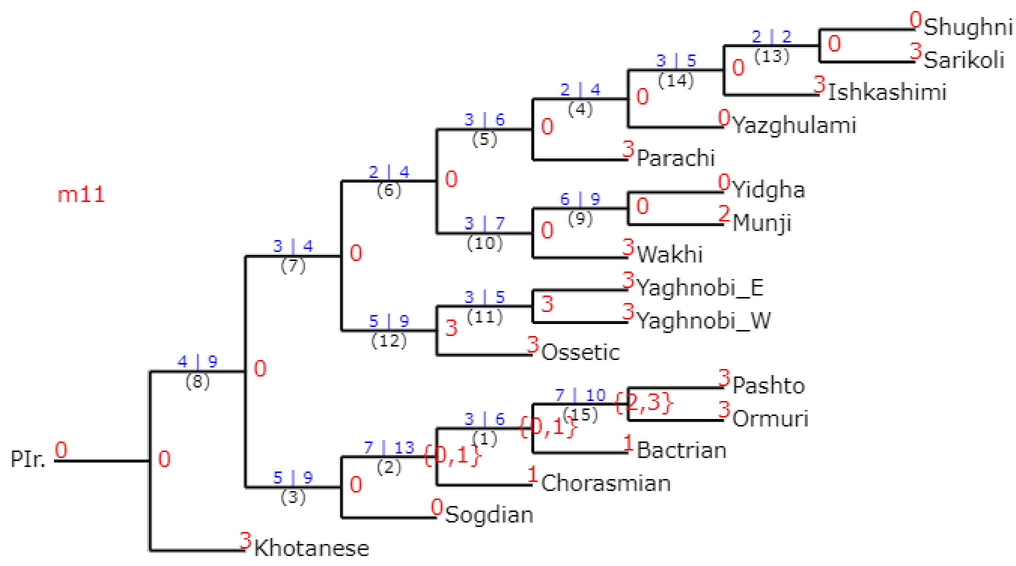
m9


m9
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

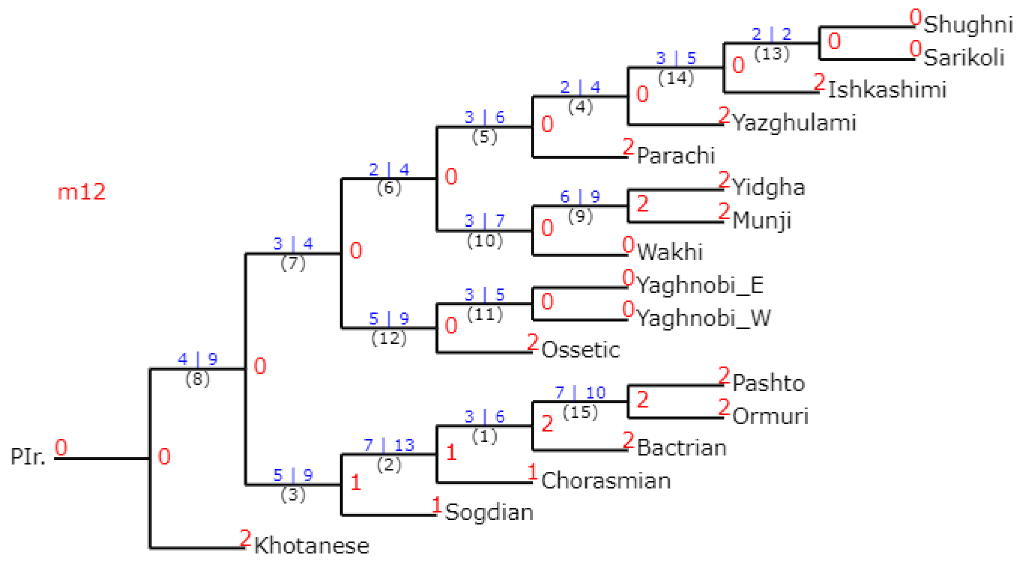
m10


m10
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

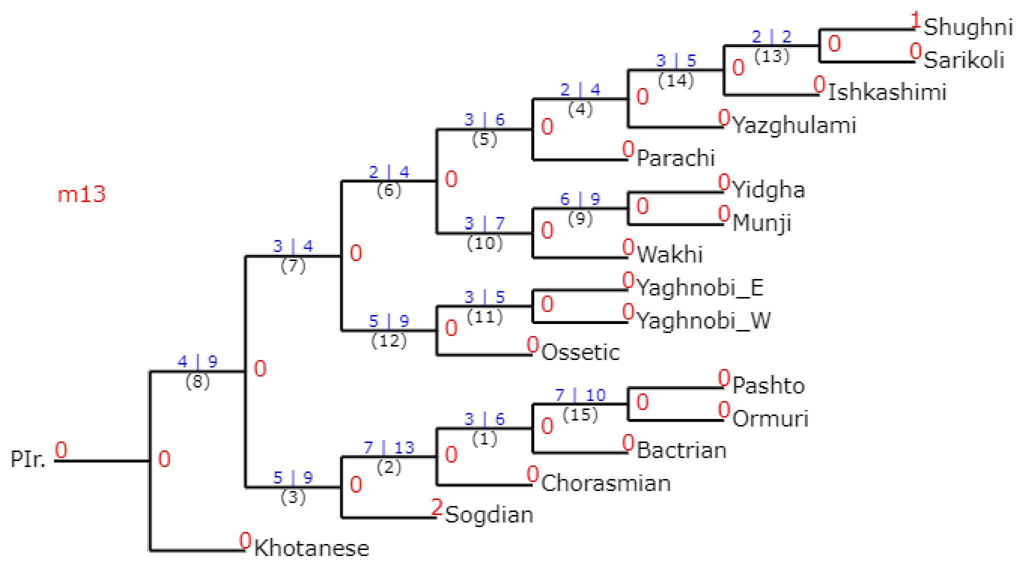
m11


m11
