Evolutionary dynamics of Indo-European alignment patterns

Gerd Carling and Chundra Cathcart
Lund University | University of Zurich

This paper employs phylogenetic modeling to reconstruct the alignment system of Indo-European. We use a data set of categorical morphosyntactic features, which take states such as 'nominative-accusative', 'active-stative', or 'ergative'. We analyze these characters with a standard Bayesian comparative phylogenetic method, inferring transition rates between character states on the basis of a phylogenetic representation of the languages in the data. Using these rates, we then reconstruct the probability of presence of traits at the root and nodes of Indo-European. We find that the most probable alignment system for Proto-Indo-European is a nominative-accusative system, with low probabilities of neutral marking and ergativity in the categories lower in grammatical hierarchies (nouns, past). Using a test of phylogenetic signal, we find that characters pertaining to categories higher in hierarchies show greater phylogenetic stability than categories lower in hierarchies. We examine our results in relation to theories of Proto-Indo-European alignment as well as to general typology.

Keywords: historical linguistics, evolutionary linguistics, Proto-Indo-European, linguistic typology, grammar evolution, alignment, syntactic reconstruction

1. Introduction

1.1 The study: An overview

The current study aims to investigate the Indo-European alignment system using a Bayesian comparative phylogenetic method. In doing so, we are interested in several aspects of alignment reconstruction and change. First, we aim to reconstruct the most probable system of the root of the tree, which represents the Indo-European proto-language. Second, we aim to reconstruct the evolutionary dynamics of traits in the Indo-European phylogenetic tree, in order to conclude

https://doi.org/10.1075/dia.19043.car | Published online: 23 July 2021
Diachronica 38:3 (2021), pp. 358–412. ISSN 0176-4225 | E-ISSN 1569-9714

whether general, typological correlations play a role in the dynamics of alignment evolution over time. Finally, we aim to observe whether evolutionary trends in the tree reflect the attested and reconstructed change in the Indo-European family.

Reconstruction of Indo-European alignment has a long history in the scientific literature; both within the discipline of comparative-historical linguistics as well as within the discipline of typology (see §1.3). An important aspect of this study is therefore to compare the result produced by our comparative phylogenetic model with different theories of alignment reconstruction. In previous literature on the topic, different models of reconstruction depend on the interpretation of attested language data, the reconstruction of paradigms, observed tendencies of change, as well as general typological implications. We discuss these arguments in the light of the results gained from our study. In addition, we observe the variable evolutionary dynamics of alignment traits. We discuss this variable behavior within the framework of general theories on grammatical hierarchies, in order to investigate whether there is a connection between general typological tendencies and the evolutionary behavior of alignment traits.

In this introduction, we describe the background of this study, providing an overview of the different models of reconstruction, which we label the comparative-historical, typological and comparative phylogenetic reconstruction methods (§1.2). We give a literature overview on different reconstructions of Proto-Indo-European alignment, including the supporting arguments for the various models, in §1.3. We then discuss the issue of grammatical or marking hierarchies and frequency in alignment, as a background to a discussion on general typological tendencies and their relevance to the evolutionary dynamics of traits (§1.4). In addition, we summarize, with respect to Indo-European specifically, different trends in alignment change in the family (§1.5). In §2, we first present our data, including targeted languages, the coding model of the original data set and the extraction and recoding model of the data used in the present paper (§2.1). Second, we describe the Bayesian phylogenetic comparative method, including reconstruction of ancient states (§2.2). Third, we describe the model used for assessing phylogenetic strength of characters (§2.3). In §3, we describe the results from various aspects, starting with an overview (§3.1). Subsequently, we discuss probabilities at internal nodes of the tree where no data are observed, including the proto-language state (§3.2), probabilities at the proto-language state in the light of grammatical hierarchies (§3.3), phylogenetic strength (§3.4), and a summary of results (§3.5). Section 4 contains a concluding discussion, where results are viewed in the light of general theories of language change as well as in contrast to various models of syntactic reconstruction.

1.2 Comparative, typological and phylogenetic models of reconstruction

The model of syntactic reconstruction introduced in the 19th century is mainly based on reconstruction on the basis of grammatical morphemes, where forms and meanings are systematized so that comparative sets of paradigms can be reconstructed for a proto-language. Grammatical morphemes, like lexemes, can be reconstructed by the linguistic comparative method, hereafter simply referred to as the comparative method. However, establishing the meaning and syntactic function of reconstructed morphemes is more problematic, mainly because regularity and directionality in syntactic change is difficult to establish with certainty (cf. Barðdal 2014). In general, the comparative method cannot be applied to the reconstruction of syntactic features which are not bound by morphology, such as word order (Harris & Campbell 1995; Harris 2008). Scholars take different positions on the reconstruction of syntactic features which are connected to morphology. Those who believe in a morphosyntactic reconstruction model argue that if a morpheme in combination with a syntactic function or a construction can be projected to the proto-language by the comparative method and has survived in a majority of languages, then there is reason to reconstruct it to the proto-language (Barðdal & Eythórsson 2012; Barðdal 2014; Campbell & Harris 2002: 615). Opponents to syntactic reconstruction often invoke the correspondence problem: we may reconstruct a pattern to an ancestor 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: 363–367; Walkden 2013). An additional problem is grammaticalization: features may continue to exist in languages, whereas the form is lost and recreated by grammaticalization (Hopper & Traugott 2003: 39–70). There are numerous examples, one being the case paradigm of Tocharian, which was almost entirely lost due to phonological change but which recreated several Indo-European cases, such as ablative, locative, dative, and genitive, by means of grammaticalization (Carling 2012).

