Reconstructing the evolution of Indo-European grammar

Gerd Carling[a] and Chundra Cathcart[b]

[a]Centre for Languages and Literature, Lund University
[b]Department of Comparative Language Science/ Center for the
Interdisciplinary Study of Language Evolution, University of Zurich

Abstract

Abstract:*

This study uses phylogenetic methods adopted from computational biology in order to reconstruct features of Proto-Indo-European morphosyntax. We estimate the probability of presence of typological features in Proto-Indo-European on the assumption that these features change according to a stochastic process governed by evolutionary transition rates between them. We compare these probabilities to previous reconstructions of Proto-Indo-European morphosyntax, which use either the comparative-historical method or implicational typology. We find that our reconstruction yields a strong support for a canonical model (synthetic, nominative-accusative, head-final) of the proto-language and a low support for any alternative model. Observing the evolutionary dynamics of features in our data set, we conclude that morphological features have slower rates of change, whereas syntactic traits change faster. Additionally, more frequent, unmarked traits in grammatical hierarchies have slower change rates when compared to less frequent, marked ones, which indicates that universal patterns of economy and frequency impact language change within the family.

Keywords: Indo-European linguistics, historical linguistics, phylogenetic linguistics, typology, syntactic reconstruction

---

1 INTRODUCTION

1.1 A CENTURY OF INDO-EUROPEAN SYNTACTIC RECONSTRUCTION

More than a century has passed since the pioneering work on syntactic reconstruction by the Neogrammarians (Brugmann & Delbrück 1897, 1893; Delbrück 1900; Wackernagel 1920), dealing with core issues in Indo-European grammar, such as case, word order, alignment, agreement, person agreement, position of the verb, and the behavior of clitics. The system reconstructed for Indo-European in these works was fundamentally based on a comparative-historical reconstruction of morphological and syntactic features of ancient Indo-European languages, with a strong focus on a systematic comparison of Old Indo-Aryan, in particular Vedic Sanskrit, with Latin and Greek. This model — which we label 'canonical' — used Sanskrit as a template for syntactic reconstruction. In Hirt's words: 'Delbrück's point of departure is Sanskrit. If something is not present in Sanskrit, it does not belong to Indo-European' (Hirt 1934:5, our translation). A new era of syntactic reconstruction, which is reflected in Hirt's skeptical view, began with the decipherment of Hittite in 1915 and the discovery of the Anatolian branch of Indo-European. Due to the old age of Anatolian sources, some linguists considered the complex synthetic structure of Old Indo-Aryan and Greek a secondary development, holding the view that Anatolian reflects a more archaic system. Accordingly, the discovery of Anatolian gave rise to alternative theories of Proto-Indo-European grammar, involving ergative (Uhlenbeck 1901; Vaillant 1936) or isolating (Hirt 1934) structure, and resulted in the concept of Indo-Hittite (Sturtevant 1962). Any grammatical reconstruction post-dating Indo-Hittite has to consider the role of Anatolian systems in relation to Proto-Indo-European. Currently, 'Greco-Aryan' and 'Anatolian' models serve as complementary to each other in Indo-European grammar, e.g. for the reconstruction of the verbal system (Clackson 2007:114-42).

Another important approach to syntactic reconstruction emerged during the 1970s, stemming from research on typological implicational universals (Greenberg 1963, 1978), which was adapted to a model for reconstructing syntax (Lehmann 1974). Even though this model met with skepticism from some comparative-historical scholars (Winter 1984), it had an important continuation in the typological diachronic approach of Nichols (1992, 1995, 1998). This approach has grown in importance in the era of computational typology (Bickel 2007; Wichmann 2014), giving birth to several alternative models for explaining typological change (Baker 2011; Croft et al. 2011; Dryer 2011; Dunn et al. 2011; Levy &

Daumé 2011; Longobardi & Roberts 2011; Plank 2011; Cathcart et al. 2018).

Since the pioneering work of Greenberg (1963), most syntactic reconstruction has been influenced by implicational typology. By merging typology and comparative-historical syntax, several important contributions to syntactic reconstruction have been published in recent decades, targeting the syntax of Indo-European as well as that of other families (Harris & Campbell 1995). This area of research goes under the name of diachronic typology (Viti 2015). An important approach continues the active-stative reconstruction model for Proto-Indo-European (Bauer 2000; Gamkrelidze & Ivanov 1984; Schmidt 1979). Other targeted domains have been the reconstruction of an active-stative verbal paradigm (Jasanoff 1978), the collective-count plural in the case system (Melchert 2000), dative-subject constructions (Barðdal & Eythórsson 2009), and various aspect of modality, tense, voice, aspect, particles, gender (Meier-Brügger et al. 2010:374-412). Along with these works, there are a number of excellent overviews on principles of syntactic reconstruction (Barðdal 2014; Ferraresi & Goldbach 2008; Roberts 2007), or monographs and handbooks compiling recent progress in various areas of syntactic reconstruction (Josephson & Söhrman 2008; Kulikov & Lavidas 2015; Ledgeway & Roberts 2017; Viti 2015). In §§3–5, where we evaluate the results of our reconstruction, we discuss this literature in further detail.

1.2 Outline of the current study

Our study analyzes comparative concepts of Indo-European morphosyntax, including the linguistic categories of ALIGNMENT, VERBAL MORPHOLOGY, NOMINAL MORPHOLOGY, TENSE TYPOLOGY, and WORD ORDER. We analyze data from 125 languages, including ancient, medieval and modern languages from the Indo-European family (Table 1). Our data set is extracted from the typological subsection of the database DiACL – Diachronic Atlas of Comparative Linguistics (Carling et al. 2018; Carling 2019), a collection of linguistic data from languages of Eurasia and other regions. The original data set of 108 binary features is re-coded to yield 65 categorical (i.e. non-binary) features.

We select a well-known and well-studied family with a long history of scholarship as the basis for our investigation. The aim of the study is twofold: first, we wish to assess the extent to which phylogenetic comparative methods, which can be used to estimate the probability of morphosyntactic features in Proto-Indo-European, agree with the results of previous models of syntactic reconstruction for the Indo-European family. Second, we aim to make inferences regarding the evolutionary dynamics and variability of different

morphosyntactic features during the course of the history of the Indo-European languages. In §2, we describe the model, method, and data forming the basis for the current study. In §3, we evaluate the results of the reconstruction for Proto-Indo-European and envisage further research that could emerge from the data and the model. In §3.5, we give the results of a statistical study, where we compare our reconstructed results to three different models to comparative-historical syntax. In §4, we discuss the evolutionary dynamics and variability of the transitions of traits. Finally, in §5, we discuss our results, both in the light of previous reconstructions of Indo-European grammar as well as from the perspective of general theories of grammar evolution. Technical description of the methods used in this paper can be found in appendices A–F located in the online supplementary material; sections S1–9 of the online supplementary material contain full details of the data employed and results. The raw data set is available open access via the DiACL database (`https://diacl.ht.lu.se/`). All code, metadata, and data are available at the following links: `https://github.com/chundrac/rec-evo-IE-gram`, `https://zenodo.org/record/4275010`.

[Table 1 about here.]

2   Theory, model, data, coding, method, and analysis

2.1   Comparative-historical, typological and phylogenetic models of reconstruction

The model of morphosyntactic reconstruction introduced by the scholars of Indo-European in the nineteenth century is based primarily on the comparative-historical method, systematizing forms and meanings of morphemes in a manner that sets of paradigms, rules and syntactic patterns can be reconstructed to a proto-language. Even though morphemes can be reconstructed as a result of the comparative-historical method, the reconstruction of their syntactic function is nontrivial, due to the uncertainty of regularity and the problem of establishing directionality of syntactic change (Barðdal 2014). Nevertheless, this method of reconstruction is utilized by a number of scholars, even though there is agreement that the method should not be applied to properties which are unconstrained by morphology, such as word order (Harris & Campbell 1995; Harris 2008). Proponents of the comparative-historical reconstruction model argue that if a specific pattern aided by morphological reconstruction has survived in a majority of languages, then there is reason to reconstruct it to the proto-language (Campbell & Harris 2002:615). Critics of this

model point to the directionality problem: we may reconstruct a pattern to an ancestral state of several daughter languages carrying the same pattern, but in case of a disagreement we do not know enough about the directionality of syntactic change to reconstruct one variant over another (Roberts 2007; Walkden 2013:363-67).

The model of reconstruction used by the typologists from the 1960s onwards is based upon a different principle: if language-internal implicational dependencies between typological features, so-called UNIVERSALS, can be identified, then these observations can be used as an argument for reconstructing typological properties to a proto-language. A major obstacle to the adaptation of this model is how to deal with language-internal conflicts between features with respect to assumed dependencies, both in attested as well as in reconstructed languages. An example is the controversy over Indo-European word order, where reconstruction based on ancient languages does not yield a uniform result with respect to the proto-language (Friedrich 1975; Lehmann 1974; Watkins 1976; Winter 1984).

In phylogenetic comparative methods, the issue of reconstruction is formulated in probabilistic terms, using phylogenetic computational algorithms originally adapted from biology (Calude & Verkerk 2016; Silva & Tehrani 2016; Jäger 2019). These models assume a specific stochastic process underlying CHARACTER EVOLUTION, which usually involves TRANSITION RATES which characterize change between values of a linguistic variable (e.g. different main clause word orders) over a phylogeny. These rates are estimated on the basis of a phylogenetic representation, often a TREE SAMPLE inferred from basic vocabulary patterns or a comparable linguistic feature, and the distribution of the feature among the daughter languages. These rates can be used to reconstruct the probability of a given value at internal nodes of the tree, including the root (i.e. the node ancestral to all others in the tree), as well as infer locations on branches of the tree where change is likely to have taken place (Maurits & Griffiths 2014; Dunn et al. 2017; Widmer et al. 2017; Cathcart et al. 2018; Blasi et al. 2019; Cathcart et al. 2020).

Figure 1 provides a schematic toy diagram of an ancestral state reconstruction problem in a phylogenetic comparative framework. Given a phylogeny with observed data at the tips of the tree, the procedure has two objectives: (1) to infer transition rates between feature values, and (2) to estimate values for unobserved internal nodes that are most likely to have preceded the values displayed by (or inferred for) their descendants. In a Parsimony framework (a model that minimizes the total number of character-state changes), this often involves restricting the number of parallel changes over the phylogeny.

In a likelihood-based framework, including its Bayesian extensions, this also involves inferring evolutionary rates that express the probability of changes between different states over various spans of time represented by the branch lengths of the phylogeny. In general, evolutionary rates are inferred while treating internal states as a nuisance factor, which are to be marginalized out for the sake of efficiency. The rates that are inferred, or their posterior distributions under the Bayesian approach, can then be used to estimate the probabilities of different states at different nodes in the tree, starting at the tips and moving toward the root (Felsenstein 2004; Yang 2014).

[Figure 1 about here.]

2.2 DATA: ORIGINAL DATA SET AND RECODING FOR THE CURRENT STUDY

We use a data set of Indo-European languages extracted from the DiACL Typology/Eurasia dataset (Carling et al. 2018). The data involve categories of grammar that have been under discussion in both diachronic syntax and general typology for a long time, and are coded according to a hierarchical model designed to represent morphosyntactic features with an adequate level of granularity. Like other similar databases, e.g. AUTOTYP (Bickel & Nichols 2002) or WALS (Dryer & Haspelmath 2013), the data set consists of comparative concepts (Haspelmath 2010), definitions of linguistic features of grammar designed for cross-linguistic comparison. The original binary hierarchical model of DiACL, related to multivariate approaches (Bickel 2007), organizes comparative concepts according to levels of increasing detail. We recoded the binary data so that the data set consists of categorical VARIABLES[1] (e.g. MAIN CLAUSE WORD ORDER) taking multiple values (e.g. SVO, SOV, VSO, V2), organized within larger morphosyntactic CATEGORIES (comprising ALIGNMENT, WORD ORDER, NOMINAL MORPHOLOGY, VERBAL MORPHOLOGY, and TENSE). This results in 65 categorical variables, 64 of which are 'informative' in that they show variation within Indo-European and are thus suitable for phylogenetic analysis (Figure 2). Figure 2a illustrates the hierarchical principle of organizing linguistic properties into Grids, Features, Variants and Values, which is used in the database DiACL. Figure 2b demonstrates how this hierarchy is mapped into categorical variables, which contain blocks of value combinations, defined as traits.

There are several advantages of using a hierarchical model for typological data, as in the original data set of our study DiACL (Carling et al. 2018) or AUTOTYP (Bickel & Nichols 2002). The most important advantage is the possibility of increasing detail, which

enables local adaptations as well as the possibility of testing grammatical relations. Another advantage is the ability to recode the binarized strings of 1 and 0 into new combinations which match a specific research question. A further advantage is the possibility of contrasting features across grammatical categories (cf. next section).

The procedure for transforming the hierarchically organized original data into our scheme of recoded categorical variables is shown in Figure 2, exemplified using alignment. The complete recoded data is given in S3a. The various coding and recoding strategies are described under the respective sections, where we discuss and evaluate results (§3).

[Figure 2 about here.]

[Figure 3 about here.]

2.3 ADDITIONAL DATA SETS

CODING OF RECONSTRUCTION MODELS   Our study compares a probabilistic model of reconstruction with insights from comparative-historical approaches to syntactic change. For this purpose, we have selected a number of well-known approaches to the reconstruction of Indo-European syntax against which to compare our results. There is a rich literature on the reconstruction of Indo-European syntax, the full treatment of which is outside of this paper's scope. For the sake of simplicity, we have limited our comparisons to representative publications that address all grammatical categories present in our data. We found three descriptions that were complete enough that we could treat the models proposed according to a coding scheme against which we could easily compare our results. These we label CANONICAL (Delbrück 1893, 1897, 1900), ISOLATING (Hirt 1934, 1937), and ACTIVE-STATIVE (Gamkrelidze & Ivanov 1984; Gamkrelidze et al. 1995). It is important to remember that alternative theories (isolating and active-stative) reconstruct a stratified Proto-Indo-European language. At the root, they reconstruct a joint Anatolian and Non-Anatolian stage (Indo-Anatolian), which later transforms into a stage which represents the predecessor of the Non-Anatolian languages. Substantial portions of the discussion within alternative theories deal with the process of system change from Indo-Anatolian to Non-Anatolian branches (Pooth et al. 2018). In order to ensure comparability of our results and previous theories of grammar reconstruction, we use an Indo-Anatolian reference phylogeny, which represents the consensus view on branching and time-depth of the Indo-European family. We take the root of our phylogeny, which serves as the joint ancestral stage of the Anatolian and Non-Anatolian

sub-branches of the family, to represent the unattested Proto-Indo-European (PIE) language (see further §§3–5 and S9).

The coding of feature variants of models, including source references, of the raw DiACL data is found in S2b; recoding of feature variants into our categorical features is found in S3a.

Coding of grammatical hierarchies in the data   Additionally, we implement a coding of grammatical hierarchies between features in our data (S7). The issue of grammatical hierarchies is of key importance to the implicational typology model of Greenberg (1966), Croft (1990, 2003), and Comrie (1981), and is implicitly connected to the frequency of grammatical categories as well as the markedness theory (Haspelmath 2006). During the course of our analyses, we found that our model displayed asymmetric results for different features, not just between basic categories (e.g. word order, nominal morphology, verbal morphology, alignment, tense), but also within categories, between features differing with respect to categories such as tense (present, past) or word class (noun, pronoun) (S2, S3). For this reason, we chose to adopt an additional model of coding in which we identify pairs of features that belong to the same grammatical category but vary with respect to other grammatical categories, which can be defined as in a grammatical hierarchical relation to each other. For the sake of simplicity and comparability, we reduce our grammatical hierarchies to pairs of features, which have been observed in previous literature (where they are often referred to as scales). There is a rich literature on grammatical as well as marking hierarchies in grammar, both from the perspective of individual languages as well as cross-linguistically (Bornkessel-Schlesewsky et al. 2015; Comrie 1981; Croft 2003; Haspelmath 2015; Mal ukov 2015). Generally, grammatical hierarchies are based on three different criteria (Croft 1990:92; Croft 2003:156-57):

1. Structural criteria, i.e. marking in grammars,

2. Behavioral criteria, i.e. the inflectional and distributional patterns in languages, and

3. Frequency, i.e. the occurrence in text, both in individual languages and cross-linguistically.

Only a handful of the grammatical hierarchies mentioned in the literature recur in our data, and there is also disagreement about the hierarchical organization of some of the categories in our data. One such example is the relation between future and present.