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

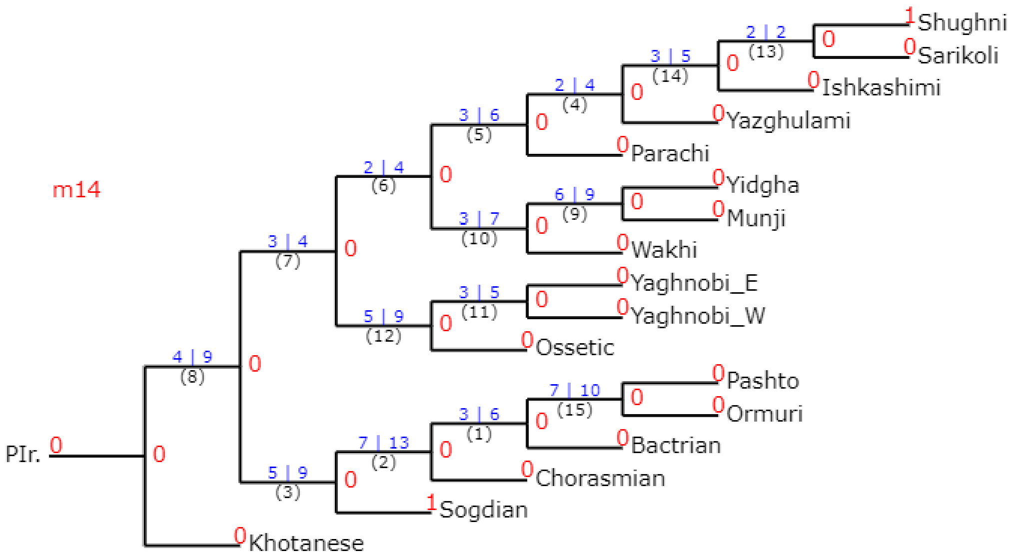
m12


m12
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

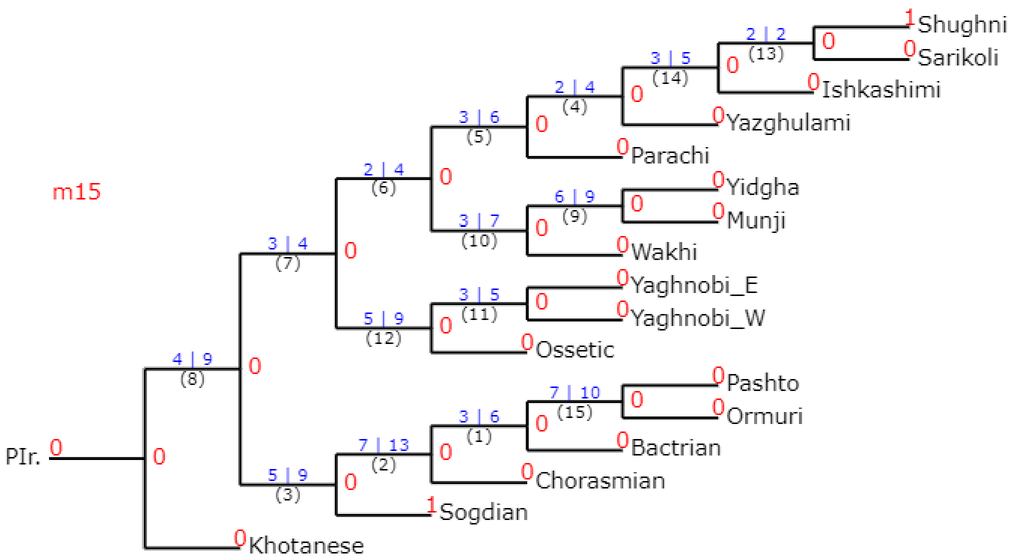
m13


m13
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

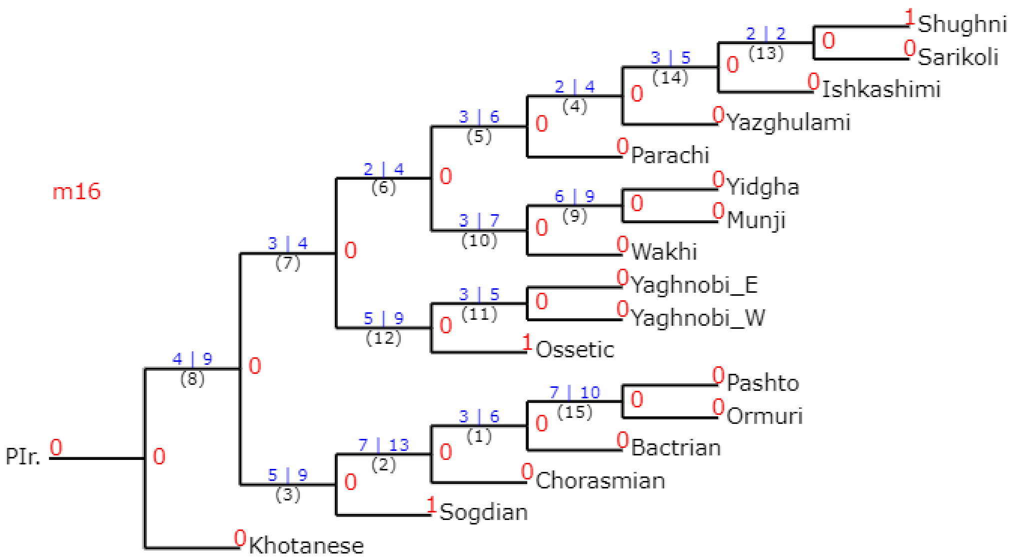