A complementary method of syntactic reconstruction emerges from Joseph Greenberg's typological model (Greenberg 1966). This method accounts for language-internal implicational dependencies between features ("universals"). If these implications can be identified, based on a mass-based study of preferably unrelated languages, then these observations can be used as an argument for reconstructing a probable typology of a proto-language (Bickel 2007; Lehmann 1973, 1974; Nichols 1992, 1995, 1998). A problem of this model is the possible occurrence of language-internal conflicts with respect to "universal" dependencies. The Proto-Indo-European word order controversy is an example of this: a reconstruction of Proto-Indo-European comes out as inconsistent if it is based

on the ancient languages (Friedrich 1975; Lehmann 1974; Watkins 1976; Winter 1984). In accordance with Greenbergian universals (Universal 24), an OV language is supposed to have Relative clause – Noun word order (Greenberg 1963; Lehmann 1973, 1974), which is the most common type in Hittite and Latin. However, in Sanskrit and Homeric Greek the most common order is Noun – Relative clause, which is inconsistent with OV, meaning that Indo-European has to be either OV or VO (following a consistency model) or reconstructed with an inconsistent order (Clackson 2007: 171–176; Hock 2013).

Diachronic typology combines reconstruction by the comparative method with typology. This model borrows methods and aims from cross-linguistic typology, and applies them to diachronic material, observing the typology of change rather than the typology of states (Hendery 2012: 2–6). However, the problem of irregularity of patterns with respect to typological consistency, as well as the problem of reconstruction of syntax without cognacy, are continuing issues also in diachronic typology (Barðdal & Eythórsson 2012; Barðdal 2014; Viti 2015: 16–17).

Computational phylogenetic comparative models (Bowern 2018; Calude & Verkerk 2016; Carling & Cathcart 2021; Jäger 2019; da Silva & Tehrani 2016) differ from the historical-comparative, typological and diachronic typological models. Most basic computational phylogenetic comparative models do not account for morphosyntactic reconstruction or assumptions regarding implicational dependencies between features. Rather, they use comparative data to infer information regarding the diachronic dynamics of individual features, and use these inferences to reconstruct most probable features for unobserved speech varieties (i.e., proto-languages) in a linguistic phylogeny. The probability of the presence of a morphological or syntactic feature at an ancestral proto-language is estimated on the basis of three key ingredients: (1) a representation of the phylogenetic relationship between the languages under study; (2) the distribution of the feature among the daughter languages; and (3) an evolutionary model of change (Cathcart et al. 2018; Dunn 2014; Greenhill et al. 2010; Maurits & Griffiths 2014). The phylogenetic representation can consist of a handcrafted tree which incorporates subgroups based on the traditional comparative-historical method, or a sample of trees inferred from unrelated linguistic data (e.g., lexical data), using Bayesian methods from computational biology (Bouckaert et al. 2012; Chang et al. 2015; Gray & Atkinson 2003). On the basis of the presupposed phylogenetic representation and the data observed for the languages in the tree (point 2 above), the model estimates the average rate of transition between traits of a categorical feature (e.g., SOV > SVO), under the assumption that the feature evolves according to a Continuous-time Markov process. This stochastic process can be used to simulate changes between states over continuous time intervals (point 3 above). For this type of model, a maximum likelihood estimate of the rates can be obtained;

alternatively, we can sample from the posterior distributions of the rates via a Markov chain Monte Carlo or a similar technique for estimating parameter values for Bayesian models. Once evolutionary rates have been estimated, the probability of the presence of the feature in question can be estimated for internal nodes of the tree, including the root — the node ancestral to all of the attested languages in the sample (Felsenstein 2004; Yang 2014) (see further §2.1).

1.3 Indo-European alignment – reconstructions by comparative syntax

Reconstruction of alignment systems has a long history in Indo-European comparative syntax, ultimately going back to the works of the Neogrammarians in the late 19th century. In the 20th century, several alternative theories for Proto-Indo-European alignment reconstruction have been proposed. The Neogrammarians, e.g., Delbrück and Brugmann (Delbrück 1893, 1897, 1900), reconstruct an alignment system that is very close to the system found in most ancient Indo-European languages with the exception of Anatolian (which was not known to them). The alignment system described here is a completely nominative-accusative system, where the nominative codes the first argument (S/A) of intransitive and transitive verbs, and the accusative codes the second argument (O) (Delbrück 1893: 360–400; Meier-Brügger et al. 2010: 401–404). However, even before the decipherment of Hittite in 1915 (Hrozný 1915), an alternative model of ergativity for Proto-Indo-European is proposed by Uhlenbeck (1901). This idea is based on the internal paradigmatic marking distribution in Proto-Indo-European, with the reconstruction of a case marking of *-s for transitive subject of animates (ergative), *-m for transitive objects of animates (absolutive), and *-Ø for inanimates (absolutive) (Pooth et al. 2019: 246–248). The theory is later continued and developed by Vaillant (1936) and Soviet scholars of the 1970s (Gamkrelidze et al. 1995; Klimov 1973a, 1973b, 1974), who take the reconstruction one step further and suggest an active-stative alignment model. In this theory, the verb has no inherent transitivity (as in nominative-accusative and ergative models) and the alignment marking is based entirely on the semantics of the verbal core. The active-stative theory is continued by western scholars, such as Lehmann (1989), and recurs in a reviewed form in, e.g., Barðdal & Eythórsson (2009, 2012), Bauer (2000), Drinka (1999), Matasović (2013), Pooth et al. (2019) and Schmidt (1979). The active-stative and ergative theories, even though they differ in claims and rigidity, build on two main arguments. First, they focus on paradigmatic distinctions, which can be reconstructed for the proto-language on morphological grounds. The semantic and functional evaluation of these distinctions are identified as typologically frequent in languages of other types than the predominant Indo-European ones, i.e., nominative-accusative system, and hence support an alternative system of the