Whereas the original hierarchy of Greenberg (Greenberg & Haspelmath 2005; Greenberg 1966) and Croft (Croft 1990:92-93) puts these traits in the order PRESENT < FUTURE, other scholars (Mal ukov 2015; Witzlack-Makarevich & Seržant 2018) place these properties in the order FUTURE < PRESENT < PAST on the basis of existing marking patterns in some languages. The issue is complex: we are aware that many languages reverse general hierarchies in their grammatical systems (Bickel 2008; see also Garrett 2008).

For this purpose, we use general grammatical hierarchies (Aissen 2003; Haspelmath 2015) as our point of reference, establishing pairwise hierarchical relations which we then implement for selected features in our data set (Table 2). The reason why we use pairwise relations and not hierarchical scales (e.g. SINGULAR < PLURAL < DUAL) is that we intend to compare grammatical hierarchies and the transition rates inferred by our model in a systematic fashion. Since our data contain features which are defined according to several categories, features may recur in hierarchical pairs. As an example, the features PRONOUN, PRESENT PROGRESSIVE: NOMINATIVE-ACCUSATIVE and NOUN, PRESENT PROGRESSIVE: NOMINATIVE-ACCUSATIVE are in a hierarchical relationship (PRONOUN < NOUN), whereas the features PRONOUN, PRESENT PROGRESSIVE: NOMINATIVE-ACCUSATIVE and PRONOUN, SIMPLE PAST: NOMINATIVE-ACCUSATIVE are also in a hierarchical relationship (PRESENT < PAST). A number of features in our data are not involved in any grammatical hierarchy relation, for a number of reasons. One reason is that they lack a hierarchical grammatical relationship to any other feature in the data. We also chose to consistently mark negative values in a fashion similar to their positive counterparts; for example, NO SYNTHETIC PRESENT PROGRESSIVE and NO SYNTHETIC FUTURE are in a hierarchical relation PRESENT < FUTURE, just as SYNTHETIC PRESENT PROGRESSIVE and SYNTHETIC FUTURE.

We chose a priori not to code any hierarchies for word order. Even though it is evident that head-final traits (OV, RELATIVE-NOUN, POSSESSOR-POSSESSED, etc.) have lower rates of change (S5), we prefer not to enter into a discussion about possible marking hierarchies or general frequencies in word order (Croft 1990:84-91).

[Table 2 about here.]

2.4 Methodology: reconstruction with phylogenetic comparative methods

The methodology on which this paper relies assumes that linguistic variables evolve under a CONTINUOUS-TIME MARKOV PROCESS (for an introduction see Liggett 2010), a stochastic model which assumes that there exist rates of change between values of categorical variables which characterize their evolution over time. Accordingly, our model infers rates of change between values of the categorical variables in our data set, using a tree sample representing genetic relationships between languages of the Indo-European family. Once these TRANSITION RATES have been inferred, they can be used to estimate the probability of a value for a given variable at phylogenetic nodes where data are unobserved; these INTERNAL NODES correspond to reconstructible protolanguages, with the ROOT of the tree corresponding to Proto-Indo-European.

Concrete details regarding the generation of the tree sample and the inference procedure can be found in the Appendix. Our tree sample (S9) is generated as follows: we assume a fixed topology that agrees with received philological wisdom, and sample branch lengths from chronologically realistic intervals, yielding a tree with a root age uniformly distributed between 7000 and 6000 years BP. The model is Bayesian; we infer posterior distributions for transition rates, using Felsenstein's PRUNING ALGORITHM (Felsenstein 1981, 2004) to compute the likelihood of these parameters for trees in the tree sample. We estimate the probability of a value for a given variable at the root of the phylogeny (i.e. for Proto-Indo-European) by randomly drawing evolutionary rates from their respective posterior samples, iteratively sampling a value at the root (Nielsen 2002; Huelsenbeck et al. 2003; Bollback 2006), and normalizing the counts for each sampled state to yield probabilities between 0 and 1.[2] In §3, we evaluate these results.

In most cases, there is a clear result in which our procedure reconstructs a feature with relative certainty, inferring a high probability for a specific value of a variable and a low probability for the remaining values. This behavior can be seen in a histogram of all reconstruction probabilities (Figure 4); the distribution of these probabilities is U-shaped, in that low (0.0–0.25) and high (0.75–1) probability ranges are more frequent than the intermediate ones (0.25–0.75). While a small number of features are reconstructed with high uncertainty, meaning that we cannot say anything concrete about the value most likely to be present in Proto-Indo-European, the evolutionary dynamics of such features are still of interest, since these features may emerge in later phases of the Indo-European history. Furthermore, the behavior of such variables helps us diagnose the overall behavior

of our model, with an eye to why it reconstructs certain patterns with high certainty. We discuss our model's results in detail below.

[Figure 4 about here.]

3 RESULTS: RECONSTRUCTION

Using the methodology outlined in the previous section, we reconstructed probability distributions across values for each variable in our data set at the root of the phylogeny. These distributions represent probabilities that particular features were present in Proto-Indo-European, under our model. A complete listing of all results, along with figures providing visualizations of the evolutionary history of all variables in our data set, are found in the supplementary material (S4, S8). In the following sections, we provide a detailed assessment of our results, organized thematically according to different domains of morphosyntax that have been discussed at length in the traditional literature on syntactic reconstruction in Indo-European, namely alignment, definiteness, gender, case, verbal typology, verbal morphology, and word order. Finally, we provide a quantitative comparison of our results against received wisdom in the form of models of reconstruction proposed by the different schools or camps of traditional Indo-European syntactic reconstruction described above, which we term the CANONICAL, ACTIVE-STATIVE, and ISOLATING models.

3.1 ALIGNMENT

For variables pertaining to alignment (S4, A1-30), our results support the reconstruction of nominative-accusative alignment in multiple systems (Figure 5). Nominative-accusative alignment is found with nouns as first argument in the present progressive, with nouns as first argument in the simple past, with pronouns as first argument in the present progressive, with pronouns in the simple past, with verbal marking in the present progressive, and in the simple past. At the same time, while nominative-accusative alignment is reconstructed with a higher probability than other alignment types across these systems, the certainty with which it is reconstructed varies. We note that nominative-accusative is more likely in the present progressive than in the simple past, both for nouns and pronouns, and more likely with pronouns as first argument than with nouns. The second most frequent type of alignment is no marking, followed by ergative (both with low probabilities).

This discrepancy is striking. Considering language-internal distributions of the clause and argument types involved, it is clear that nominative-accusative features are reconstructed with higher certainty for grammatical categories of higher cross-linguistic frequency (present, pronoun) as opposed to the more infrequent categories (past, noun). This result is of relevance to discussions of grammatical hierarchies (Croft 2003; Haspelmath 2006) (see §4). Second, we notice that the ergative appears (at a low probability) only in the simple past. This result is reminiscent of results in the domain of verbal morphology, in which the simple past shows different patterns of change from the present progressive.

The reconstruction of patterns of alignment has a long history of discussion in comparative-historical syntax. In the canonical model of Delbrück and Brugmann (Delbrück 1893, 1897, 1900), the nominative codes the first argument (S/A), independent of transitivity of the predicate, and the accusative codes the second argument (O) (Meier-Brügger et al. 2010:401-404). However, due to the reconstruction of a case marking of *-s* for AGENT and *-m* for PATIENT, ergative alignment was proposed for Proto-Indo-European at an early date (Uhlenbeck 1901). This theory was later continued by Vaillant (1936) and Soviet scholars of the 1970s (Gamkrelidze & Ivanov 1984; Klimov 1974), who reconstruct an active-stative system, based on the *-os/-*om* distinctions in nominative/accusative and a corresponding *-os/*-om* distinction between genitives of active and inactive noun classes (Gamkrelidze et al. 1995: 233-76). Several scholars continue the active-stative theory (Bauer 2000; Schmidt 1979), reconstructing the relative chronology of the Indo-European paradigm, as well as reconstructing a continuation of change from an active-stative proto-language and into sub-branches, e.g. Italic (Bauer 2000). The source of the active-stative theories is a fundamental marking distinction between animate and inanimate (active-stative), reconstructed from the marking of the cases of nouns and pronouns in Proto-Indo-European (Table 3). The distinction is also reflected in suppletion in the pronominal paradigm (Table 4). The subject case has an unmarked zero-ending, against which the object is marked (Bauer 2000; Martinet 1962:44-46).

[Table 3 about here.]

[Table 4 about here.]

Under the active-stative theory, the active alignment is also marked in the two series of verbal endings, the *-mi* (active), and *-h₂e* (inactive) conjugation, supported by the *-mi*

and -*ḫi* paradigm in Anatolian (Gamkrelidze et al. 1995:254-76). However, as pointed out by other scholars, the formal contrast in Hittite between -*mi* and -*ḫi* conjugation is not reflected in any systematic difference in meaning (Jasanoff 2003:1-40), which is a prerequisite for the active-stative theory. At the same time, active-stative interpretations remain important in many theories of explanation of the Indo-European sets of endings (Jasanoff 1978).

The active-stative theory has no support under our reconstruction, pointing in the direction of nominative-accusative prevalence in Proto-Indo-European, both in the case marking on nouns and pronouns, in verbal conjugation, as well as in the present/past distinction (Figure 5). The active-stative and ergative theories are not generally supported by all Indo-European scholars (Meier-Brügger et al. 2010:412). However, they remain of great interest to us, since they connect to the reconstruction of the Indo-European gender and case systems, which yields interesting results on the basis of our data, consistent with the reconstruction of nominative-accusative alignment.

[Figure 5 about here.]

3.2 Nominal morphology

Case marking in the NP   Our results from the domain of nominal morphology provide information regarding the position of case marking within the noun phrase in Proto-Indo-European. In the data set, we code whether languages mark case on adjectives, articles, the first element of the NP, and the head noun (S4, NM1-8). Our reconstruction provides support for the presence of case marking on head nouns (0.745) and adjectives (0.559), but not on the article, in line with the probable absence of definite articles in Proto-Indo-European. Our system does not provide support for the presence of a rule that case must be marked on the last member of an NP (0.076). The case marking on the noun is not especially controversial: as long as we reconstruct a synthetic case system of canonical type (see below, Case), we also expect case marking to appear on the nominal head in a noun phrase. However, the relatively lower degree of probability of case marking on the adjective (0.559) is not completely in line with the canonical model, which also reconstructs full case marking, with respect to case and gender, on adjectives (note the higher gender agreement value, see below; Delbrück 1893:402ff..

Definiteness   On the whole, features pertaining to definiteness are reconstructed with low probabilities, indicating that the presence of definiteness in Proto-Indo-European is

unlikely. This is the case for definiteness marked on the adjective, definiteness on the first element of the NP, definiteness on the last element of the NP, definite article, and definiteness suffix (S4, NM9-17). This result is uncontroversial with respect to all models, since it is evident from the historical record that most Indo-European branches developed definiteness marking independently, by means of grammaticalization (Bauer 2007).

GENDER   Features targeting noun class and nominal gender (S4, NM18-27) display particularly interesting results. The probabilities of the presence of more than five noun classes (genders) and an animate gender are close to zero. However, the probability for having a masculine/feminine distinction is higher (0.684) than the probability of *not* having a masculine/feminine distinction (0.316). The probability for a special neuter gender is high (0.855). Furthermore, the probability for a predicative adjective to agree with its nominal head in gender is reasonably high (0.673, Figure 6).

These results for gender are noteworthy and somewhat controversial. At an early date, Delbrück (1893:132-133) is hesitant in reconstructing a Proto-Indo-European three-gender system, equivalent to the system found in archaic Indo-European languages, such as Sanskrit or Classical Greek. Considering the formal distribution of endings and the gender syncretism found in later Indo-European branches, he proposes that the three-gender system of Indo-European emerged from a two-gender system, based on an animacy-inanimacy distinction. Hirt (1934:28) reconstructs an Indo-European proto-language with no gender marking at all on nouns. In later literature, there is consensus around an original two-gender model of Proto-Indo-European, where the feminine is secondary (Szemerényi 1989:164-65; Tichy 1993; Gamkrelidze et al. 1995:242-24; Matasović 2004; Luraghi 2011). The issue of gender/noun class is critical to arguments for reconstructing active-stative or ergative systems for Indo-European, and the animacy vs. inanimacy distinction is interpreted as an active vs. inactive, subject vs. non-subject distinction (Meier-Brügger et al. 2010:412). There are discussions of how a three-gender system emerged out of a two-gender system, i.e. how the animacy category split up into a sexus distinction, the possible distinction concrete–abstract and non-collective–collective, and the formation of a feminine gender in *-h_2, originally an abstract suffix, which was extended to the collective (Luraghi 2011; Matasović 2004; Tichy 1993). Although we reconstruct a masculine/feminine distinction with only moderately high probability, this result goes against the mainstream model in that it reconstructs a three-gender system for Proto-Indo-European.

There are several possible reasons for this result. By using comparative concepts, i.e., features with no particular connection to individual pieces of morphological matter, the coding makes no difference between the two-gender system of Hittite (which is assumed to be preserved from Proto-Indo-European) and the two-gender system of, e.g. Dutch or Swedish (which collapsed from a previous three-gender system). Our model assumes that linguistic features evolve under a continuous-time Markov process, which estimates transition rates between values of a linguistic variable over time. The three-gender system is preserved and stable in many branches of Indo-European, as well as occasionally collapsed in some of the branches (e.g. Romance, Germanic), but not in a consistent way (masculine/feminine vs. common/neuter). Accordingly, the model estimates that it is more likely for Anatolian to have collapsed a Proto-Indo-European three-gender system than to have preserved an ancient two-gender system (see Figures 7–8, which display the most probable trajectories of historical development of these features under our model on a maximum clade credibility [MCC] summary tree, and further discussion in section 5).

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

CASE   Results for features pertaining to case show a degree of agreement with the canonical system of reconstruction similar to that of the features discussed in the foregoing sections. In Indo-European studies, the topics of morphosyntactic reconstruction of nominal morphology, the case system and its functionality, and paths of syncretism in various Indo-European sub-branches have been the subject of much debate, a full discussion of which is outside of this paper's scope. Instead, we use our results as a point of departure in an attempt to assess the extent to which they dovetail with previous theories regarding nominal morphology in comparative-historical syntax.

As far as typological structure is concerned, we code languages for the presence of agglutination for number and for case in nouns and pronouns (S4, NM30-33). All agglutinating traits have low reconstruction probability, somewhat higher for nominal morphology, but close to zero for pronominal morphology (Figure 10). Consequently, the

model reconstructs absence of agglutination for Proto-Indo-European, which is more evident for pronouns than for nouns.

In comparative-historical syntax, the discussion of the typological structure of the Indo-European case paradigm relates to the discussions of alignment, described in the previous section. Agglutination is not a key issue in the canonical model, which bases its reconstruction on the synthetic Old Indo-Aryan paradigm. A variant of the active-stative theory by, e.g. Gamkrelidze et al. (1995), is found in Hirt (1934), who reconstructs an uninflected stage of Proto-Indo-European. This stage is preserved in the neuter, which has no marking. In a later stage, the distinction between *-s* and *-m* marks 'grammatical cases', i.e. alignment cases (Table 3). The genitive also represents a variant of the *-s* and *-m* forms. All other cases, 'local cases', are secondarily formed by means of postposing elements. Even though it is not prominent in Hirt's text, it is evident that Hirt presupposed an agglutinating stage of Proto-Indo-European (between an assumed isolating and a synthetic stage), at least for the case paradigm. The proposed pathway from an agglutinating stage to a synthetic stage is reminiscent of the theory of Bopp (1816) regarding the origin of the Proto-Indo-European verbal endings.

The issue connects to the number and type of Proto-Indo-European cases. There is much discussion on this topic, in particular whether Indo-European had a rich case system and was secondarily syncretic, or whether the high number of cases in Old Indo-Aryan and other ancient languages represents an innovation. An argument is that most Indo-European languages show a decay and an increase in case syncretism rather than a growth in case morphology (Szemerényi 1996:158). An exception is Tocharian, which can partly be explained by the geographic position of Tocharian, surrounded by other agglutinating languages (Schmidt 1982). The evolution of Tocharian is also paralleled in Modern Indo-Aryan languages (Carling 2012). Delbrück (1893:180-91) reconstructs a case system which is identical to Sanskrit, with a nominative/vocative, accusative, genitive, ablative, locative, instrumental, dative and ablative. This is also the paradigm that scholars following the canonical model reconstruct (Meier-Brügger et al. 2010:398-410).

As for number and types of cases (Figure 10), our reconstruction is in line with the canonical model, albeit with a degree of uncertainty. We reconstruct as system with fewer than seven cases in the nominal paradigm (0.584), but with intermediate probability. Beyond that, we find, in the nominal paradigm, an intermediate probability for a dative and a genitive (0.608) but also a low score of (0.184) for having neither a genitive nor a dative. The score for having local cases outside of the core (in our coding the core includes

cases for A, S, O, dative, and genitive) is high (0.983); the score for an accusative/objective, i.e. a case for O different from A is also high (0.715), just as the score for a vocative case (0.885).