m14


m14
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

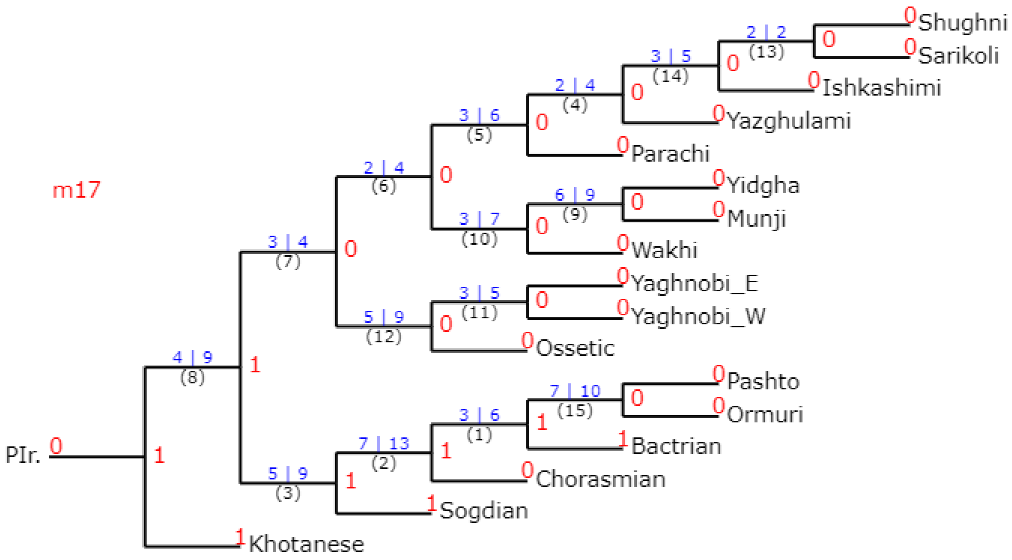
m15


m15
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

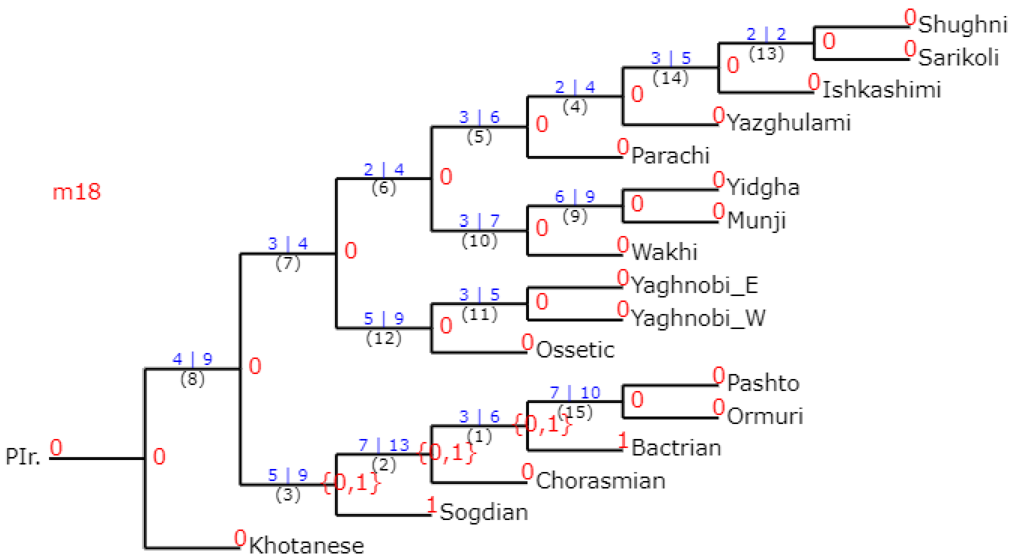
m16


m16
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