proto-language, either ergative or active-stative. Second, they point at the existence of typological properties in ancient Indo-European languages, in particular Anatolian, but also in, e.g., Latin, which are proto-typical for active languages. These are seen as residuals of a pre-state that was entirely active-stative or ergative (i.e., Proto-Indo-European). Over time, the diachronic development from Proto-Indo-European to Anatolian and Early Italic resulted in a dominance of nominative-accusative structure.

The diachronic typology of the reconstructed active-stative or ergative paradigms, i.e., how they develop from active-stative via ergativity into nominative-accusative (or from ergative to nominative-accusative) in a stratified Proto-Indo-European language and into sub-branches, is a vital part of the argument of active-stative theories (Schmalsteig 1981). In any case, an important part of the argument is that this transition was still in progress in the ancient branches, such as Anatolian and Italic (Bauer 2000). Therefore, at the core of the active-stative and ergative theories we find the identification of properties of Proto-Indo-European which are observed to correlate with active-stative and ergative alignment (for a more detailed description, see Matasović (2013), together with references to relevant literature).

The most important arguments for alternative (i.e., ergative or active-stative) alignment theories are as follows:

1. THE ABSENCE OF A VERB FOR 'HAVE' IN PROTO-INDO-EUROPEAN AND THE USE OF A MIHI EST-CONSTRUCTION (Gamkrelidze et al. 1995: 250; Klimov 1973a: 217; Lehmann 1989: 115f). This is a property which is frequently found in active-stative and ergative languages, but it is also a common feature of languages of any type.
2. AN ASSUMED DISTINCTION BETWEEN ANIMATE – INANIMATE IN THE PROTO-INDO-EUROPEAN NOMINAL PARADIGM. In the Proto-Indo-European noun, there is a *-os/*-om distinction in nominative/accusative of masculine/feminine (animate), whereas neuter (ancient inanimate) has an ending *-Ø/*-m, which makes no distinction between nominative and accusative (see Table 1). The distinction, which is mirrored by suppletion in the pronominal paradigm (Table 2), constitutes a core both to the ergative and active-stative theories (Drinka 1999; Gamkrelidze et al. 1995: 245ff.; Kortlandt 1983; Pooth et al. 2019; Uhlenbeck 1901; Vaillant 1936). In addition, the distinction is present in the reconstructed noun class paradigm, which later emerges as a gender system (Villar 1984). An important argument here is the lack of gender and distinction between animate and inanimate in Anatolian (Luraghi 2011; Villar 1984).

3. AN UNMARKED SUBJECT CASE AGAINST A MARKED OBJECT. This is another aspect of the reconstructed paradigm, which is seen as an indication of an ergative or active-stative Proto-Indo-European state. The marked *s-nominative (in the masculine/feminine) is supposed to reflect an ergative case, against an unmarked absolutive (Bauer 2000: 44–46; Martinet 1962).
4. THE OCCURRENCE OF OBLIQUE SUBJECTS IN PROTO-INDO-EUROPEAN. Oblique subjects, or non-canonical (quirky) case marking, occur in several Indo-European languages, such as Germanic or Latin (Cennamo 2009; Matasović 2013). By the syntactic reconstruction model of Construction Grammar these are traced back to Proto-Indo-European (Barðdal & Eythórsson 2009, 2012). Non-canonical marking, in particular with stative verbs, is frequent in active-stative and ergative languages and is therefore seen as an indication of an ergative system and a preceding active-stative system (Barðdal & Eythórsson 2009; Matasović 2013; Pooth et al. 2019).
5. THE PROTO-INDO-EUROPEAN VERBAL PARADIGM. Proto-Indo-European had a binary paradigm, with double setups of endings, which is seen as evidence in favor of active-stative structure in the proto-language (Gamkrelidze et al. 1995: 254ff.). Most important is the active vs. middle-passive paradigms, where the middle-passive is supposed to be related to the perfect tense by the endings (Lehmann 1989). In the active-stative theory, the *-mi (active), and *-h_2e (inactive) conjugation mark a semantic alignment system on the verb, and this is preserved in the *-mi* and *-ḫi* paradigm setup in Anatolian (Gamkrelidze et al. 1995: 254–276; Lehmann 1989; Meiser 2009). An important issue is the interpretation of verbal semantics of the *-mi* and *-ḫi* paradigms of Anatolian (Jasanoff 1978), the active and middle-passive voice (Luraghi 2012), and transitivity split in Hittite, visible in auxiliary formations as well as in clitics (Garrett 1996).
6. ALIENABLE/INALIENABLE POSSESSION AND INCLUSIVE/EXCLUSIVE PRONOUNS. This theory assumes that Proto-Indo-European had a distinction between alienable and inalienable possession, which is reflected in Hittite (Gamkrelidze et al. 1995: 251–252), as well as inclusive and exclusive pronouns (Gamkrelidze et al. 1995: 253–254), reflected among others in the double forms of the plural of the pronoun (*wei-/*mes-). This distribution is considered to be an argument in favor of active-stative structure. In general, the arguments are considered to be weak (Matasović 2011).