The pronominal paradigm shows similar tendencies, with some important exceptions: the probability for more than seven cases is close to zero (0.031), the probability for an Agent-Object distinction higher (0.934), the probability for a dative moderate (0.429), and for non-core local cases high (0.922). In addition, the probability of pronominal vocatives is low. As with alignment, we notice that the results of the pronominal paradigm are more distinct, i.e. the difference between the preferred versus non-preferred variant is larger.

In sum, our reconstruction yields medium to high probabilities for a canonical system, with a nominative, accusative, genitive, dative, vocative and one or several local cases, which do not exceed seven in number. The pronominal paradigm is similar, but with the exception that the inference procedure reconstructs a very low probability for a system of more than seven cases, as well as for the vocative.

[Figure 10 about here.]

3.3 Verbal morphology and tense

Verbal morphology The category verbal morphology targets agreement or person concord, i.e. the inflectional morphology of verbs with respect to their syntactic environment (Bickel 2007:169-171). A basic matrix (cf. Baerman & Brown 2013) including the variants full agreement (i.e. with reference to person and number) gender agreement (with respect to gender), and no agreement are matched against the core constituents S/A, O, and the case of the Recipient (dative). The coding captures the typological variation in syncretism between full and no agreement (see Table 5). Only the A agreement has results that are of interest to us here; the probability for dative and O agreement, which is found in branches of Indo-European, is low for the proto-language (Figure 11).

[Table 5 about here.]

Our results (Figure 11) display two tendencies of interest to us: the higher probability of full A agreement (0.657) against syncretic A agreement (0.212) in the present progressive, and the higher probability of syncretic A agreement (0.207) against full A agreement (0.031) in simple past. Again, we see a pattern in which the more frequent

category (present) is reconstructed with higher certainty than the less frequent category (past).

Our data set lacks a more fine-grained distinction between the various categories (e.g. voice, aspect, modality) than are present in our reconstructed Indo-European system and is therefore not fully comparable with the set of endings reconstructed for Indo-European via the comparative method. Much of the system complexity of ancient Indo-European languages is also lost in several branches of the modern languages (Clackson 2007:114-156), a transformation over time that our data only reflect to a certain degree.

[Figure 11 about here.]

Typological marking of tense Our data set's category tense takes as its focus the typological marking of present progressive and future, which is linguistically relevant from an areal perspective. In the Indo-European family, tense is historically integrated with the category of aspect: the original, tense-aspect system of Proto-Indo-European is thought to have developed in many branches into a system which is mainly tense-based (Hewson & Bubeník 1997). Aspect features are not included in the original data of DiACL, which spans more families than Indo-European (Carling et al. 2018).

Tense (S4, T1-14) includes two properties designed to capture the typological profile of the forms used to mark present progressive and future. Looking at the present progressive (S4, T1-4), the data distinguish whether a language uses a synthetic form or an analytic construction (with an auxiliary). The model reconstructs a high probability for a synthetic form (0.922), which is entirely consistent with the canonical model. There is an ongoing discussion in comparative-historical syntax as to whether some of the synthetic constructions of Indo-European, such as infinitives, might have an analytic origin (Meier-Brügger et al. 2010:320-21), but this is not relevant in connection to the present progressive, which makes our result uncontroversial.

The results pertaining to the future (S4, T5-14), are also uncontroversial. The data distinguish whether a language uses an analytic construction (with auxiliary), a participle, a particle, a synthetic form, or an aspectual form. Our reconstruction yields a low probability for an analytic future formed by an auxiliary (0.372) but lower probabilities for all other variants. The future in Indo-European is an old issue. Delbrück (1897:242-55) doubted that the future of Sanskrit was derived from a future in Proto-Indo-European, due to its formal similarity with subjunctive and aorist. In further discussions of the verbal system, even within the canonical (Greco-Aryan) model, there is consensus that the

future of Greek and Indo-Aryan is a secondary development (Meier-Brügger et al. 2010:236-42; Rix & Kümmel 2001:10-30; Szemerényi 1989:244-47). Our result cannot contribute to this reconstruction; the probabilities are in general low.

3.4 Word order

Since Greenberg (1966), word order (constituent order, order of meaningful elements) has played a central role in linguistic typological research (Comrie 1981; Dryer 1992; Siewierska 1998). In diachronic syntax, reconstruction of word order remains a controversial issue. At their core, Greenberg's observations targeted implicational relations among word order types, in later literature defined as ORDER OF HEAD AND DEPENDENT (Lehmann 1973), a concept which also includes typological properties beyond word order, following upon the order types (Nichols 1992, 1995, 1998). In Nichols' model, the various dependency types are seen as stable both geographically as well as diachronically, something that is indicated by the fact that the types have regional skewing patterns (Nichols 1995). Word order harmony remains an issue also in computational typology, where the main source of controversy is whether word order patterns are mainly lineage-specific or areal (Baker 2011; Bickel 2011; Croft et al. 2011; Cysouw 2011; Donohue 2011; Dunn et al. 2011). To avoid confusion and not enter into too much detail in the scientific literature on word order, we use the terms 'head-initial' and 'head-final' to refer to the issue of constituent order in sentences and phrases.

In diachronic syntax, the reconstruction of word order is characterized as being beset by methodological difficulties (Roberts 2007:175-98). The source of the uncertainty is the fact that word order in most cases cannot be implicitly connected to any morphosyntactically reconstructable material. Therefore, reconstruction of word order has to be based on, at first, actual observation in attested languages, and second, connections to other properties in language, which may be either reconstructable or not. Consistency and harmony, as well as stability in word order patterns, are central in the model of reconstruction proposed by Lehmann and Nichols (Lehmann 1973, 1974; Nichols 1992), and this is also one of the major sources of criticism against the reconstruction of word order (Lightfoot 2002; Watkins 1976; Winter 1984). For a discussion of these issues and defenses against some of these criticisms, see Campbell & Harris 2002; Harris & Campbell 1995. It is also clear that a diachronic shift from one type to another, e.g. from head-final to head-initial, is a complex evolution where archaic structures are retained and coexist side-by-side with more recent, changed ones (Bauer 1995).

Word order in Proto-Indo-European is the subject of a century-long debate, beginning with Delbrück (2010:38-111) and the study on the position of clitics by Wackernagel (1920). The latter remains one of the few unquestioned reconstructed syntactic features of Proto-Indo-European (Clackson 2007:168). In recent decades, two competing positions on Proto-Indo-European word order have emerged, both of which require a consistency approach. Much of the critique of the word order theories revolves around problems of establishing a default order in ancient languages, which form the basis for a proto-language reconstruction (Winter 1984). Other researchers highlight the general problems of word order reconstruction due to the inherent problem of reconstructing variation and change (Lightfoot 2002; Pires & Thomason 2008). The mainstream position, also given by Delbrück, assumes verb-finality (OV) and head-final order for noun phrases (Clackson 2007; Hock 2013; Lehmann 1973, 1974, 1993, 2002; Mallory & Adams 1997:165-71). The competing position, which bases its discussion on problems of Proto-Indo-European relative clauses, assumes VO and head-initial order for Indo-European (Friedrich 1975).

Our data set contains standardized coding of word order in ancient languages and therefore, the results relate to the discussion of word order reconstruction based on evidence from ancient languages. As a rule, the coding policy aims to capture the DOMINANT word order in a language, but in uncertain cases, the coding system allows for polymorphic coding, i.e. coding a value 1 for two or several variants (Carling et al. 2018). Word order is also split into a relatively high level of granularity, e.g. distinguishing different clause types (S4, WO24-28, 42-46).

Considering the results of our reconstruction (S4, WO1-50), we have to bear in mind that our model does not take into account implicational dependencies between variables (e.g. head-final or head-initial; Dunn et al. 2011; Murawaki 2018; Pagel & Meade 2006). The probability of values of a categorical variable is estimated independently of other variables.

Our model reconstructs SOV order with high probability in main clauses (0.905) as well as subordinate clauses (0.899). Furthermore, our model produces reconstructions of high probability for postpositions (0.849), possessor-noun order (0.585), adjective-noun order (0.870), OV order with participles (0.894), as well as OV order with infinitives (0.806; see Figure 12). These results are compatible with the mainstream view on Proto-Indo-European as a head-final language.

Our results for clitic pronouns (S4, WO5-19) are less simple to interpret. For clitic

pronouns, we distinguish 2nd position, OV, and VO (if the language does not have clitic pronouns, the variable is not applicable to the language), with finite verb, infinitive, and participle. Our model reconstructs distributions of high uncertainty for all categorical features pertaining to clitics. Here, it is obvious that the situation in languages is too complex, with too many gains and losses at hand, for a clear picture to emerge. This result is problematic, considering the safe reconstruction of the position of clitics in Indo-European (Krisch 1990).

Finally, we have an interesting result: a reconstructed high probability of noun-relative clause (0.627) over relative clause-noun (0.292; see Figure 12). The position and construction type of the relative clause was a major source of conflict between Lehmann and Friedrich, and has been extensively discussed in Indo-European syntax (Hock 2013; Watkins 1976). In accordance with the consistency theory by Greenberg (Greenberg 1963), continued by Lehmann (1973, 1974), an OV language is more likely to have relative clause-noun order (Harris & Campbell 1995:363-67), which is also the case in several of the archaic languages, such as Sanskrit and Homeric Greek. However, Hittite (paralleled in, e.g. Latin) has the reverse order relative clause-noun, which is inconsistent with OV (Clackson 2007:171-76). This result can only be taken to be provisional; due to the simplified definition of relative clauses in our data (NRel/RelN, which does not distinguish, e.g. correlative relative clauses, type of clause relation, i.e. paratactic or hypotactic, or restrictive/ non-restrictive types), our result does not bear fully on the issue of Proto-Indo-European relative clauses in the degree of detail with which they are treated in the comparative-historical literature (Hock 2013).

[Figure 12 about here.]

3.5 Comparison of results with traditional models of reconstruction

As described above, our coding scheme for different models of traditional Indo-European reconstruction, which we term canonical, isolating, and active-stative, is based on three sources: Brugmann-Delbrück (Delbrück 1893, 1897, 1900), Hirt (Hirt 1934), and Gamkrelidze-Ivanov (Gamkrelidze & Ivanov 1984; Gamkrelidze et al. 1995). We identify the values reconstructed to Proto-Indo-European by the different models for the variables in our data set to the extent that information is available, though not all variables in our data set are addressed in these sources. The different models have somewhat differing conceptualizations of the nature of the Proto-Indo-European language.
Brugmann-Delbrück do not regard Proto-Indo-European as a diachronically stratified

language; rather, they reconstruct a uniform language, based on Old Indo-Aryan, Greek, Latin, and other ancient Indo-European languages. Compared to the others, their model of Proto-Indo-European is simpler: they reconstruct a highly synthetic stage, which in all branches of the family becomes simplified and less synthetic, losing a number of categories. The other models have a different take on this issue. Scholars reconstructing active-stative and isolating systems presuppose that Proto-Indo-European was a language with several diachronic layers, which changed from a hypothetical early active-stative or isolating stage to a later synthetic stage, found in all ancient languages except for Anatolian. The principles and reasons for this change are important in both the isolating and active-stative models (Gamkrelidze et al. 1995:270-71; Hirt 1934:29-36) as well as in other publications where alternative models are reconstructed for Proto-Indo-European (Bauer 2000; Pooth et al. 2018). The phylogenetic model we employ does not allow us to explicitly stratify the proto-language into layers. It allows us to reconstruct probabilities at the root as well as at ancestral nodes of the tree (see Figures 7–9 and S8); we take the root of the phylogenetic tree to represent the earliest layer of the proto-language in the sources mentioned before, but do not consider any subsequent changes or areal differentiations within Proto-Indo-European. We use a tree sample that is compatible with the Indo-Anatolian hypothesis (S9); the root represents the earliest layer of Proto-Indo-European in all models, before Anatolian split of and subsequent changes began in the Anatolian and Non-Anatolian sub-branches.

The values reconstructed to Proto-Indo-European for each model are found in S4. We assess the extent to which our results agree with the views of each comparative-historical reconstruction model on the basis of the likelihood of each model's reconstructed values for each variable in our data set (where applicable), i.e. the probability with which our model reconstructs the value to Proto-Indo-European for the variable in question. These likelihoods are found in Figure 13; higher values indicate greater agreement. Whereas the active-stative and isolating models differ from our results for some of the domains discussed above, such as nominal and verbal morphology (isolating) or alignment (active-stative), both of these models come close to our reconstruction for other domains such as word order, and in that future tense and definiteness are absent in the reconstructions. At the same time, it is clear that our results show the most agreement with the canonical model of reconstruction (median likelihood = .796), followed by the active-stative (median likelihood = 0.657) and isolating (median likelihood = 0.652) models; agreement with the canonical model is significantly higher than both other models

according to a pairwise Wilcoxon signed-rank test for paired samples ($p < 0.01$ with Benjamini-Hochberg correction for multiple comparisons; variables for which a reliable value was not found for all models are excluded). This indicates that our results most clearly resemble the canonical model of Proto-Indo-European, close to the reconstruction outlined by Brugmann and Delbrück in the nineteenth century.

[Figure 13 about here.]

4 Results: evolutionary dynamics

4.1 Reconstructed probabilities, feature distributions, and transition rates

Our reconstructions are estimated from transition rates inferred on the basis of our tree sample and the features in our data set; these rates characterize the behavior of pairwise transitions between all values of each variable in our data set. Specifically, a transition rate represents the average number of times that a change from a value x (e.g. SOV main clause word order) to a value y (e.g. SVO main clause word order) occurs within a 1000-year span. In this section, we assess the extent to which the frequencies of individual features, as well as transition rates pertaining to them transition rates of individual features — or as a proxy, their diachronic stability or instability — influence the reconstructions produced by our phylogenetic model. It may be the case that only highly stable and frequent features have a chance of being reconstructed to the proto-language with high probability, and less frequent features or features in greater flux are unlikely to be reconstructed. If our model essentially carries out a majority rules-style method of reconstruction, then its utility is severely diminished, as a phylogenetic model is not needed to reconstruct the most frequent pattern. If, however, it picks up on more nuanced patterns of change and incorporates these dynamics into the reconstructions it produces, then the value of methods of this sort is evident. Additionally, an analysis of rates of change alongside reconstruction probabilities provides a better understanding of temporal dynamics within Indo-European.

Does our model simply reconstruct the most frequent feature? If our model simply reconstructs via a majority-rules approach, then there is no real reason to use a phylogenetic model, since our method attends only to the frequency distributions of features across languages, and not patterns of genetic relatedness between said languages.

Furthermore, if this is the case, then certain Anatolian features, if they are rare within Indo-European, face a natural disadvantage and will not be reconstructed, which leads to a result in line with the canonical model of Indo-European reconstruction.

In order to assess the degree of sensitivity of our reconstructions to the frequency distributions of features in our data set, we computed the relative frequency in our data set for features reconstructed with highest probability by our model. We carried out this procedure for all languages, as well as only the ancestral languages in our sample. Figure 14 shows the probabilities with which 'winning' features (black dots) as well as 'non-winning' features (gray dots) are reconstructed, plotted against their relative frequencies in our data set. Visually, it is clear that the relationship between these quantities is noisy; at best, the correlation is moderate (all languages: Spearman's $\rho = 0.58, p < .001$; ancestral languages: $\rho = 0.4, p < .001$; statistical tests include only 'winning features' reconstructed with highest probability to ensure that data points are independent). This indicates that our system's reconstructions are somewhat (but not overwhelmingly) sensitive to the distribution of features within our data set and within more archaic ancestral languages. Certain highly frequent features are not necessarily reconstructed, if our system infers that the feature is likely to have come about many times in parallel. Additionally, an infrequent feature may be reconstructed if is more likely to have survived into certain languages than come about in parallel. This can be illustrated using our system's reconstruction of adpositions: prepositions are frequent in our data set, but infrequent among older languages, and are accordingly reconstructed with low probability.

[Figure 14 about here.]