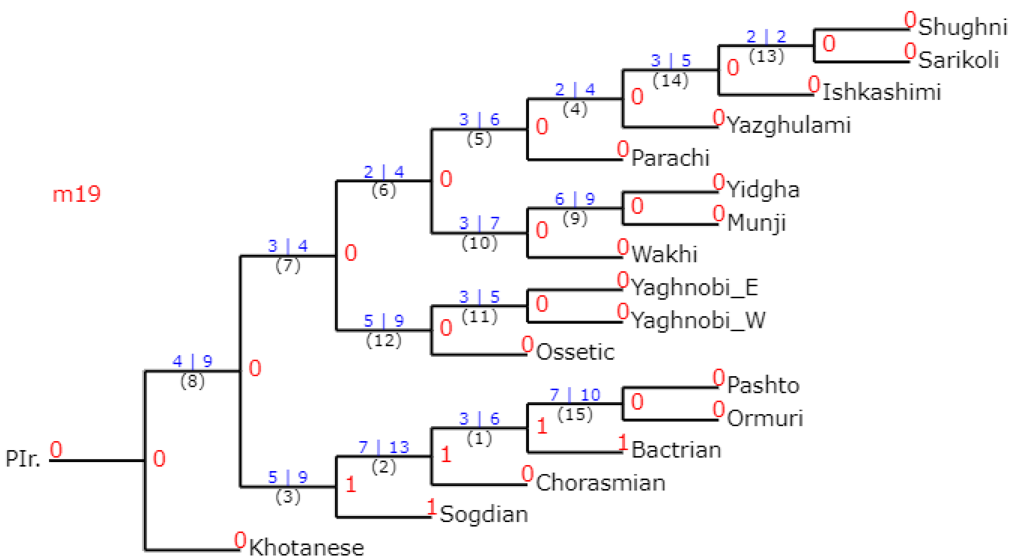
m17


m17
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

m18


m18
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

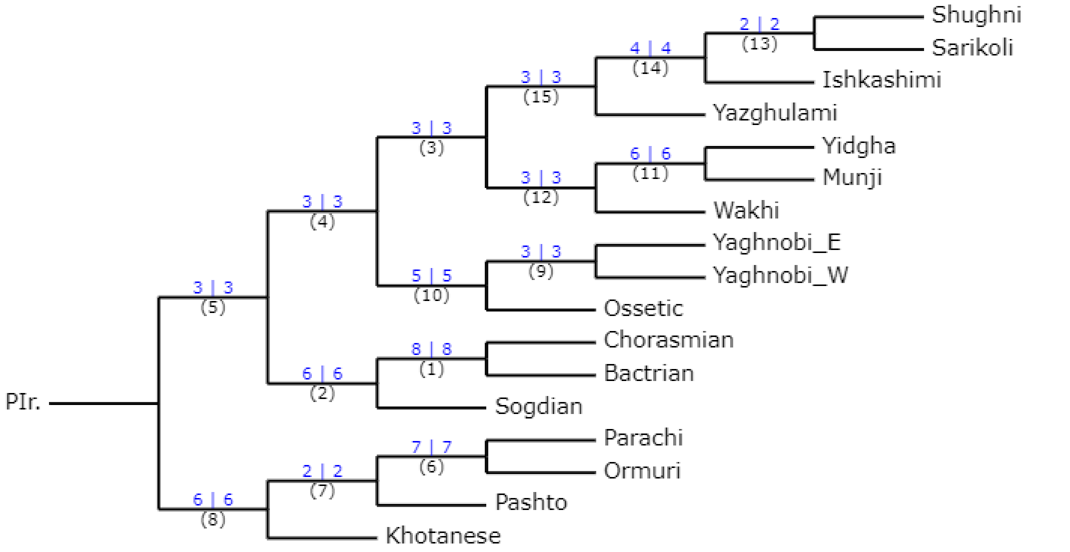
m19


m19
PIr.
Shughni
Sarikoli
Ishkashimi