As mentioned before, important arguments in favor of both the active-stative and ergative theories are supposed remnants of systems of this type in ancient Indo-European languages, most importantly the Anatolian *-mi* and *-ḫi* paradigm setup

as well as the presence of an ergative case marker (Garrett 1990, 1996) and the lack of gender (Matasović 2004; Villar 1984).

The active-stative and ergative theories are not accepted by all scholars (cf. Clackson 2007: 176–180). Most importantly, both theories imply a typologically ideal state, which is not preserved as such in any of the Indo-European branches. The described typological tendencies, which imply ergativity or active-stative typology (absence of 'have', nominal classification based on animacy, alienable/inalienable possession, inclusive/exclusive pronouns), are not unique to active-stative or ergative systems, but are also found frequently in nominative-accusative systems (Matasović 2011: 2–3). Even languages with ergative systems, such as Hittite, are typically split-ergative rather than completely ergative (Goedegebuure 2013). The occurrence of ergativity in these languages typically reflects animacy hierarchies, where ergativity is restricted to the nominal system, whereas the pronominal system shows a nominative-accusative tendency (Silverstein 1976).

Several of the functional reconstructions of the ergative and active-stative paradigmatic reconstructions are weak. For instance, the formal contrast in Hittite between the *-mi* and *-ḫi* conjugations is not reflected by a systematic difference in meaning (Jasanoff 2003: 1–40), which makes it difficult to propose a semantic core for the Proto-Indo-European verb. Finally, yet importantly, the observed occurrences in ancient languages may be innovations under areal pressure from non-Indo-European languages, such as the Hittite ergative case (Clackson 2007; Garrett 1990: 178). These arguments lead some scholars to refute the active-stative and ergative theories purely on typological grounds (Rumsey 1987).

Table 1. Markedness in the Proto-Indo-European case paradigm, underlying the ergative and active-stative theories (Bauer 2000: 45; Szemerényi 1989: 169)

	Masculine/Feminine	**Neuter**
Nominative	*-Ø/*-s	*-Ø/*-m
Accusative	*-m	*-Ø/*-m

Table 2. Suppletion in the Proto-Indo-European pronominal paradigm (Bauer 2000: 45; Szemerényi 1989: 169)

	1st person	**3rd person Masculine/Feminine**	**3rd person Neuter**
Nominative	*ego	*so/*sa	*tod
Accusative	*me	*to	*tod

1.4 Marking and animacy hierarchies in alignment

In languages with ergative or active-stative systems, referential marking hierarchies of grammatical relations, in particular with respect to animacy, influence the alignment coding (Bickel 2008; Comrie 1981: 117ff.; Dixon 1994: 83ff.; Silverstein 1976). Few attested systems are entirely ergative or active-stative without accusative tendencies, and even accusative languages, such as Hittite, may display ergative tendencies (Garrett 1990; Goedegebuure 2013; Lopuhaä-Zwakenberg 2019). Marking hierarchies play a role in ergative and active-stative theories of Proto-Indo-European (Bauer 2000: 44–47), in which synchronic patterns are seen as possible carriers of earlier stage residues of alternative (ergative, active-stative) alignment. In general, semantic qualities as well as grammatical relations can underlie a marking hierarchy. Marking hierarchies are based on semantically conditioned factors, where grammatical categories higher in the animacy scale are also higher in the hierarchy (Comrie 1981: 120–129), but marking hierarchies are also correlated with frequency and economy of feature types (Croft 2003: 123; Witzlack-Makarevich & Seržant 2018). Hierarchies can be observed by means of argument marking synchronically, such as subject and object, present and past. It is relatively common for a language to confirm marking hierarchies as a variation rather than as a default pattern (Verbeke & De Clercq 2016; Witzlack-Makarevich & Seržant 2018 for an overview) (see Table 3 for a list of marking hierarchies of relevance to alignment).

Table 3. Marking hierarchies of grammatical and semantic properties observed in the literature. After (Bickel 2008; Comrie 1981; Croft 2003; Dixon 1979; Malchukov 2015; Witzlack-Makarevich & Seržant 2018)

Property type	Hierarchy
Person	1st person < 2nd person < 3rd person < proper nouns
Animacy	human < animate < inanimate
Uniqueness	proper nouns < common nouns
Number	singular < plural < dual
TAM	future < present < imperfect < aorist < perfect

Generally, the underlying principle of marking hierarchies is a tendency towards A (transitive subject) marking in the higher-ranking categories and a tendency towards O (transitive object) marking in the lower-ranking categories (Dixon 1994: 95ff.). This is a tendency which nevertheless has a substantial number of exceptions (Bickel 2008). Since grammatical hierarchies play a role in synchronic states of languages as well as in diachronic typology, we assume that they

also play a role in the evolutionary dynamics of language change. We will return to the concept of grammatical hierarchies from the perspective of evolutionary reconstruction and the relevance to our results in §3.3 and §3.4.