A few concrete examples serve to exemplify the consequences of this behavior, with respect to the reconstruction of features found in Anatolian to Proto-Indo-European. Figure 7 shows that a masculine/feminine gender distinction is reconstructed by our model to Proto-Indo-European, despite the fact that it is absent in the archaic Anatolian subgroup; because the gender distinction is predominant in Nuclear Indo-European (i.e. all non-Anatolian languages) and because it has been lost several times, our model assigns high probability to a scenario where Anatolian lost the gender distinction during its development (our model also assigns a small amount of probability to a scenario in which the gender distinction was lost independently within Anatolian). In Figure 8, we see that neuter gender is reconstructed to Proto-Indo-European because it is shared by a

substantial number of Core Indo-European (i.e. all of Indo-European to the exclusion of Anatolian and Tocharian) languages and Anatolian, to the exclusion of Tocharian. In Figure 9, we see that our system does not reconstruct prepositions to Proto-Indo-European with high probability; incidentally, although prepositions are reconstructed with high probability for Core Indo-European, our model finds it unlikely that prepositions were lost independently in the history of Anatolian as well as Tocharian (but finds it likely that they were lost on some Indo-Iranian lineages). Hence, we see that Anatolian features have a good chance of being reconstructed to Proto-Indo-European if they are reconstructed to the proto-language of at least one other higher-order or archaic Indo-European branch.

Transition rates   Given the transitions between each pair of values (e.g. ergative → accusative) within a variable, we estimate the overall entry or gain rate and the overall exit or loss rate for individual features according to the formulae given in the supplementary material, Appendix F. The full list of inter-feature transition rates as well as the gain and loss rates for each feature are found in S5–6. A scatterplot of gain and loss rates for each feature, organized according to overarching grammatical categories, is found in Figure 15. The size of individual data points indicates the probability with which a given feature is reconstructed to Proto-Indo-European by our model, with larger size indicating higher probability. The plot is divided according to the median gain and loss rates for our variables; this allows us to divide features into the following four classes of features:

1. High gain rate, high loss rate (upper right quadrant): features of high instability, in frequent flux, gained and lost frequently. These include features pertaining to the presence of case on adjectives, clitics, distinctions between dative and genitive marking, absence of case on nouns, and different alignment systems in the simple past.

2. High gain rate, low loss rate (lower right quadrant): features of high stability to which languages are frequently attracted; gained often and rarely lost. These include features pertaining to the presence of case on nouns, case difference between A and O for nouns as well as pronouns, masculine/feminine distinction, Noun-Relative word order, Possessor-Noun word order, Present progressive by auxiliary, and absence of neuter gender and vocative case.

3. Low gain rate, high loss rate (upper left quadrant): 'recessive' features (cf. Nichols 1993) quickly repulsed by languages when they do occur. These include features pertaining to the presence of future tense by participle, future tense by particle, more than seven cases, more than seven pronominal cases, more than five genders, tripartite alignment, ergative alignment in pronouns, active-stative alignment, double oblique alignment, V2 word order, and VSO word order.

4. Low gain rate, low loss rate (lower left quadrant): highly stable features that arise infrequently. These include features pertaining to the presence of adjective-noun word order, agglutination for case, agreement on prepositions, case on the last member of a NP, definite articles, definite suffixes on adjectives, definite suffixes on nouns, neuter gender, a noun class for animates, and synthetic future tense.

We follow the literature on the diachronic dynamics underlying typological stability and instability (cf. Greenberg 1978; Nichols 2003; Dediu 2010) in our interpretation of these patterns.

[Figure 15 about here.]

For two of the classes, patterns of reconstruction are highly consistent. Stable features that are frequently gained and infrequently lost (lower right quadrant) are virtually always reconstructed with high certainty. Recessive features that are rarely gained and frequently lost (upper left quadrant) are consistently reconstructed with low probability. For the remaining classes, the behavior of our model is more variable. Infrequently gained but stable features (lower left quadrant) are reconstructed with both high and low probability. For instance, SOV word order (in both main and subordinate clauses), adjective-noun word order, neuter gender, postpositions, and vocative case are reconstructed with high probability, whereas features pertaining to definiteness and agglutination, as well as SVO word order (in both main and subordinate clauses) are reconstructed with low probability. The same variability is found for features that fluctuate (upper right quadrant): features pertaining to nominative-accusative alignment, the genitive/dative distinction, and case on adjectives are reconstructed with probability greater than .5, whereas features pertaining to case mergers, clitic word order and the presence of ergativity in the simple past tense are reconstructed with low probability.

It is noteworthy that features pertaining to nominal morphology and tense (inflectional typology) tend to exhibit slow change and features pertaining to alignment and verbal

morphology (agreement) show more rapid patterns of change. Word order is well represented among swift-changing features. These results partly confirm that traits that are immediately bound by morphology, such as nominal morphology and tense, have slower rates of change, in contrast to traits that are not bound by morphology, such as word order and alignment, which have swifter rates of change. Verbal morphology, i.e. agreement patterns, as well as the most stable and frequent word orders, constitute an exception to these tendencies.

All in all, these results show that the reconstructions produced by our model are not simply artifacts of feature distributions across our data set; on the contrary, they reflect multifaceted patterns of change; for certain types of stability or instability, features are reconstructed with either high or low probability, while for other patterns, there is more variability among reconstruction probabilities. These patterns are summarized in Table 6.

[Table 6 about here.]

4.2 Transition rates and grammatical hierarchies

Earlier, we mentioned the existence of asymmetries in the certainty with which features are reconstructed to Proto-Indo-European that correspond to differences in grammatical hierarchies. This asymmetry can be found also in transition rates pertaining to the features in questions. As described in §2.3, we organize our features into hierarchical pairs which belong to the same grammatical category but which vary with respect to other features of the grammar. The categories we identify in our data are restricted to the following features (UNMARKED/MORE FREQUENT < MARKED/LESS FREQUENT; see Table 2):

PRONOUN < NOUN

PRESENT < PAST

AGENT < OBJECT

AGENT/OBJECT < OBLIQUE

MASCULINE/FEMININE < NEUTER

Since loss rates are an inverse measure of the longevity of a given feature (i.e. shorter-lived features are lost at a higher rate), we can measure whether the loss rates differ

significantly across the unmarked/marked feature pairs that we identify in our data set. We find that the loss rates of marked traits are higher than those of unmarked traits ($V = 851, p < 0.001$, according to a one-sided Wilcoxon signed-rank test), indicating that marked traits are lost more frequently than unmarked, more frequent traits.

This is an interesting result, but it is not entirely unexpected. The idea of grammatical or marking hierarchies in the traditional sense (Comrie 1981; Croft 1990; Greenberg 1966) is based on the notion that higher-ranking categories as a rule are more frequent in languages. The idea that more frequent and basic categories, both in grammar and lexicon, are more conservative and archaic, due to their everyday use, has a long history in Indo-European linguistics (Meillet 1948:135). Many lexemes, which are typically part of Swadesh lists, such as kinship words, body parts, numerals, fire, water, and so forth, as a rule preserve more archaic paradigms, including change in stem consonants (e.g. *-r-/-n-*, *-l-/-n-*) or ablauting patterns (qualitative, quantitative) (Meier-Brügger et al. 2010:336-48). The reflexes in daughter languages of the most frequent verbs, such as PIE *$h_1$es-* 'to be' or PIE *$h_1$ey-* 'to go', are typically irregular, preserving archaic inflection patterns and categories (Rix & Kümmel 2001:232-33, 41-42). On the other hand, analogy and other types of changes that harmonize and simplify language structures, making them easier to memorize, are more frequently found among words and categories of lower frequency. By means of phylogenetic methods, we know that there is, at least in the lexicon of basic vocabulary, a correlation between frequency and substitution rates: the most frequent meanings are concepts with generally lower substitution rates (Pagel et al. 2007). Transferred to a scenario of grammatical hierarchies, we expect the unmarked categories, representing the more frequently used categories, to have lower loss rates and longer periods between transitions, whereas we expect the marked categories, representing less frequent categories, to have higher loss rates and shorter periods between transitions.

[Figure 16 about here.]

5 Discussion

In the preceding sections, we presented the results of Proto-Indo-European reconstruction using phylogenetic comparative methods, and provided a careful analysis of our model's behavior in order to better understand the mechanisms underlying the results that it produces. We found broad support for the canonical model of Indo-European syntactic reconstruction, largely because the features reconstructed under alternative models

undergo evolutionary dynamics that make them unlikely to survive into the languages that attest them; they are more likely to have been innovated in parallel. The behavior of the model we use rests on the assumption that we can make inferences about the behavior of linguistic features in prehistory on the basis of their behavior during attested history; calibrating these dynamics according to attested patterns of change is made possible by the use of ancestry constraints. The methodology that we use relies on a number of simplifying assumptions regarding the nature of change. One of these is the assumption of RATE UNIFORMITY, namely that rates of change between values of a linguistic variable are the same on all lineages of Indo-European. There are a number of methods which relax this assumption in order to incorporate RATE VARIATION or HETEROTACHY (Tuffley & Steel 1998; Heath et al. 2011; Beaulieu & O'Meara 2014), but the utility of these methods may be limited relative to their increased computational complexity; phylogenetic linguistic analyses assuming rate homogeneity dovetail well with independent evidence (Blasi et al. 2019), and incorporating heterotachy produces no or little improvement (Chang et al. 2015; Blasi et al. 2020). Our data consist of typological variables rooted in comparative concepts that can be easily operationalized. Other approaches to syntactic reconstruction, rooted in particular syntactic theories, have different assumptions regarding the levels of representation that should be reconstructed to the proto-language (Hale & Kissock 2015); at the same time, similar work makes use of bioinformatic algorithms to address questions of linguistic prehistory (Longobardi et al. 2013). Furthermore, certain syntactic theories make strong predictions regarding syntactic change over many generations of first language acquisition (Berwick & Niyogi 1996; Yang 2000) which, if correct, can potentially find support in phylogenetic models. While some of the assumptions we make may be challenged by other scholars, both in terms of methodology and the nature of the data we employ, we believe that our contribution is of great utility to Indo-European studies, as the assumptions of our probabilistic model are explicit, the results we present are replicable, and we provide a means of explicit evaluation against the hypotheses explored in this paper.

Our results shed light on interesting cross-linguistic diachronic tendencies. At the same time, our data are confined to one family, Indo-European, and the problem of embracing uniformitarianism within one family leads to the problem of sample diversity pointed out by Levy & Daumé (2011). When investigating differences in diachronic dynamics across hierarchies, we build upon insights from a cross-linguistic sample while assessing results derived from one family only, which is not possible without adapting a uniformity-of-state

framework to language evolution: rules that govern language structure are similar in the present and in the past, and all languages reflect some basic universal principles (Croft 2003:233; Roberts 2007:174; Walkden 2019). We are forthcoming in the admission that restricting data to one family gives a limited picture. Some of the observed patterns are obviously uniquely Indo-European, such as the loss of synthetic structure. At the same time, the markedness rubric we employ is derived from cross-linguistic typological observations beyond the scope of Indo-European, allowing us to avoid circular reasoning. On the basis of trends within Indo-European, we confirm broader hypotheses regarding cross-linguistically universal tendencies, namely that factors such as economy and frequency interfere in the processes of language evolution (Croft 2003; Haspelmath 2015). Furthermore, our finding that features pertaining to unmarked, more frequent categories are lost more frequently than those of marked, less frequent categories dovetails nicely with the common finding that frequently used items are resistant to various types of change (e.g. Diessel 2007).

It is remarkable how close our comparative phylogenetic reconstruction approaches to a canonical reconstruction model of Indo-European syntax (Delbrück 1893, 1897, 1900; Krahe et al. 1972). Despite all the variation and change in the grammar in the Indo-European family, our model reconstructs a highly synthetic, mainly head-final language, with nominative-accusative alignment, independent of tense and animacy degree of the first argument, case marking on nouns, no definite article, three genders (masculine, feminine, neuter), predicative gender agreement, a non-agglutinating case system with less than seven cases but with a nominative, accusative, dative, genitive, and vocative, also in pronouns, a synthetic present, no future, full agreement in present tense of verbs, but not in the past tense, postpositions, OV infinitive word order, SOV in main and subordinate clauses, possessor-noun, adjective-noun, noun-relative clause, OV participle word order, and WH-initial word order. The outcome is striking. We see the structure of a grammatical system which has been retained to a high degree through many branches of Indo-European, and which is remarkably constant, despite several millennia of language contact and change, loss of categories, emergence of new categories by grammaticalization, and substantial typological changes, for instance in word order patterns.

Alternative models (see 1.1, 2.3) assume far-reaching typological changes between Proto-Indo-European and the Non-Anatolian sub-branches of the tree. These changes are supported by internal reconstruction based on Proto-Indo-European paradigmatic correlations in combination with a comparison with other unrelated language families (see

§3). Given a nuanced understanding of Indo-European chronology (Bouckaert et al. 2012; Chang et al. 2015; Meid 1975; Schlerath 1981), as well as both attested as well as estimated information regarding the timespans characterizing change between typological features of the type that we investigate here (Croft 2003:252; Haspelmath 2018; Hock & Joseph 1996:183-184), it is increasingly clear why there is limited support for the alternative theories. On the basis of what can be inferred regarding change between alignment systems of languages in our sample, for instance, a relatively rapid development from ergative or active-stative alignment (as assumed by the active-stative model) or via grammaticalization from an isolating system with as yet undeveloped agreement marking (as assumed by the isolating model) to a nominative-accusative system is less likely than the retention of nominative-accusative alignment. All these models take into account a large amount of data (morphological, typological) that may be connected directly or indirectly to the typological traits investigated in our study.

In the realm of alignment, nominative-accusative alignment is dominant in many contemporary and most historical states of the family (Carling 2019:31-50), and also reconstructed to the proto-language (Figure 5). Absence of agreement marking (features involving NO MARKING) which implies isolating structure, is reconstructed to the proto-language with low probability. In conclusion, nominative-accusative alignment is stable and the only noteworthy trend over time is a higher rate of transition from nominative-accusative to no marking (SYNTHETIC > ISOLATING). Development of other systems (ergative, active-stative, tripartite) from nominative-accusative are marginal and have low transition rates. Similar patterns can be seen across other data types: in general, within Indo-European, there is greater evidence for developments in the directions SYNTHETIC > ISOLATING and SYNTHETIC > AGGLUTINATING. Developments in the opposite direction are not impossible, but are unlikely to have taken place between Proto-Indo-European and its descendants. An exception is word order, where features exhibit varying degrees of stability and instability.

Our results are of key relevance to larger discussions regarding typological stability, as well as the suitability of typological data for language classification (Dediu & Levinson 2012; Dediu & Cysouw 2013; Dunn et al. 2011; Plank 2011). We observe two overarching feature classes characterized by different patterns of change. Either the change is slow and overwhelmingly unidirectional, moving from synthetic to isolating, which is found mainly in the paradigmatic categories, i.e. nominal morphology and parts of verbal morphology, where a synthetic system is broken down in the direction of an isolating system. This type

of change conforms to a model of a unidirectional, cyclic typological change, occurring at a slow change rate (Croft 2003:227ff.). Alternatively, the evolutionary trend is oscillating, with high amounts of gains and losses. This occurs in the syntactic (non-paradigmatic) categories, mainly by word order, alignment, and partly by verbal morphology traits. This type generally conforms to a theory of a punctuated and non-directional change, which may take any direction depending on a combination of internal pressure and areality-induced change (Dixon 1997). Given this result, the search for a phylogenetic signal in the evolution of non-homologous, structural linguistic features is difficult without considering areality and ancient language data.

Finally, we employed a relatively uncontroversial and neutral model of change that has been used in state-of-the art work in phylogenetic linguistics. Simple models like the Continuous-Time Markov Process assume that a feature can be born and die (or that change between traits of a multistate character can occur) with a given rate, and there are no a priori restrictions on the values taken by these rates, as long as they are positive. These models are appropriate for grammatical and lexical data, and are suitable in situations when the system may leave and return to states, as in the case of variants of word order, agreement, case, or different lexical meanings. More complex models like the stochastic dollo character (SDC; Nicholls & Gray 2006), which assume that features are born only once in the history of a language family, are useful for features which cannot return to identical states, such as morphological traits bound to specific forms, irreversible outcomes of sound change (e.g. mergers), or features which may undergo grammaticalization. Using an SDC model would likely yield different results for our data, at least for gender. At the same time, many scholars agree that SDC models are unsuitable for not only comparative concepts, which aim to capture features cross-linguistically, independent of linguistic matter, but other types of linguistic characters (Chang et al. 2015), though modifications thought to be more appropriate for linguistic data have been proposed (Bouckaert & Robbeets 2017). However, we are confident that a Bayesian approach has the capacity beyond a comparative-historical model to contribute in a meaningful way to the theoretical discussions about trends in diachrony, directionality of syntax, rates of gains and losses, stability of features and categories, as well as correlations to important aspects of typology such as frequency, economy, hierarchies, and general trends in grammar change. Future work can potentially contrast the results of different evolutionary models in applications like the one undertaken in this paper; researchers wishing to argue for a specific evolutionary model

over others (along with its concomitant result) may employ posterior predictive checks (Box 1980; see also supplementary material, Appendix D) to demonstrate that their model is a better fit to the data than others.