Yazghulami
Parachi
Yidgha
Munji
Wakhi
Yaghnobi_E
Yaghnobi_W
Ossetic
Pashto
Ormuri
Bactrian
Chorasmian
Sogdian
Khotanese

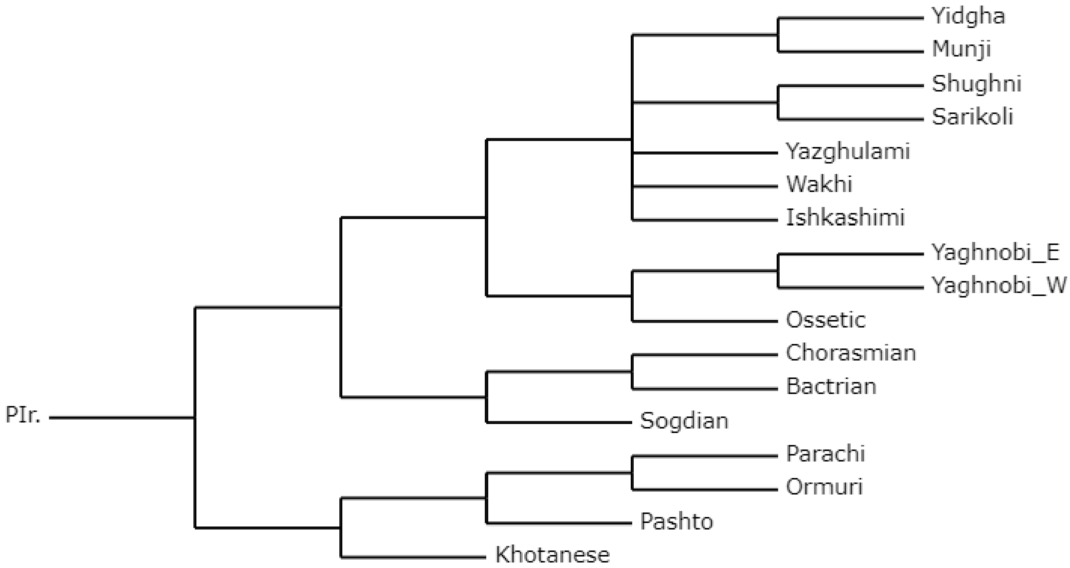
7.9. Appendix 9: Alternative analyses

No weights, but costs kepts (NWC)

"Tree 6" NWC best tree (score 404)