1.5 General trends in the family

An important issue in alignment change is the principle of transition from one system to another, a phenomenon that we will consider from an evolutionary perspective. Diachronic change in alignment systems is a process that we are relatively well informed about: it has been observed historically and can be partly or completely reconstructed in several languages. Typically, a change from nominative-accusative to ergative and the reverse process, from ergative to nominative-accusative, involves a syntactic reinterpretation: in the case of accusative → ergative the process frequently involves a reinterpretation of the passive, in the case of ergative → accusative the process involves a reinterpretation of the antipassive. Further innovations include the emergence of periphrastic systems, emergence of new cases, or a generalization of marking patterns from one construction to others. There is a rich literature on the issue of morphosyntactic causes and grammaticalization pathways for the emergence of ergativity, based on observations from different languages (Dahl & Stroński 2016a; Dahl 2016; Dench 1982; Dixon 1994: 182–206; McGregor 2009: 498–500; Verbeke & De Cuypere 2009).

Looking at Indo-European alignment, there are two main transition tendencies, which have been observed from the earliest attested stages of a branch and further into the daughter languages. First, we have the alignment transition from the Old-Indo-Aryan nominative-accusative system into the (split) ergative and tripartite systems of New Indo-Aryan. Even though the pathways of change are not entirely clear, an important issue is the reinterpretation of the passive participle in *-ta* (which agrees with O in transitive verbs and S by transitive verbs) into an ergative construction (Bubenik 2016; Dahl & Stroński 2016b; Verbeke & De Cuypere 2009). Second, we have the transition from a nominative-accusative system into a neutral system, which occurred in languages of the west (e.g., English, Celtic languages) but also occasionally in eastern languages (e.g., Pashto) (Brugman & David 2014; Carling 2019: 31–51; Ronan 2011). The process of this change, starting with a neutralization of the nominative – accusative distinction, can for instance be observed in Late Latin (Cennamo 2009). We will return to the issue of system transition from an evolutionary perspective and the relevance to our results in §3.3–3.4.

2. Data, model and method

2.1 Data: Languages and coding models

The data for the current study are extracted from the database DiACL (Carling 2017), subset DiACL/ Typology/ Eurasia (Carling et al. 2018), which contains typological properties of the Eurasian linguistic area. The coded features of the data describe alignment in terms of comparative concepts, following a "Comrian" model of generalization (Haspelmath 2011), similar to WALS (Dryer & Haspelmath 2013). However, features are not described as properties, e.g., 'ergative', 'nominative-accusative', etc., rather, they are defined as hierarchically organized features, which are Boolean (1/0) at the lowest level. This is more similar to the AUTOTYP and the multivariate model (Bickel & Nichols 2002; Bickel 2010). For the current paper, we extract the hierarchical and binary coding of the data, which we recode into multistate characters (combinations of 1/0), labeled as categorical features (see further below). The dataset contains data from most ancient Indo-European languages, something which increases the reliability of the comparative phylogenetic reconstruction (see further §2.2). We use all 126 Indo-European languages available in the dataset (see Appendix 1) and extract features pertaining to the domain "Alignment" in the database (see Appendix 2 for a complete list). After the recoding, the data consists of multistate characters, labeled as categorical features, with variants, labeled traits (see Appendix 3). The raw dataset is openly available and can be downloaded from the DiACL database (Carling 2017).[1]

For the coding of alignment we follow the model of the data, which codes properties of grammatical relations according to how languages mark S (intransitive subject), A (transitive subject), and O (transitive object). The model defines an alignment matrix based on argument properties (Bickel 2011; Dixon 1994, 2010), which identifies the basic aspects of accusative, ergative, active, and tripartite marking from the core arguments (Dahl & Stroński 2016b). The basic alignment matrix of the data lists four correlations, A=O, A=Sa, O=So, and Sa=So (Dixon 1994: 23–48, 2010: 126ff.; Haspelmath 2011), which are used to describe the coding relations of A, S and O with verbs of various transitivity (intransitive/ transitive) or semantic (active/stative, unergative/unaccusative) types (Sa/So) (Table 4). These correlations are tested for differences in marking related to the word class of the first argument (noun/pronoun) and the tense of the verbal predicate (present progressive/past tense) (Table 5). A second coding variant tests

1. The recoded set as well as code used in this paper can be accessed via https://github.com/chundrac/evo-dyn-ie-align and https://zenodo.org/record/4118097.

alignment properties with verbs, targeting marking differences in verbal agreement with reference to S, A, and O (Table 6, Appendix 2).

The state combinations represent different alignment configurations, such as accusative, tripartite, ergative, active, or neutral marking (Table 7).

Table 4. Basic correlations in the data for defining alignment systems

Variant	Explanation (for marking of A, S, and O)
A=O	The subject (A) of a transitive verb bears the same marking as the object (O) of a transitive verb.
A=Sa	The subject (A) of a transitive-active verb bears the same marking as the subject (S) of an intransitive-active verb.
Sa=So	The subject (S) of an intransitive-active verb bears the same marking as the subject (S) of an intransitive-stative verb.
So=O	The subject (S) of an intransitive-stative verb bears the same marking as the object (O) of a transitive-active verb.