6 CONCLUSION

The current paper had several foci. We reconstructed the evolutionary history of selected aspects of Indo-European morphosyntax by means of a model which infers patterns of diachronic development of linguistic features over a phylogeny. This allowed us to infer the most probable value of a given linguistic variable in the unattested Proto-Indo-European language. We used a dataset of binary coded comparative concepts, recoded as categorical features, which also contained data from extinct and historical Indo-European languages. We focused on five categories of grammar: alignment, nominal morphology, verbal morphology, tense, and word order. The result at the proto-language state we compared to previous reconstructions of Proto-Indo-European grammar by means of the comparative-historical method and diachronic typology. The methodology that we used allowed us to compare ideas from the traditional comparitive-historical linguistic literature with our model's output. We found that phylogenetic reconstruction produced a consistent and coherent system, which corresponded to a highly synthetic, mainly head-final language, with nominative-accusative alignment, independent of tense and animacy degree of the first argument, case marking on nouns, no definite article, three genders (masculine, feminine, neuter), predicative gender agreement, a non-agglutinating case system with less than seven cases but with a nominative, accusative, dative, genitive, and vocative, also in pronouns, a synthetic present, no future, full agreement in present tense of verbs, but not in the past tense, postpositions, OV infinitive word order, SOV in main and subordinate clauses, possessor-noun, adjective-noun, noun-relative clause, OV participle word order, and WH-initial word order. This reconstruction matched a canonical model of Proto-Indo-European grammar, as it was described by the Neogrammarians already in the nineteenth century.

We also analyzed the inferred inter-feature transition rates on which our reconstructions are based. Our analysis shed light on different tendencies of change across features. In general, traits that were reconstructed to the proto-language had relatively low loss and gain rates, which implies that the reconstructed typological system is consistent and stable in the family. The most noteworthy tendency was a change from synthetic to isolating structure. In addition, a general tendency was for morphological

(paradigmatic) categories (nominal morphology and tense) to have low change rates, whereas syntactic categories (alignment and word order) to have higher change rates. Verbal morphology opposed this tendency with high change rates. Finally, we divided our grammatical traits (excluding word order) into hierarchical pairs by different members of categories available in our data, such as tense (present, past), word class (noun, pronoun), or gender. We found that the unmarked, more frequent traits are lost less frequently than marked, less frequent traits; this difference was significant.

In sum, our result supported the theory that grammar evolution is both divergent, down to the level of highest granularity, as well as following general, universal principles. Over the 6000–7000 year cycle represented in our data, morphological traits tended to show a unidirectional path of change, fundamentally moving from synthetic to isolating, whereas word order, alignment and person agreement properties showed more non-directional and unpredictable paths of change, with higher rates of gains and losses. The results were in line both with a cyclic and a punctuated model of change. Our results also indicated that the variability of change in grammar over time is governed by general, 'universal' tendencies, such as grammatical hierarchies and frequency.

REFERENCES

AISSEN, JUDITH. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21.435–483.

BAERMAN, MATTHEW, and DUNSTAN BROWN. 2013. Syncretism in verbal person/number marking. *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

BAKER, MARK C. 2011. The interplay between universal grammar, universals, and lineage specificity. *Linguistic Typology* 15.473–482.

BAMMESBERGER, ALFRED. 1986. *Untersuchungen zur vergleichenden Grammatik der germanischen Sprachen. Bd 1, Der Aufbau des germanischen Verbalsystems.* Heidelberg: Winter.

BARÐDAL, JÓHANNA. 2014. Syntax and syntactic reconstruction. *The Routledge handbook of historical linguistics*, ed. by Claire Bowern and Bethwyn Evans, 343–373. London - New York: Routledge.

BARÐDAL, JÓHANNA, and THÓRHALLUR EYTHÓRSSON. 2009. The origin of the oblique subject construction: an Indo-European comparison. *Grammatical change in Indo-European languages*, ed. by Vit Bubenik, John Hewson, and Sarah Rose, 179–193. Amsterdam-Philadelphia: John Benjamins.

BAUER, BRIGITTE. 2000. *Archaic syntax in Indo-European: the spread of transitivity in Latin and French.* Berlin: Mouton de Gruyter.

BAUER, BRIGITTE. 2007. The definite article in Indo-European. Emergence of a new grammatical category? *Nominal Determination: Typology, context constratis, and historical emergence*, ed. by Elisabeth Stark, Elisabeth Leiss, and Werner Abraham, 103–139. Amsterdam-Philadelphia: John Benjamins.

BAUER, BRIGITTE L. M. 1995. *The emergence and development of SVO patterning in Latin and French: Diachronic and psycholinguistic perspectives.* Oxford: Oxford University Press.

BEAULIEU, JEREMY M., and BRIAN C. O'MEARA. 2014. Hidden Markov models for studying the evolution of binary morphological characters. *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice*, ed. by László Zsolt Garamszegi, 395–408. Heidelberg, New York, Dordrecht, London: Springer.

BERWICK, ROBERT C, and PARTHA NIYOGI. 1996. Learning from triggers. *Linguistic Inquiry* 27.605–622.

BICKEL, BALTHASAR. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11.239–251.

BICKEL, BALTHASAR. 2008. On the scope of the referential hierarchy in the typology of grammatical relations. *Case and Grammatical Relations. Studies in honor of Bernard Comrie*, ed. by G. Corbett Greville and Michael Noonan, 191–210. Amsterdam - Philadelphia: John Benjamins.

BICKEL, BALTHASAR. 2011. Statistical modeling of language universals. *Linguistic Typology* 15.401–413.

BICKEL, BALTHASAR, and JOHANNA NICHOLS. 2002. Autotypologizing databases and their use in fieldwork. *Proceedings of the International LREC Workshop on Resources*

and Tools in Field Linguistics, Las Palmas, 26 - 27 May 2002, ed. by Peter Austin, Helen Dry, and Peter Witternburg. Nijmegen: ISLE and DOBES.

BLASI, DAMIÁN E; STEVEN MORAN; SCOTT MOISIK; PAUL WIDMER; DAN DEDIU; and BALTHASAR BICKEL. 2020. Languages, evolution and statistics: human sound systems were shaped by changes in bite configuration. Response to Tarasov & Uyeda (2020). *BioRxiv*. Online: `https://www.biorxiv.org/content/10.1101/2020.02.27.965400v1.abstract`.

BLASI, DAMIÁN E; STEVEN MORAN; SCOTT R MOISIK; PAUL WIDMER; DAN DEDIU; and BALTHASAR BICKEL. 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* 363(6432).

BOLLBACK, JONATHAN P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7.88. Online: `https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-88`.

BOPP, FRANZ. 1816. *Über das Conjugationssystem der Sanskritsprache*. Frankfurt am Main: Andreas.

BORNKESSEL-SCHLESEWSKY, INA; A. L. MAL UKOV; and MARC RICHARDS. 2015. *Scales and hierarchies: a cross-disciplinary perspective*. Berlin: Mouton De Gruyter.

BOUCKAERT, REMCO; PHILIPPE LEMEY; MICHAEL DUNN; SIMON J. GREENHILL; ALEXANDER V. ALEKSEYENKO; ALEXEI J. DRUMMOND; RUSSELL D. GRAY; MARC A. SUCHARD; and QUENTIN D. ATKINSON. 2012. Mapping the origins and expansion of the indo-european language family. *Science* 337.957–960.

BOUCKAERT, REMCO R, and MARTINE ROBBEETS. 2017. Pseudo Dollo models for the evolution of binary characters along a tree. *BioRxiv*. Online: `https://www.biorxiv.org/content/10.1101/207571v1`.

BOX, GEORGE E. P. 1980. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)* 143.383–430.

BRUGMANN, KARL, and BERTHOLD DELBRÜCK. 1893. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen: kurzgefasste Darstellung der Geschichte des Altindischen, Altiranischen (Avestischen u. Altpersischen), Altarmenischen, Altgriechischen, Albanesischen, Lateinischen, Oskisch-Umbrischen, Altirischen,*

Gotischen, Althochdeutschen, Litauischen und Altkirchenslavischen. Bd 3, Vergleichende Syntax der indogermanischen Sprachen, T. 1. Strassburg: Trübner.

BRUGMANN, KARL, and BERTHOLD DELBRÜCK. 1897. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen: kurzgefasste Darstellung der Geschichte des Altindischen, Altiranischen (Avestischen u. Altpersischen), Altarmenischen, Altgriechischen, Albanesischen, Lateinischen, Oskisch-Umbrischen, Altirischen, Gotischen, Althochdeutschen, Litauischen und Altkirchenslavischen. Bd 4, Vergleichende Syntax der indogermanischen Sprachen, T. 2.* Strassburg: Trübner.

CALUDE, ANDREEA S., and ANNEMARIE VERKERK. 2016. The typology and diachrony of higher numerals in indo-european: a phylogenetic comparative study. *Journal of Language Evolution* 1.91–108.

CAMPBELL, LYLE, and ALICE C. HARRIS. 2002. Syntactic reconstruction and demythologizing 'myths and the prehistory of grammars'. *Journal of Linguistics* 38.599–618.

CARLING, GERD. 2012. Development of form and function in a case system with layers: Tocharian and Romani compared. *Tocharian and Indo-European Studies* 13.57–76.

CARLING, GERD. 2019. *Mouton Atlas of Languages and Cultures. Vol. 1: Europe and West, Central and South Asia.* Berlin - Boston: Mouton de Gruyter.

CARLING, GERD; FILIP LARSSON; CHUNDRA A. CATHCART; NIKLAS JOHANSSON; ARTHUR HOLMER; ERICH R. ROUND; and ROB VERHOEVEN. 2018. Diachronic Atlas of Comparative Linguistics (DiACL) – A Database for Ancient Language Typology. *PLOS ONE* 13(10). Online: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205313`.

CATHCART, CHUNDRA A. 2018. Modeling linguistic evolution: a look under the hood. *Linguistics Vanguard* 4. Online: `https://doi.org/10.1515/lingvan-2017-0043`.

CATHCART, CHUNDRA A.; GERD CARLING; FILIP LARSSON; NIKLAS JOHANSSON; and ERICH R. ROUND. 2018. Areal pressure in grammatical evolution. *Diachronica* 35.1–34.

CATHCART, CHUNDRA A.; ANDREAS HÖLZL; GERHARD JÄGER; PAUL WIDMER; and BALTHASAR BICKEL. 2020. Numeral classifiers and number marking in Indo-Iranian: a phylogenetic approach. *Language Dynamics and Change*, 1 – 53.

CHANG, WILL; CHUNDRA A. CATHCART; DAVID HALL; and ANDREW GARRETT. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91.194–244.

CLACKSON, JAMES. 2007. *Indo-European linguistics: an introduction.* Cambridge, UK: Cambridge University Press.

COMRIE, BERNARD. 1981. *Language universals and linguistic typology: syntax and morphology.* Oxford: Blackwell.

CROFT, WILLIAM. 1990. *Typology and universals.* Cambridge: Cambridge University Press.

CROFT, WILLIAM. 2003. *Typology and universals.* Cambridge: Cambridge University Press.

CROFT, WILLIAM; TANMOY BHATTACHARYA; DAVE KLEINSCHMIDT; D. ERIC SMITH; and T. FLORIAN JAEGER. 2011. Greenbergian universals, diachrony, and statistical analyses. *Linguistic Typology* 15.433–453.

CYSOUW, MICHAEL. 2011. Understanding transition probabilities. *Linguistic Typology* 15.415–431.

DEDIU, DAN. 2010. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society of London B* 278.474–79.

DEDIU, DAN, and MICHAEL CYSOUW. 2013. Some structural aspects of language are more stable than others: A comparison of seven methods. *PLOS ONE* 8(1).e55009. Online: `https://doi.org/10.1371/journal.pone.0055009`.

DEDIU, DAN, and STEPHEN C. LEVINSON. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLOS ONE* 7(10).e45198. Online: `https://doi.org/10.1371/journal.pone.0045198`.

DELBRÜCK, BERTHOLD. 1893. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen : kurzgefasste Darstellung der Geschichte des Altindischen, Altiranischen (Avestischen u. Altpersischen), Altarmenischen, Altgriechischen,*

Albanesischen, Lateinischen, Oskisch-Umbrischen, Altirischen, Gotischen, Althochdeutschen, Litauischen und Altkirchenslavischen. Bd 3, Vergleichende Syntax der indogermanischen Sprachen, T. 1. Strassburg: Trübner.

DELBRÜCK, BERTHOLD. 1897. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen : kurzgefasste Darstellung der Geschichte des Altindischen, Altiranischen (Avestischen u. Altpersischen), Altarmenischen, Altgriechischen, Albanesischen, Lateinischen, Oskisch-Umbrischen, Altirischen, Gotischen, Althochdeutschen, Litauischen und Altkirchenslavischen. Bd 4, Vergleichende Syntax der indogermanischen Sprachen, T. 2.* Strassburg: Trübner.

DELBRÜCK, BERTHOLD. 1900. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen : kurzgefasste Darstellung der Geschichte des Altindischen, Altiranischen (Avestischen u. Altpersischen), Altarmenischen, Altgriechischen, Albanesischen, Lateinischen, Oskisch-Umbrischen, Altirischen, Gotischen, Althochdeutschen, Litauischen und Altkirchenslavischen. Bd 5, Vergleichende Syntax der indogermanischen Sprachen, T. 3.* Strassburg: Trübner.

DELBRÜCK, BERTHOLD. 2010. *Vergleichende Syntax der indogermanischen Sprachen, Cambridge Library Collection - Linguistics*, vol. 2. Cambridge: Cambridge University Press. Online: `https://www.cambridge.org/core/books/vergleichende-syntax-der-indogermanischen-sprachen/78513CA5BC38E27D42F1C6788BB59781`.

DIESSEL, HOLGER. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25.108–127.

DIXON, R. M. W. 1997. *The rise and fall of languages.* Cambridge: Cambridge University Press.

DONOHUE, MARK. 2011. Stability of word order: Even simple questions need careful answers. *Linguistic Typology* 15.381–391.

DRYER, MATTHEW, and MARTIN HASPELMATH (eds.) 2013. *World Atlas of Linguistic Structures.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

DRYER, MATTHEW S. 1992. Greenbergian word order correlations. *Language* 68.81–138.

DRYER, MATTHEW S. 2011. The evidence for word order correlations. *Linguistic Typology* 15.335–380.

DUNN, MICHAEL; TONYA KIM DEWEY; CARLEE ARNETT; THÓRHALLUR EYTHÓRSSON; and JÓHANNA BARÐDAL. 2017. Dative sickness: A phylogenetic analysis of argument structure evolution in Germanic. *Language* 93.e1–e22. Online: `https://muse.jhu.edu/article/651291`.

DUNN, MICHAEL; SIMON J. GREENHILL; STEPHEN C. LEVINSON; and RUSSELL D. GRAY. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473.79–82.

FELSENSTEIN, JOSEPH. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17.367–76.

FELSENSTEIN, JOSEPH. 2004. *Inferring phylogenies.* Sunderland, Mass.: Sinauer.

FERRARESI, GISELLA, and MARIA GOLDBACH. 2008. *Principles of syntactic reconstruction.* Amsterdam: John Benjamins.

FRIEDRICH, PAUL. 1975. *Proto-Indo-European syntax: the order of meaningful elements.* Butte, Mont.: Institute for the Study of Man.

GAMKRELIDZE, TAMAZ VALERIANOVIČ, and VJAČESLAV VSEVOLODOVIČ IVANOV. 1984. *Indoevropejskij jazyk i indoevropejcy: rekonstrukcija i istoriko-tipologičeskij analiz prajazyka i protokul tury = Indo-European and the Indo-Europeans: a reconstruction and historical typological analysis of a protolanguage and a proto-culture.* Tbilisi: Izd. Tbilisskogo univ.

GAMKRELIDZE, TAMAZ VALERIANOVIČ; VJAČESLAV VSEVOLODOVIČ IVANOV; and WERNER WINTER. 1995. *Indo-European and the Indo-Europeans: a reconstruction and historical analysis of a proto-language and a proto-culture.* Berlin: Mouton de Gruyter.

GARRETT, ANDREW. 2008. Paradigmatic uniformity and markedness. *Explaining linguistic universals: Historical convergence and universal grammar*, ed. by Jeff Good, 124–143. Oxford: Oxford University Press.

GREENBERG, JOSEPH H. 1963. *Universals of language: report of a conference held at Dobbs Ferry, New York, April 13-15, 1961.* Cambridge, Mass.: MIT Press.