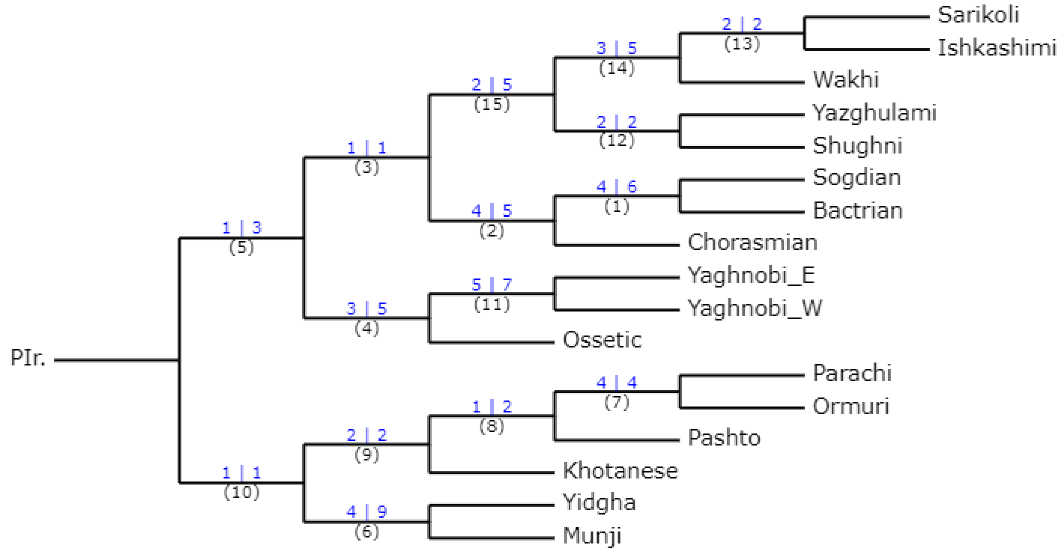


NWC Majority Consensus tree

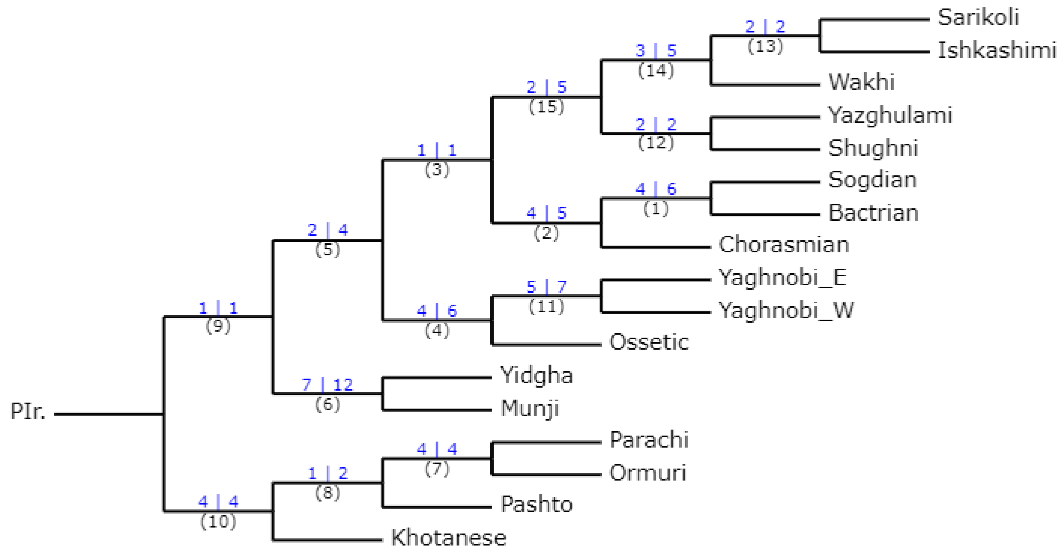


Weights kept, but costs removed (WNC)

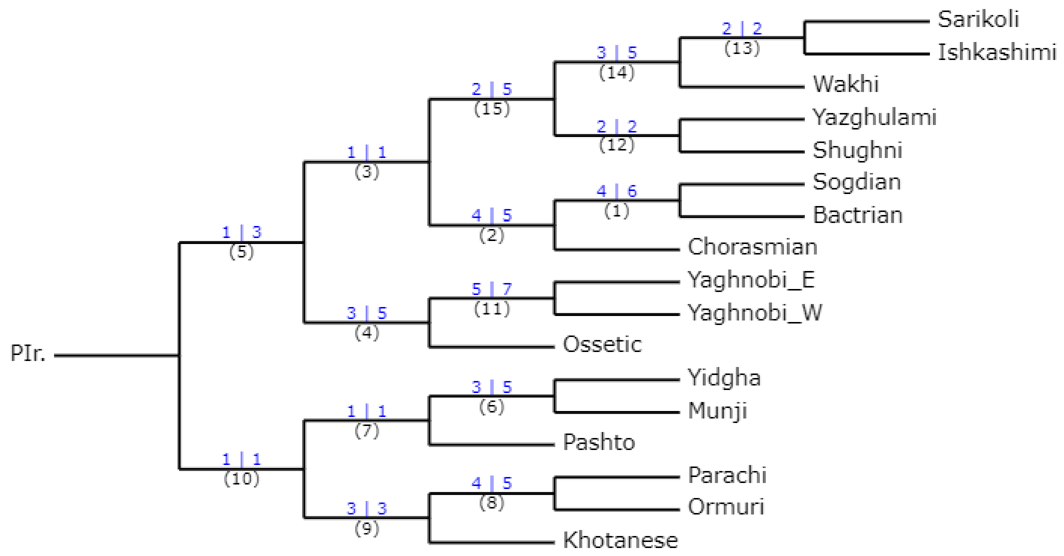
WNC best tree (a) (score 250)