Table 5. Model for coding of basic alignment types (by a matrix of properties) against word class of the first argument and tense of the predicate

Matrix:	With	Word class of first argument:	in	Tense:
A=O		Noun		Present progressive
A=Sa		Pronoun		Simple past
O=So				
Sa=So				

As mentioned above, we observe combinations of 1 and 0 as states of a multistate character, i.e., a categorical feature, in our data. To facilitate reading, we define these states as traits, to which we give a unique ID and a descriptive label, such as "Noun, Present progressive: Tripartite" (Table 7, Appendix 3). Nothing prevents the occurrence of illogical or impossible combinations, e.g., A=O and A=Sa. Whenever illicit traits are found in the data or reconstructions, they turn out to be the result of mistakes in the coding. The illicit combinations at reconstructed states disappear as soon as we clean the coding in the attested languages.

In general, categorical features in attested languages can be defined by a relatively limited number of traits. The number of licit combinations, seen from a global perspective, could potentially be larger than in our sample. As an example, active-stative traits, which of course are perfectly licit, are absent in our data set. This has a straightforward explanation: we have extracted languages belonging to the Indo-European family, and the typological patterns reflect the attested

Table 6. List of features of the alignment subsection (see Appendix 2 for the full set of binary feature variants)

Categorical feature	Explanation
Compare PROG-PAST	What is the marking relation between subject and object in present progressive and simple past?
Noun: Present progressive	In present progressive: how is the marking of subject and object of nouns realized?
Noun: Simple past	In simple past: how is the marking of subject and object of nouns realized?
Pronoun: Present progressive	In present progressive: how is the marking of subject and object of pronouns realized?
Pronoun: Simple past	In present progressive: how is the marking of subject and object of pronouns realized?
Reflexive pronoun in transitive clause	What is the alignment of reflexive pronouns?
Verb: Present progressive	In present progressive, how is alignment realized on the verb?
Verb: Simple past	In simple past, how is alignment realized on the verb?

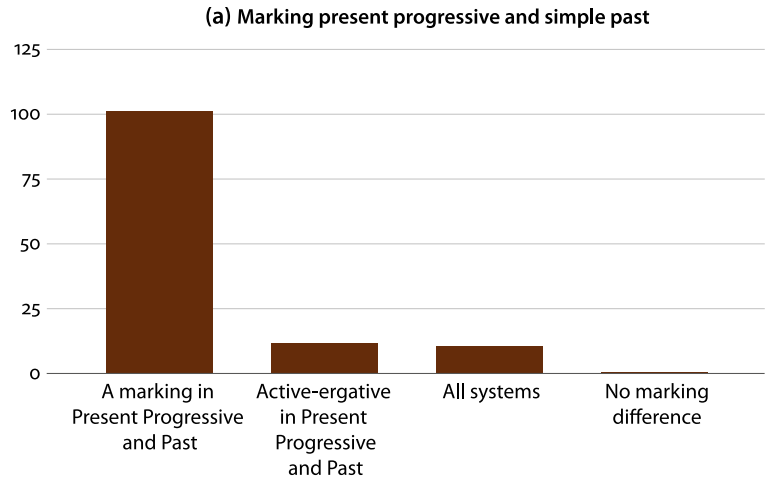
languages of this family. Figure 1 gives an overview of the distribution of the distribution of features and characters analyzed in this paper.

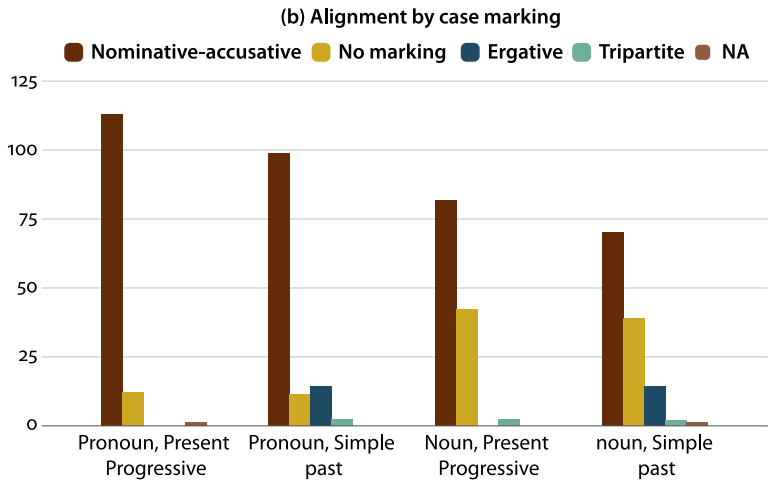
Table 7. Overview of state combinations of the data and the trait labels for a multistate character "Alignment type"

Type	A=O	A=Sa	O=So	Sa=So	Alignment type
(a)	0	1	0	1	nominative-accusative
(b)	1	1	1	1	neutral marking
(c)	0	0	0	1	tripartite
(d)	0	0	1	1	ergative
(e)	0	1	1	0	active

2.2 The Bayesian phylogenetic comparative reconstruction model

For each multistate character, the likelihood of a set of evolutionary transition rates between each pair of states under a given tree topology and observed data can be computed according to Felsenstein's Pruning Algorithm (Felsenstein 1981). We use a tree sample, which can be seen in Figure 2, that is generated as follows: we assume a fixed topology that agrees with received wisdom (Chang et al. 2015),


(a) Marking present progressive and simple past
125
100
75
50
25
0
A marking in Present Progressive and Past
Active-ergative in Present Progressive and Past
All systems
No marking difference