GREENBERG, JOSEPH H. 1966. *Language universals: with special reference to feature hierarchies.* The Hague: Mouton.

GREENBERG, JOSEPH H. 1978. *Universals of human language.* Stanford: Stanford University Press.

GREENBERG, JOSEPH H., and MARTIN HASPELMATH. 2005. *Language universals: With special reference to feature hierarchies.* Berlin - New York: De Gruyter.

HALE, MARK, and MADELYN KISSOCK. 2015. Syntactic reconstruction: The correspondence problem revisited. Paper presented at the 17th Diachronic Generative Syntax Conference (DIGS 17), Reykjavik, May 31. Online: `http://modlang-phonetica.concordia.ca/Iceland-Talk.pdf`.

HARRIS, ALICE C. 2008. Reconstruction in syntax. Reconstruction of patterns. *Principles of syntactic reconstruction*, ed. by Gisella Ferraresi and Maria Goldbach, 73–95. Amsterdam - Philadelphia: John Benjamins.

HARRIS, ALICE C., and LYLE CAMPBELL. 1995. *Historical syntax in cross-linguistic perspective.* Cambridge: Cambridge University Press.

HASPELMATH, MARTIN. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42.25–70.

HASPELMATH, MARTIN. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86.663–687.

HASPELMATH, MARTIN. 2015. Descriptive scales versus comparative scales. *Scales and hierarchies. A cross-disciplinary perspective*, ed. by Ina Bornkessel-Schlesewsky, Andrej Mal ukov, and Marc D. Richards, 45–58. Berlin - Boston: De Gruyter.

HASPELMATH, MARTIN. 2018. Revisiting the anasynthetic spiral. *Grammaticalization from a typological perspective*, ed. by Heiko Narrog and Berndt Heine. Oxford: Oxford University Press.

HEATH, TRACY; MARK HOLDER; and JOHN HUELSENBECK. 2011. A Dirichlet Process Prior for estimating lineage-specific substitution rates. *Molecular Biology and Evolution* 29.939–55.

HEWSON, JOHN, and VÍT BUBENÍK. 1997. *Tense and aspect in Indo-European languages: theory, typology, diachrony.* Amsterdam: Benjamins.

HIRT, HERMANN ALFRED. 1934. Das Nomen. Kasuslehre. *Indogermanische Grammatik. T. 6, Syntax, 1: syntaktische Verwendung der Kasus und der Verbalformen,* 29–75. Heidelberg: Carl Winter.

HIRT, HERMANN ALFRED. 1937. *Indogermanische Grammatik. T. 7, Syntax, 2: die Lehre vom einfachen und zusammengesetzten Satz.* Heidelberg: Carl Winter.

HOCK, HANS HENRICH. 2013. Proto-Indo-European verb-finality. Reconstruction, typology, validation. *Journal of Historical Linguistics* 3.49–76.

HOCK, HANS HENRICH, and BRIAN D. JOSEPH. 1996. *Language history, language change, and language relationship: an introduction to historical and comparative linguistics.* Berlin: Mouton de Gruyter.

HUELSENBECK, JOHN P.; RASMUS NIELSEN; and JONATHAN P. BOLLBACK. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52.131–58.

JASANOFF, JAY H. 1978. *Stative and middle in Indo-European.* Innsbruck: Institut für Sprachwissenschaft.

JASANOFF, JAY H. 2003. *Hittite and the Indo-European verb.* Oxford: Oxford University Press, 2003.

JOSEPHSON, FOLKE, and INGMAR SÖHRMAN. 2008. *Interdependence of diachronic and synchronic analyses.* Amsterdam: John Benjamins.

JÄGER, GERHARD. 2019. Computational historical linguistics. *Theoretical Linguistics* 45.151–182.

KLIMOV, GEORGIJ ANDREEVICH. 1974. On the character of languages of active typology. *Linguistics* 12.11–76.

KRAHE, HANS; HANS SCHMEJA; and WOLFGANG MEID. 1972. *Grundzüge der vergleichenden Syntax der indogermanischen Sprachen.* Innsbruck: Institut für vergleichende Sprachwissenschaft.

Krause, Wolfgang, and Werner Thomas. 1960. *Tocharisches Elementarbuch. B. 1, Grammatik.* Heidelberg: Carl Winter.

Krisch, Thomas. 1990. Das Wackernagelsche Gesetz aus heutiger Sicht. *Sprachwissenschaft und Philologie. Jacob Wackernagel und die Indogermanistik heute*, ed. by Heiner Eichner and Helmut Rix, 64–81. Wiesbaden: Harrassowitz.

Kulikov, L. I., and Nikolaos Lavidas. 2015. *Proto-Indo-European syntax and its development.* Amsterdam: John Benjamins.

Ledgeway, Adam, and Ian G. Roberts. 2017. *The Cambridge handbook of historical syntax.* Cambridge: Cambridge University Press.

Lehmann, Winfred P. 1973. A structural principle of language and its implications. *Language* 49.47–66.

Lehmann, Winfred P. 1974. *Proto-Indo-European syntax.* Austin: Univ. of Texas P.

Lehmann, Winfred P. 1993. *Theoretical bases of Indo-European linguistics.* London: Routledge.

Lehmann, Winfred P. 2002. *Pre-Indo-European.* Washington, D.C.: Institute for the Study of Man.

Levy, Roger, and Hal Daumé. 2011. Computational methods are invaluable for typology, but the models must match the questions. *Linguistic Typology* 15.393–399.

Liggett, Thomas M. 2010. *Continuous time Markov processes: An introduction, Graduate Studies in Mathematics*, vol. 113. Providence, RI: American Mathematical Society.

Lightfoot, David W. 2002. Myths and the prehistory of grammars. *Journal of Linguistics* 38.113–36.

Longobardi, Giuseppe; Cristina Guardiano; Giuseppina Silvestri; Alessio Boattini; and Andrea Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3.122–152.

Longobardi, Giuseppe, and Ian Roberts. 2011. Non-arguments about non-universals. *Linguistic Typology* 15.483–495.

Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45.435–64.

Mallory, James P., and Douglas Q. Adams. 1997. *Encyclopedia of Indo-European culture.* London: Fitzroy Dearborn.

Mal ukov, Andrej. 2015. Towards a typology of split ergativity: A TAM-hierarchy for alignment splits. *Scales and hierarchies: A cross-disciplinary perspective*, ed. by Ina Bornkessel-Schlesewsky, Andrej Mal ukov, and Marc D. Richards, 275–96. Berlin - New York: Mouton de Gruyter.

Martinet, André. 1962. *A functional view of language: being the Waynflete lectures delivered in the College of St. Mary Magdalen, Oxford 1961.* Oxford: Clarendon Press.

Matasović, Ranko. 2004. *Gender in Indo-European.* Heidelberg: Winter.

Maurits, Luke, and Thomas L. Griffiths. 2014. Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences* 111.13576–13581.

Meid, Wolfgang. 1975. Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen. *Flexion und Wortbildung*, ed. by Helmut Rix, 204–219. Wiesbaden: Ludwig Reichert.

Meier-Brügger, Michael; Matthias Fritz; and Manfred Mayrhofer. 2010. *Indogermanische Sprachwissenschaft.* Berlin: de Gruyter.

Meillet, Antoine. 1948. *Linguistique historique et linguistique générale.* Paris: Champion.

Melchert, Craig. 2000. Tocharian plurals in -nt and related phenomena. *Tocharian and Indo-European Studies* 9.53–75.

Murawaki, Yugo. 2018. Analyzing correlated evolution of multiple features using latent representations. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4371–4382. Brussels: Association for Computational Linguistics. Online: https://www.aclweb.org/anthology/D18-1468.

Nicholls, Geoff K., and Russell D. Gray. 2006. Quantifying uncertainty in a stochastic Dollo model of vocabulary evolution. *Phylogenetic methods and the prehistory*

of languages, ed. by Peter Forster and Colin Renfrew, 161–71. Cambridge: McDonald Institute for Archaeological Research.

NICHOLS, JOHANNA. 1992. *Linguistic diversity in space and time.* Chicago: University of Chicago Press.

NICHOLS, JOHANNA. 1993. Ergativity and linguistic geography. *Australian Journal of Linguistics* 13.39–89.

NICHOLS, JOHANNA. 1995. Diachronically stable structural features. *Historical linguistics 1993. Selected papers from the 11th International Conference on Historical Linguistics. Los Angeles 16-20 August 1993*, ed. by Henning Andersen, 337–355. Amsterdam - Philadelphia: John Benjamins.

NICHOLS, JOHANNA. 1998. The Eurasian spread zone and the Indo-European dispersal. *Archaeology and language II. Archaeological data and linguistic hypotheses*, ed. by Roger Blench and Matthew Spriggs, 220–266. New York: Routledge.

NICHOLS, JOHANNA. 2003. Diversity and stability in languages. *The Oxford handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 283–310. Oxford: Blackwell.

NIELSEN, RASMUS. 2002. Mapping mutations on phylogenies. *Systematic Biology* 51.729–39.

PAGEL, MARK; QUENTIN D. ATKINSON; and ANDREW MEADE. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449.717–720.

PAGEL, MARK, and ANDREW MEADE. 2006. Bayesian analysis of correlated evolution of discrete characters by Reversible-Jump Markov Chain Monte Carlo. *The American Naturalist* 167.808–25.

PIRES, ACRISIO, and SARAH G. THOMASON. 2008. How much syntactic reconstruction is possible? *Principles of syntactic reconstruction*, ed. by Gisella Ferraresi and Maria Goldbach, 27–72. Amsterdam: Benjamins.

PLANK, FRANS. 2011. Call for debate re word-order universals. *Linguistic Typology* 15.333–334.

Pooth, Roland; Peter Kerkhof; Leonid Kulikov; and Jóhanna Barðdal. 2018. The origin of non-canonical case marking of subjects in Proto-Indo-European: Accusative, ergative, or semantic alignment. *Indogermanische Forschungen* 124.245–64.

Rix, Helmut, and Martin Kümmel. 2001. *LIV: Lexikon der indogermanischen Verben: die Wurzeln und ihre Primärstammbildungen.* Wiesbaden: Reichert.

Roberts, Ian G. 2007. *Diachronic syntax.* Oxford: Oxford University Press.

Schlerath, Bernfried. 1981. Ist ein Raum/Zeit-Modell für eine rekonstruierte Sprache möglich? *Zeitschrift für vergleichende Sprachforschung* 95.175–202.

Schmidt, Karl Horst. 1979. Reconstructing active and ergative stages of Pre-Indo-European. *Ergativity: Towards a theory of grammatical relations.*, ed. by Frans Plank, 333–345. London: Academic Press.

Schmidt, Karl Horst. 1982. Typusrelevanter Sprachwandel flektierend zu agglutinierend und seine Korrelationen. *Études Finno-Ougriennes* XV.335–346.

Siewierska, Anna. 1998. *Constituent order in the languages of Europe.* Berlin: Mouton de Gruyter.

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

Sturtevant, Edgar H. 1962. The Indo-Hittite hypothesis. *Language* 38.105–110.

Szemerényi, Oswald. 1989. *Einführung in die vergleichende Sprachwissenschaft.* Darmstadt: Wissenschaftl. Buchgesellschaft.

Szemerényi, Oswald. 1996. *Introduction to Indo-European linguistics.* Oxford: Clarendon Press.

Tichy, Eva. 1993. Kollektiva, Genus femininum und relative Chronologie im Indogermanischen. *Historische Sprachforschung / Historical Linguistics* 106.1–19.

Tuffley, Chris, and Mike Steel. 1998. Modeling the covarion hypothesis of nucleotide substitution. *Mathematical Biosciences* 147.63–91.

UHLENBECK, C. CORNELIS. 1901.   Agens und Patiens im Kasussystem der indogermanischen Sprachen. *Indogermanische Forschungen* 12.170–171.

VAILLANT, A. 1936. *L' ergatif indo-européen.* C. Klincksieck.

VITI, CARLOTTA. 2015. Historical syntax: Problems, materials, methods, hypotheses. *Perspectives on historical syntax*, ed. by Carlotta Viti, 3–34. Amsterdam - Philadelphia: John Benjamins.

WACKERNAGEL, JACOB. 1920. *Vorlesungen über Syntax mit besonderer Berücksichtigung von Griechisch, Lateinisch und Deutsch.* Basel: in Kommissionsvlg von E. Birkhäuser.

WALKDEN, GEORGE. 2013.   The correspondence problem in syntactic reconstruction. *Diachronica* 30.95–122.

WALKDEN, GEORGE. 2019. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics* 4.52. Online: `http://doi.org/10.5334/gjgl.888`.

WATKINS, CALVERT. 1976.   Towards Proto-Indo-European syntax: problems and pseudoproblems. *Papers from the parasession on diachronic syntax*, ed. by Sanford Steever, Carol A. Walker, and Salikoko S. Mufwene, 305–326. Chicago: Chicago Linguistic Society.

WICHMANN, SØREN. 2014. Diachronic stability and typology. *The Routledge handbook of historical linguistics*, ed. by Claire Bowern and Bethwyn Evans, 212–224. London - New York: Routledge.

WIDMER, MANUEL; SANDRA AUDERSET; JOHANNA NICHOLS; PAUL WIDMER; and BALTHASAR BICKEL. 2017. NP recursion over time: Evidence from Indo-European. *Language* 93.799–826.

WINTER, WERNER. 1984. Reconstructional comparative linguistics and the reconstruction of undocumented stages in the development of languages and language families. *Historical syntax*, ed. by Jacek Fisiak, 613–625. Berlin - New York: Mouton de Gruyter.

WITZLACK-MAKAREVICH, ALENA, and ILJA A. SERŽANT. 2018. Differential argument marking: Patterns of variation. *Diachrony of differential argument marking*, ed. by Ilja A. Seržant and Alena Witzlack-Makarevich, 1–40. Berlin: Language Science Press.

YANG, CHARLES D. 2000. Internal and external forces in language change. *Language Variation and Change* 12.231–50.

YANG, Z. 2014. *Molecular evolution: A statistical approach.* Oxford: Oxford University Press.

NOTES

[1]There are a number of terms for data of this type; in biology, it is common to refer to such data as a a multistate character (e.g. eye color) which can be realized as one of several traits (e.g. green eyes). We use the terms variable and value for the purpose of terminological and conceptual comparability with other work in quantitative linguistics (rather than biology), at times using the terms FEATURE and TRAIT interchangeably with value.

[2]Phylogenetic rate inference and reconstruction requires practitioners to define the PRIOR PROBABILITY of different values of a variable at the root of the phylogeny. A common practice in biology is to use the stationary probability of the CTM process, which gives the probability of the system taking a particular value as time approaches infinity. Felsenstein (2004:252) states that this prior is appropriate, but only if we assume that the model of evolution has been operating for a very long time. An alternative approach is to assume that the equal prior probability of each value at the root, or to treat the root prior as an unknown parameter to be inferred. The issue of how to treat the root prior is not widely discussed in phylogenetic linguists (many studies do not mention the issue at all), with some exceptions (Maurits & Griffiths 2014). In our main analyses, we follow other work (Cathcart et al. 2018; Blasi et al. 2019; Cathcart et al. 2020) in employing the stationary probability as the root prior. At the same time, because use of the stationary probability may bias our reconstructions, we run our models under two additional inference regimes, one using a uniform (i.e. equiprobable) root prior, and one where the root prior is treated as a parameter to be inferred. We find that results critical to the evaluation of our model against different traditional models are not affected by this choice (a full analysis of this issue is found in the supplementary material, Appendix E).
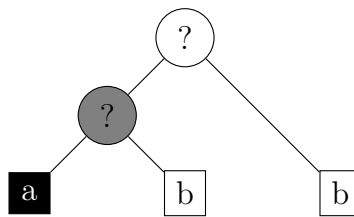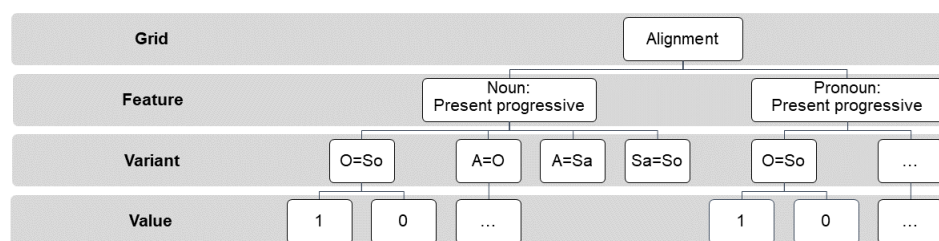
LIST OF FIGURES

Figure 1: Visualization of ancestral state reconstruction in a phylogenetic comparative model. The probability that a trait is present at a given node is estimated based on the evolutionary rates inferred for the trait, as well as the probability that the trait is present in descendant nodes.