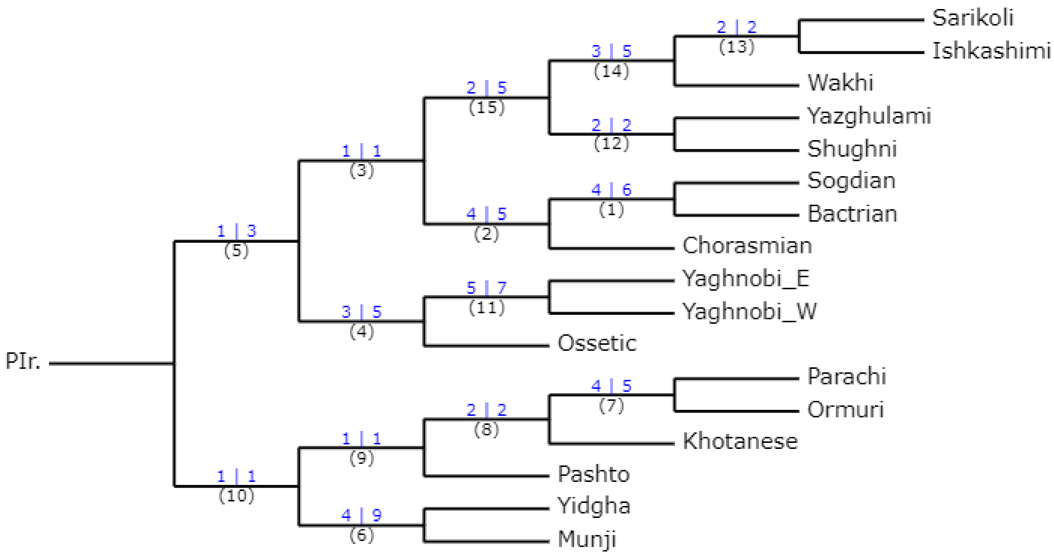
WNC best tree (b) (score 250)



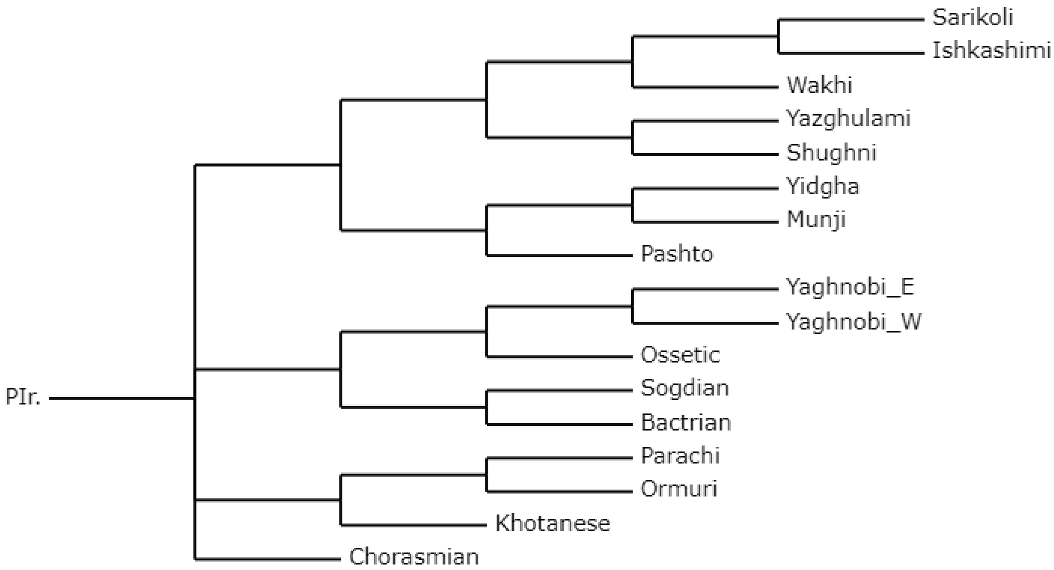
WNC best tree (c) (score 250)