Grid | Alignment

Feature | Noun: Present progressive | Pronoun: Present progressive

Variant | O=So | A=O | A=Sa | Sa=So | O=So | ...

Value | 1 | 0 | ... | 1 | 0 | ...

a.

Categorical feature Alignment| Noun: Present progressive	Variant O=So	Variant A=O	Variant A=Sa	Variant Sa=So	Trait
Categorical feature Alignment| Noun: Present progressive	Value 0	Value 0	Value 0	Value 1	Noun, Present progressive: Tripartite
	Value 0	Value 0	Value 1	Value 1	Noun, Present progressive: Nominative-accusative
	Value 1	Value 1	Value 1	Value 1	Noun, Present progressive: No marking

b.

Figure 2: Organization of typological data before (a) and after (b) recoding, illustrated for the categorical feature 'Alignment, Noun: Present progressive' (S32a).

Figure 3: Map of language locations in the data, distinguished by time period
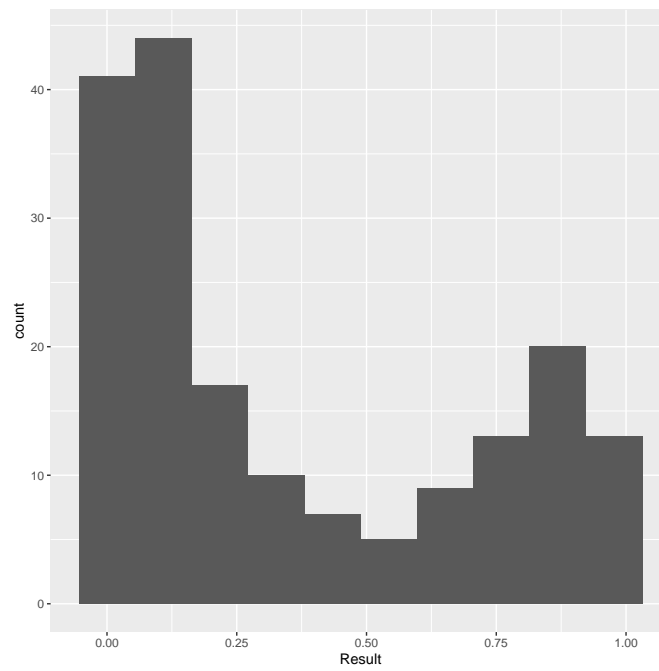
Figure 4: Histogram of reconstruction probabilities at the proto-language state (see S4). The lowest and (0.0–0.25) and the highest (0.75–1) probability ranges are more frequent than the intermediary ones (0.25–0.75).

Probability range	1–0.9	0.9–0.8	0.8–0.7	0.7–0.6	0.6–0.5	0.5–0.4	0.4–0.3	0.3–0.2	0.2–0.1	0.1–0.0
Pronouns										
Present Progressive	nom-acc									neutral
Simple Past			nom-acc						ergative	neutral, tripartite
Nouns										
Present Progressive				nom-acc			neutral			tripartite
Simple Past					nom-acc		neutral		ergative	tripartite

Figure 5: Overview of probability ranges at the proto-language state for alignment

Probability range	1–0.9	0.9–0.8	0.8–0.7	0.7–0.6	0.6–0.5	0.5–0.4	0.4–0.3	0.3–0.2	0.2–0.1	0.1–0.0
Gender on predicative adjective				+			−			
Masculine-feminine distinction					+		−			
Neuter			+						−	

Figure 6: Overview of probability ranges at the proto-language state for gender
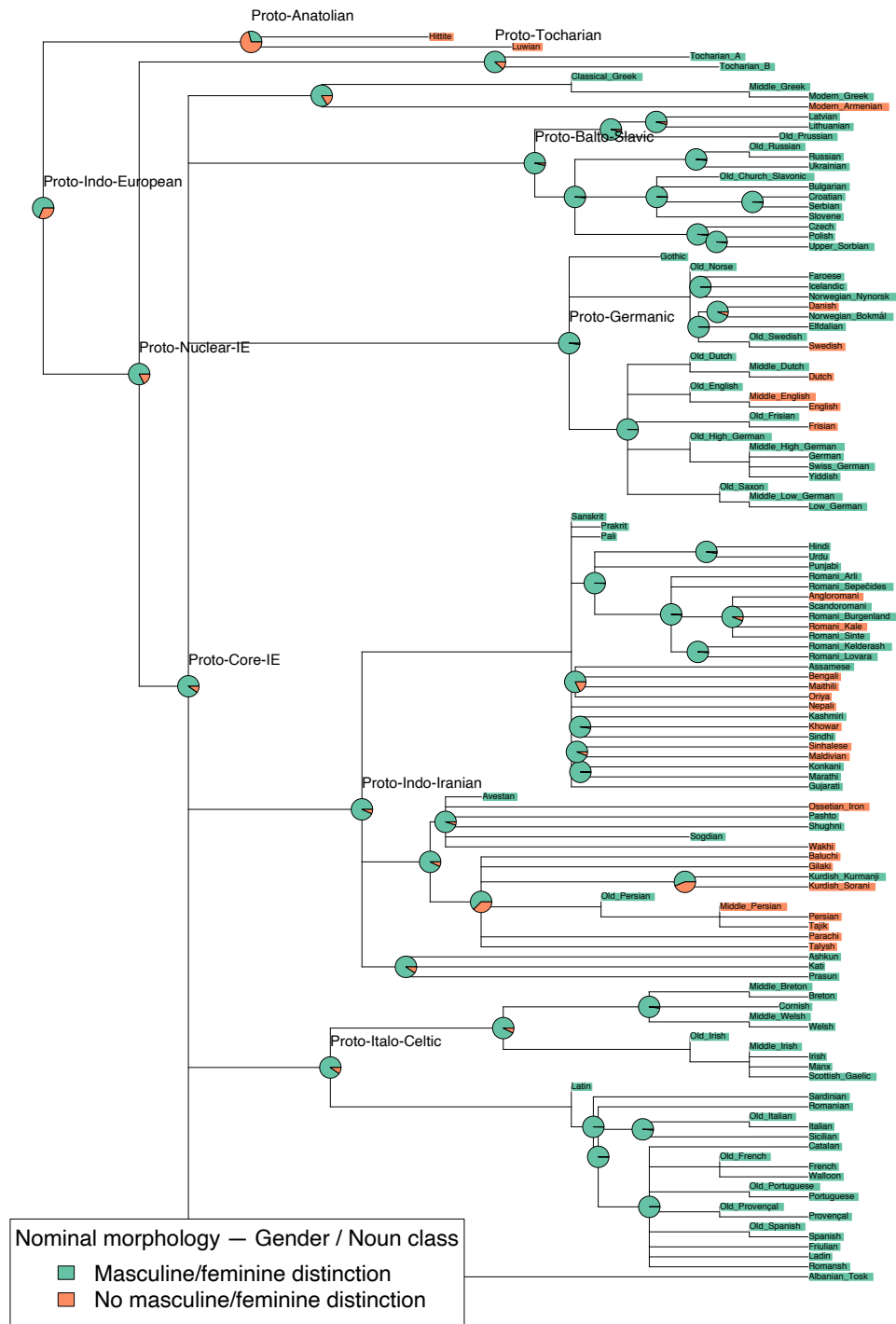
Figure 7: Maximum clade compatibility (MCC) tree with pie charts showing reconstructed probabilities of a masculine and feminine gender distinction at root and internal nodes of tree
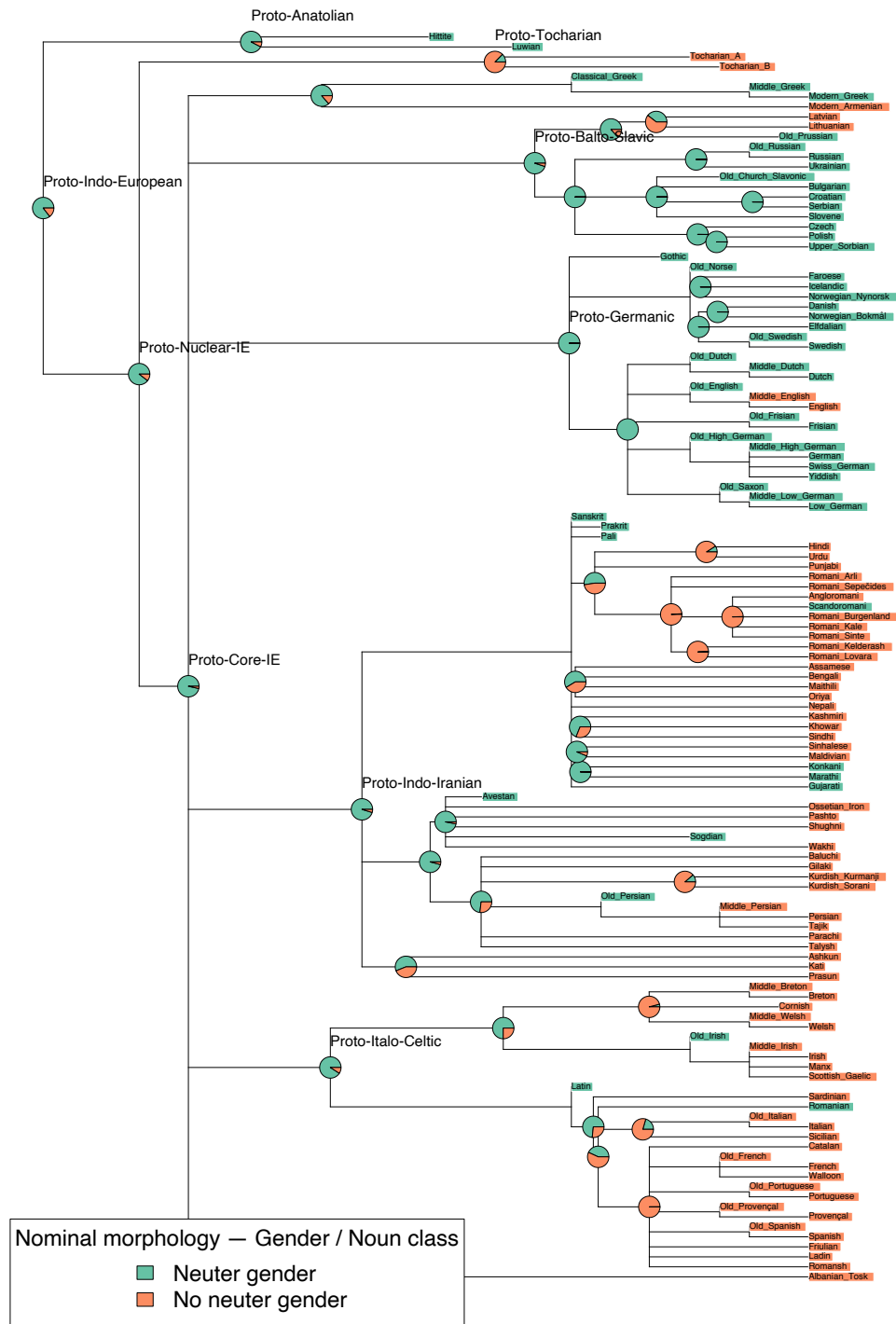
Figure 8: Maximum clade compatibility (MCC) tree with pie charts showing reconstructed probabilities of neuter gender at root and internal nodes of tree
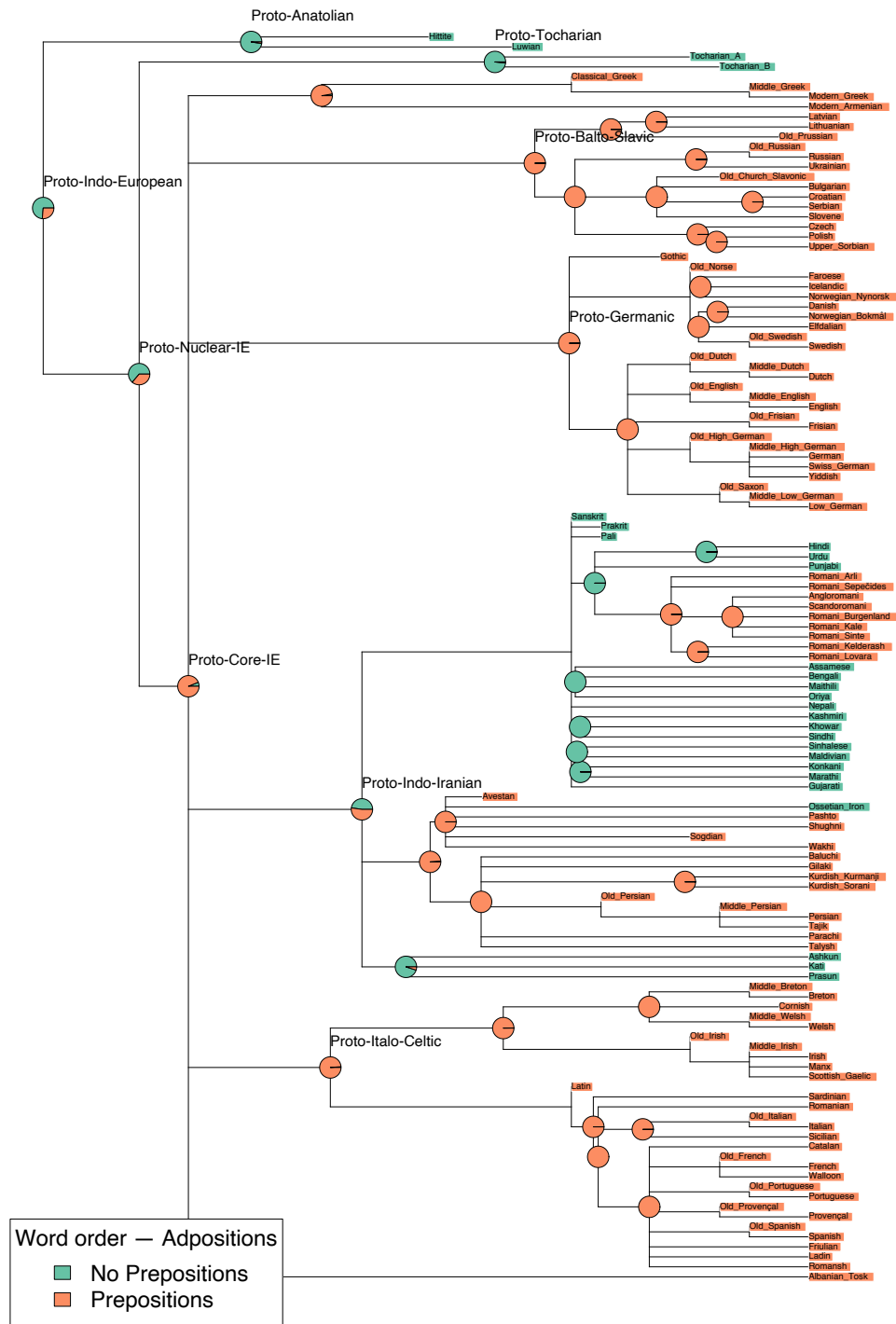
Figure 9: Maximum clade compatibility (MCC) tree with pie charts showing reconstructed probabilities of prepositions at root and internal nodes of tree

Probability range	1–0.9	0.9–0.8	0.8–0.7	0.7–0.6	0.6–0.5	0.5–0.4	0.4–0.3	0.3–0.2	0.2–0.1	0.1–0.0
Pronouns										
Agglutination: case	−									+
Agglutination: number	−									+
Difference A and O	+									−
Difference O/Dative					−	+				
Peripheral cases	+									−
More than 7 cases	−									+
Vocative	−									+
Nouns										
Agglutination: case			−					+		
Agglutination: number			−					+		
Difference A and O			+					−		
Genitive and/or dative						+gen/+dat	−gen/-dat			+gen/-dat, −dat/+gen
Peripheral cases	+									−
More than 7 cases				+			−			
Vocative		+							−	

Figure 10: Overview of probability ranges at the proto-language state for case

Probability range	1–0.9	0.9–0.8	0.8–0.7	0.7–0.6	0.6–0.5	0.5–0.4	0.4–0.3	0.3–0.2	0.2–0.1	0.1–0.0
Present progressive										
A agreement					full			syncretic		no
O agreement	no									syncretic, full
DAT agreement	no									syncretic, full
Simple past										
A agreement							syncretic	full	no	
O agreement			no							syncretic
DAT agreement		no								syncretic, full

Figure 11: Overview of probability ranges at the proto-language state for verbal agreement