WNC best tree (d) (score 250)

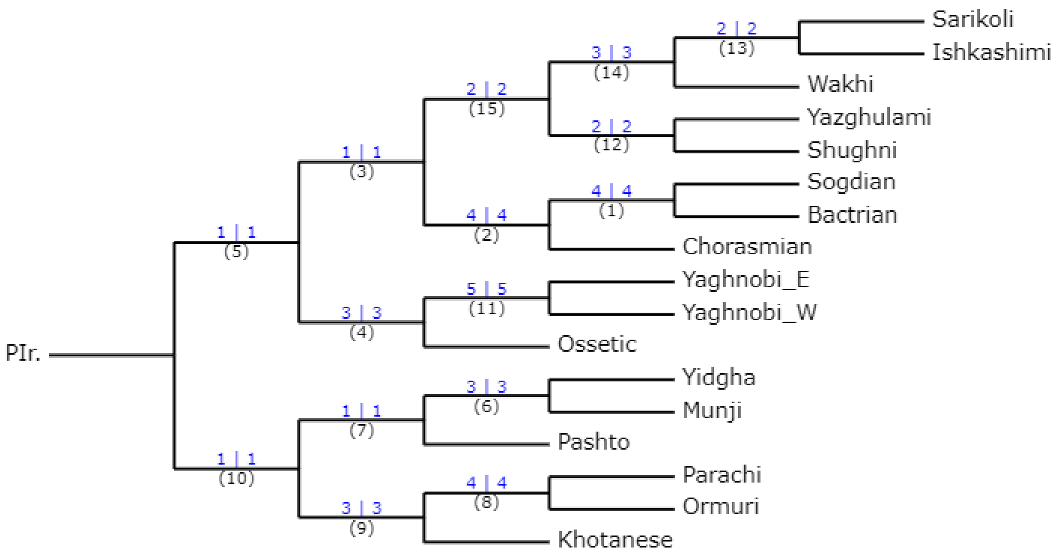


"Tree 8" WNC Majority Consensus tree

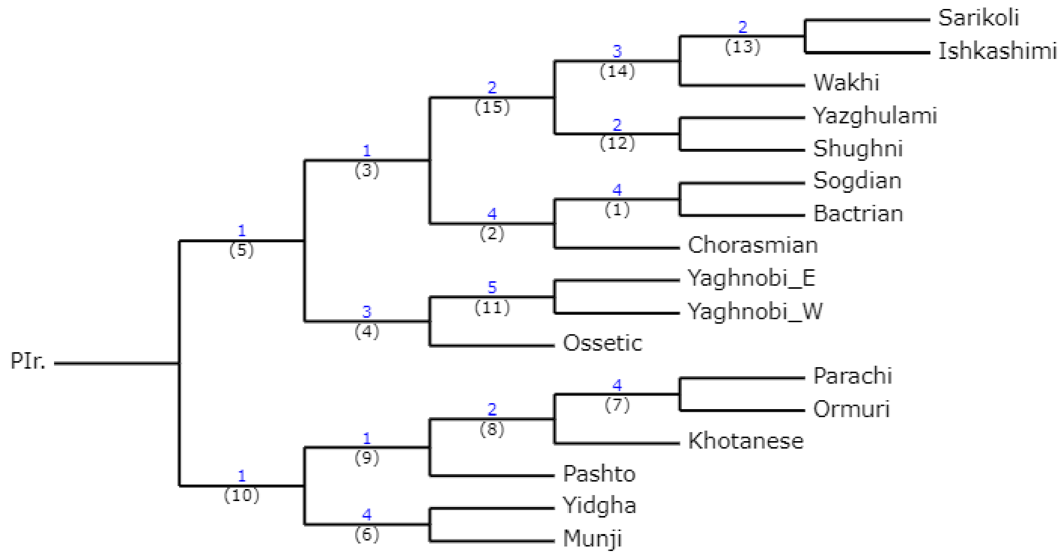


No weights, no costs

NWNC best tree (a) (score 159)



NWNC best tree (b) (score 159)



NWNC Majority Consensus tree