Probability range	1–0.9	0.9–0.8	0.8–0.7	0.7–0.6	0.6–0.5	0.5–0.4	0.4–0.3	0.3–0.2	0.2–0.1	0.1–0.0
NP word order										
Adpositional word order		+postp	−prep					+prep	−postp	
Noun-possessor				poss-n			n-poss			
Noun-adjective		adj-n							n-adj	
Clause word order										
Main clause		SOV								VSO, V2, SVO
Subordinate clause		SOV							SVO	VSO, V2
Infinitive WO			OV						VO	irrelevant
Participle WO		O-part								part-O
WH-verb			−					+		
WH-initiality			+					−		
Noun-relative clause				noun-rel				rel-noun		irrelevant

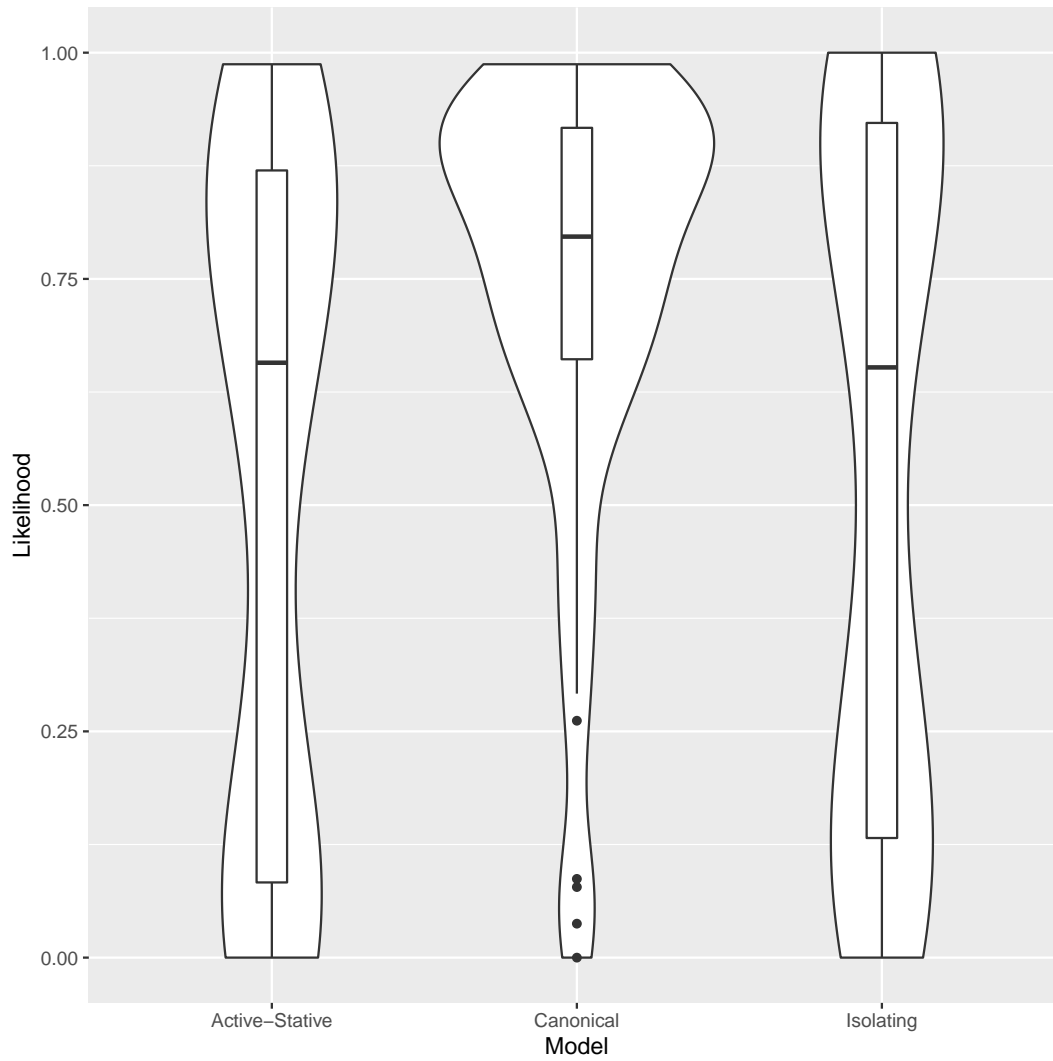Figure 12: Overview of probability ranges at the proto-language state for word order

Figure 13: Violin plot displaying likelihoods of reconstructions proposed by each comparative-historical model, labeled 'canonical', 'isolating' and 'active-stative', under the results produced by our model. Higher values signify greater agreement.
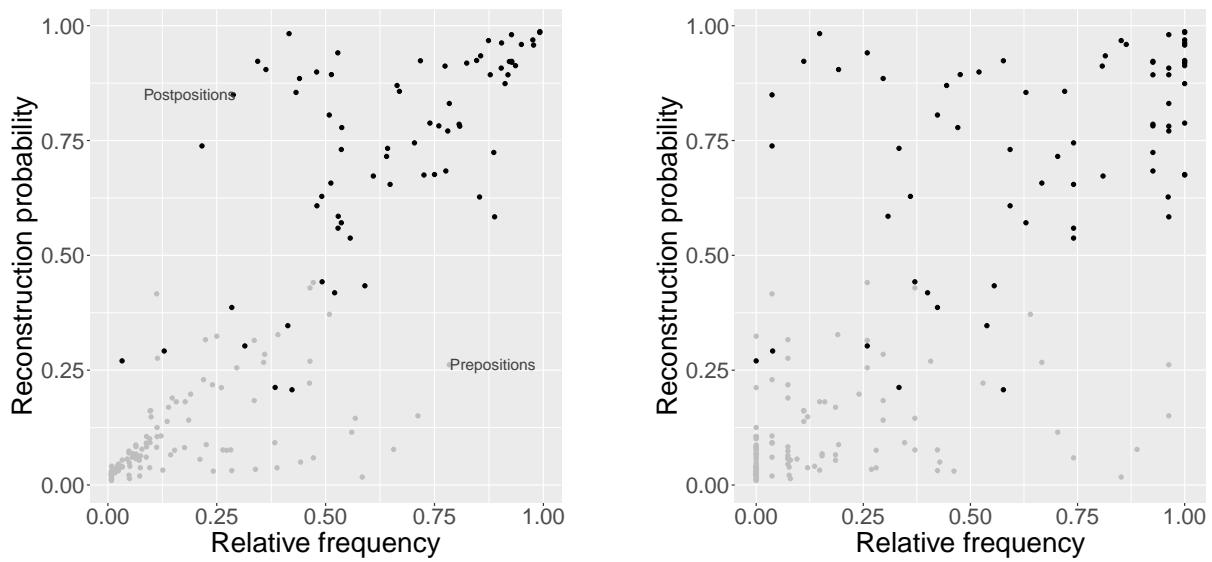
Figure 14: Reconstruction probabilities of features reconstructed with maximum probability to PIE, plotted according to their relative frequency across all languages (left) and ancestral languages (right) in the data set, with selected features annotated. Black dots represent 'winning' features reconstructed with maximum probability, while gray dots represent features not reconstructed with maximum probability.
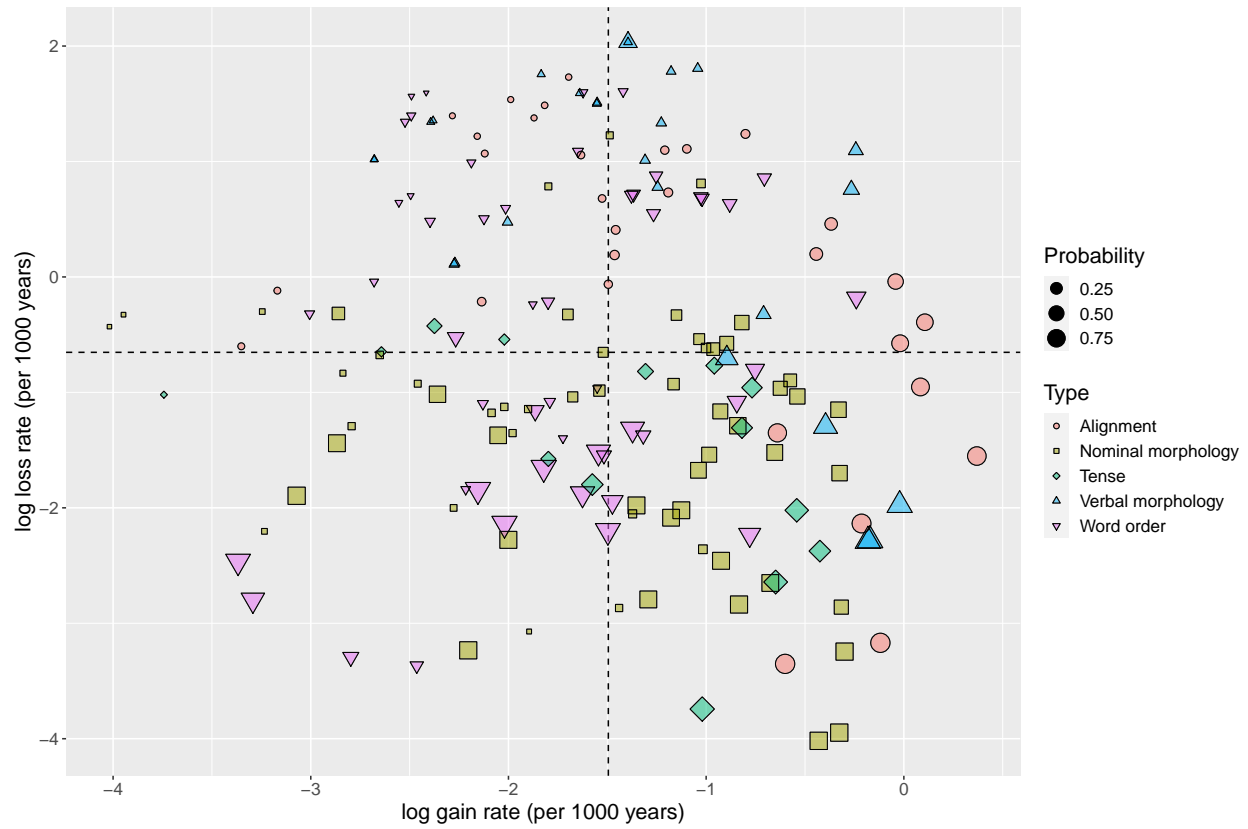
Figure 15: Features in the data set plotted according to their log gain and log loss rates. Shapes represent grammatical categories to which features belong; sizes represent the probability with which features are reconstructed to PIE.
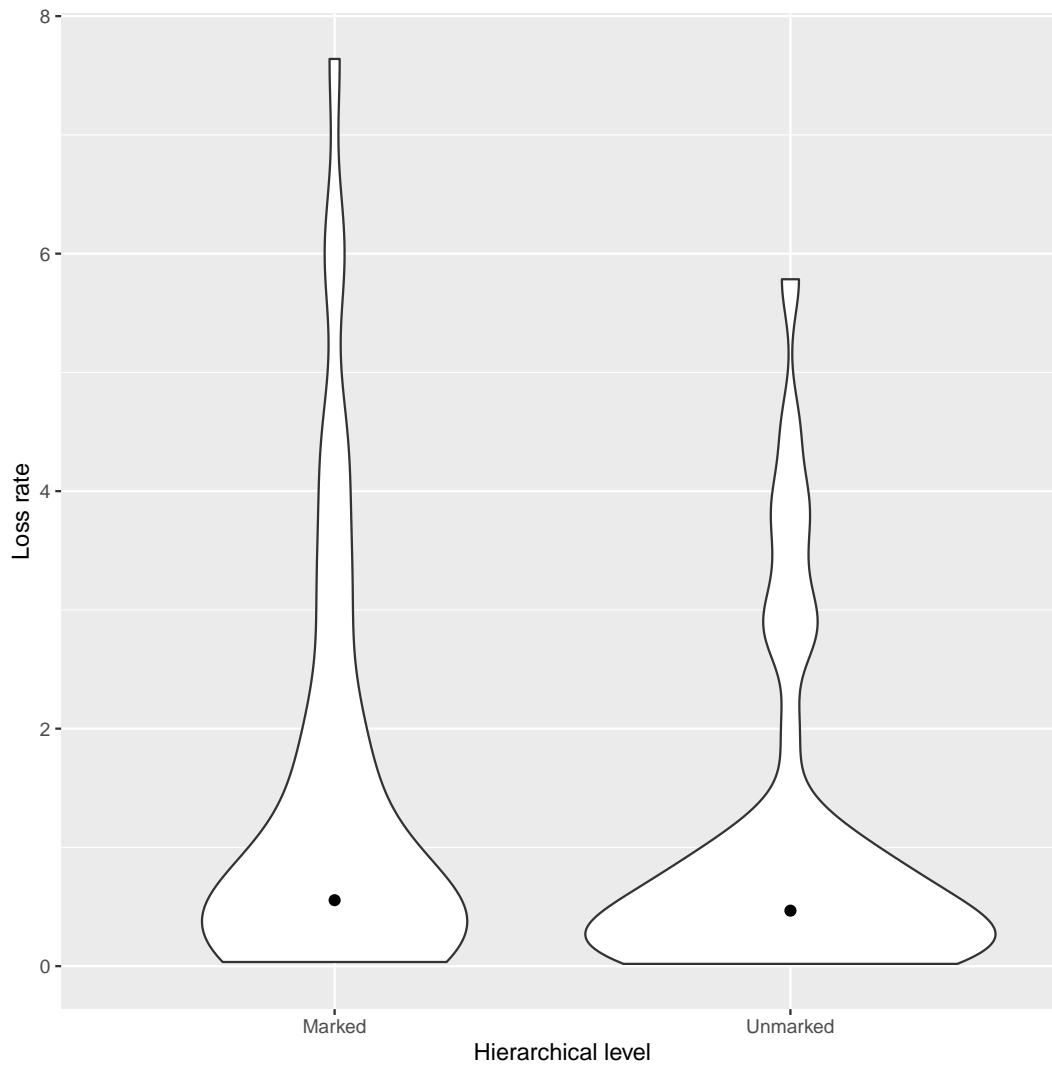
Figure 16: Distributions of loss rates for marked (more frequent) and unmarked (less frequent) features. Median values are indicated by dots. Unmarked (more frequent) traits are lost on average less frequently than marked (less frequent) ones.

List of Tables

Family	Type	Timeframe	Number
Indo-European	Archaic	2000 – 500 BCE	3
	Ancient	500 BCE – 500 CE	5
	Medieval	500 – 1500	29
	Modern	1500 – 2000	79
	Romani	1500 – 2000	9
TOTAL			**125**

Table 1: Number and type of languages in the data set of the current study (see S1)

Category	Hierarchy of features (unmarked/more frequent < marked/less frequent)
NP type	pronoun < noun
Tense	present < future
Tense	present < past
Grammatical relation	agent < object
Grammatical relation	agent/object < oblique
Gender	masculine/feminine < neuter

Table 2: Pairwise coded marking hierarchies in our data (see S6 for a complete list of coded relations), based on Croft (1990, 2003); Greenberg (1966); Greenberg & Haspelmath (2005).

	Masculine/Feminine	Neuter
Nominative	-s/-∅	-∅/-m
Accusative	-m	-∅/-m

Table 3: Markedness in the Proto-Indo-European case paradigm in the active-stative theory (Bauer 2000:45; Szemerényi 1989:169)

	1st person	3rd person M./F.	3rd person N.
Nominative	*ego	*so/*sa	*tod
Accusative	*me	*to	*tod

Table 4: Suppletion in the PIE pronominal paradigm (Bauer 2000:45; Szemerényi 1989:169)

Full agreement		1	0	0
No agreement		0	0	1
		Tocharian A *pälk-* 'shine' (Krause & Thomas 1960)	Gothic *bindan* 'bind' (Bammesberger 1986)	Swedish *sitta* 'sit'
Sg.	1	pälkäm	binda	sitter
	2	pälkät	bindis	sitter
	3	pälkäṣ	bindiþ	sitter
Pl.	1	pälkmäs	bindam	sitter
	2	pälkäc	bindiþ	sitter
	3	pälkiñc	bindand	sitter

Table 5: Coding patterns and example paradigms for A-agreement patterns, default paradigm of present progressive, in Tocharian A, Gothic, and Swedish

Gain rate	Loss rate	Interpretation	Reconstruction probability
High	High	Unstable trait in flux	Low/High
High	Low	Stable, attractor feature	High
Low	High	Recessive feature	Low
Low	Low	Infrequent but stable feature	Low/High

Table 6: Interpretation of behaviour patterns of traits depending on probability of presence at the proto-language state and gain/loss rate