(b) Alignment by case marking
Nominative-accusative
No marking
Ergative
Tripartite
NA
125
100
75
50
25
0
Pronoun, Present Progressive
Pronoun, Simple past
Noun, Present Progressive
noun, Simple past

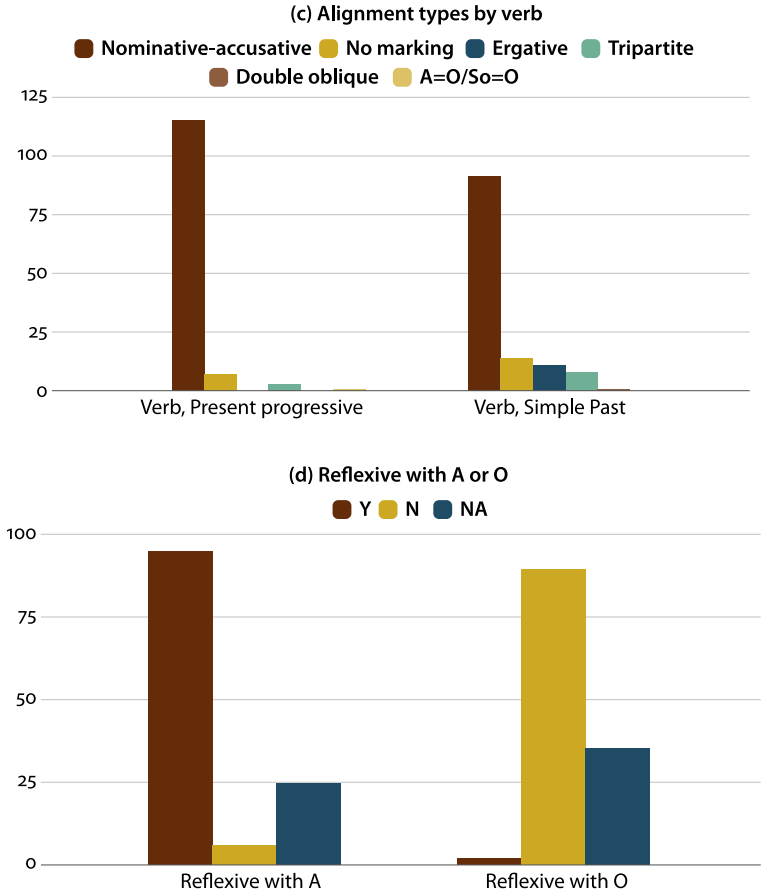


Figure 1. Bar plots summarizing the occurrence (in absolute numbers) of various trait types of some categorical features in the data, including Marking of Present progressive and Simple past (a), Alignment by case marking (b), Alignment by Verb (c), and Reflexive with A or O (d)

and uniformly sample branch lengths from chronologically realistic intervals. This yields a tree with a root age uniformly distributed between 7000 and 6000 years BP. We scale branch lengths by dividing them by 1000.

We infer transition rates using the No-U-Turn Sampler of RStan (Carpenter et al. 2017) to infer the posterior distributions of these rates for each tree in the sample. We run the sampler for 2000 iterations and discard the first half of samples as burn in; all parameters converge according to the potential scale reduction factor (Gelman & Rubin 1992). We aggregate these posterior samples over the sample of trees. We place a Uniform (0,5) prior over rates of change, constrain-

ing our model such that transitions between two states do not occur more than 5 times within a given 1000 year period.

We estimate the probability of a given character state (i.e., a trait) at the root of the tree by randomly drawing evolutionary rates from the posterior sample, iteratively sampling a state at the root (Bollback 2006; Huelsenbeck et al. 2003; Nielsen 2002), and normalizing the counts for each sampled state to yield probabilities between 0 and 1. The results of the Bayesian evolutionary reconstruction are in the form of a probability of presence (between 0 and 1) of a state within a categorical feature at the root and nodes of the tree (see Appendix 3 and Figure 3). We evaluate the results in §3.3.

In most cases, there is a clear result, with a high probability (0.70–1.00) for a specific state, and a low probability (0.10–0.30) for the remaining states of a multistate character. In other instances, there is high uncertainty across reconstruction probabilities, indicating that no state in the character can be securely reconstructed.

2.3 The model for assessing phylogenetic stability across characters

In addition to the reconstruction process described in the previous section, we assess the phylogenetic stability of each character (e.g., categorical feature) in our data set using the method of Borges et al. (2018), which provides an estimation of the uncertainty of reconstructed states at internal nodes of the tree (cf. also Zhou & Bowern 2015).[2] Higher overall uncertainty or entropy corresponds to lower phylogenetic stability (in theory, a feature that varies across internal nodes of the tree but is reconstructed with high certainty at each node will also display low entropy). This method computes a linear version of the Shannon entropy on the basis of the probabilities of each of a character's K states reconstructed for each node n in the tree:

$$e_k^n = \begin{cases} p(state^n = k) & if p(state^n = k) \leq \frac{1}{K} \\ \frac{1}{1-K}\, p(state^n = k) - \frac{1}{1-K} & if p(state^n = k) > \frac{1}{K} \end{cases}$$

The entropy of each node can be obtained by summing over all K states, yielding a normalized entropy measure:

2. The second author thanks Natalia Chousou-Polydouri for bringing this method to his attention.

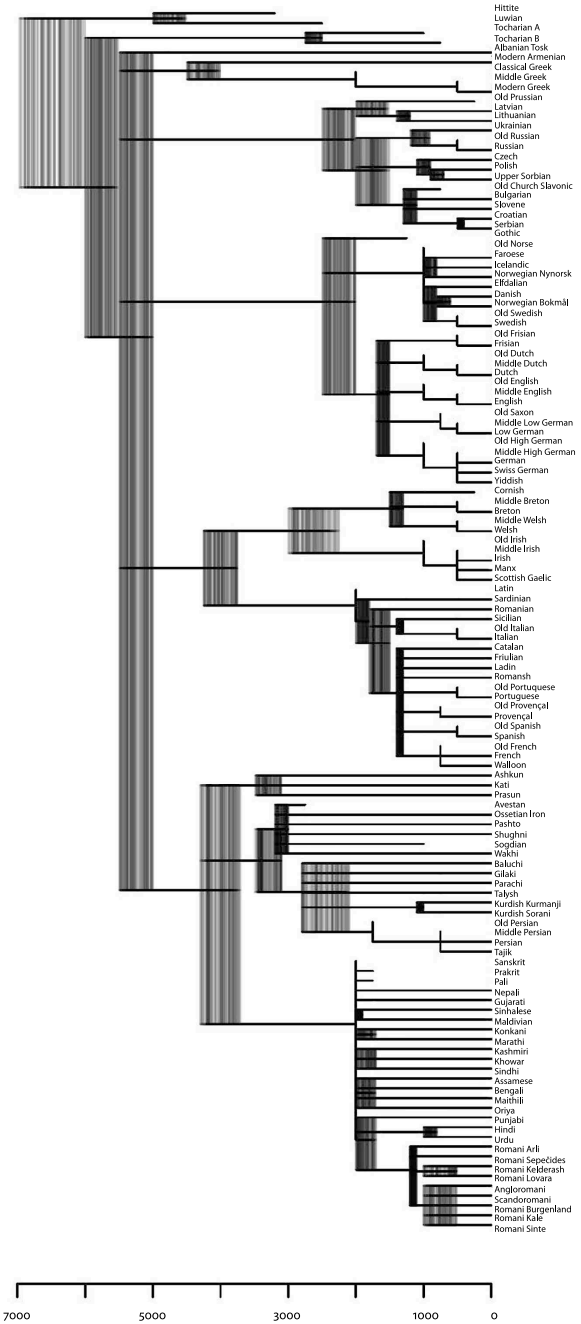


Figure 2. Phylogenetic reference tree for the Bayesian evolutionary reconstruction, based on a fixed tree topology

$$e_n = \sum_{k=1}^{k} e_k^n$$

A small constant is added to zero values of e_n. Because values of $\boldsymbol{e}$ are defined on the (0,1) interval, they can be modeled using the Beta distribution, assuming the following generative process:

$\lambda_o \sim Gamma(1,1)$; $\alpha \sim Exponential\ (\lambda_o)$; $\beta \sim Exponential\ (\lambda_o)$; for each internal node n, $e_n \sim Beta\ (\alpha, \beta)$

The parameters of the generative process can be inferred via Bayesian inference. We use the NUTS sampler of Rstan (with the same number of iterations and burn in as described above) to infer the posterior distributions of α and β. The δ statistic is computed for each sample by dividing β by α. Because Beta distributions with higher values of β relative to α generate values closer to zero, higher values of δ correspond to lower overall entropy and higher phylogenetic stability. Distributions of δ computed from posterior distributions of α and β for each character are found in Figure 4, in decreasing order (i.e., from highest to lowest phylogenetic stability). We evaluate the results in §3.4.

3. Results

3.1 Result overview

In this section, we overview and evaluate the results of the evolutionary reconstruction. As outlined in the introductory section (§1.2–1.5), there are several potentially interesting aspects of our results, which we consider under different headings.

For each trait in the data, to which we have given a unique ID (A1–30) and a label to increase transparency, we reconstruct the probability of presence at the root (for the method, see §2.2). Table 8 (columns from left to right) lists the multistate characters (i.e., categorical features) (1–9), the probabilities of presence for each trait at the root, the trait labels, and the trait ID numbers.

We highlight three different aspects of particular interest in our results, which have been mentioned in the introductory section (§1.2–1.5). The first aspect is the reconstructed trait probabilities at the proto-language state and the consequence of the results for the reconstruction of Proto-Indo-European morphosyntax (§3.2). The second aspect is the internal variation in probability levels between traits at the proto-language state, seen in the light of grammatical hierarchies in

alignment (§3.3). The third aspect is the phylogenetic strength of characters and the implication for the reconstruction and grammatical hierarchies (§3.4).

3.2 Proto-language probabilities in the light of reconstruction by the comparative method and diachronic typology

The results give a clear indication of nominative-accusative alignment (see Table 8). In the contrast between present and past, A marking is most frequent for both (Table 8: 1b). Nominative-accusative alignment is most probable with nouns as first argument in present progressive (Table 8: 2b), nouns as first argument in simple past (Table 8: 3b), pronouns as first argument in present progressive (Table 8: 4a), pronouns in simple past (Table 8: 5b), verbal marking in present progressive (Table 8: 6b), and verbal marking in simple past (Table 8: 7b). We also note that the reflexive pronoun in transitive clauses aligns with the A (Table 8: 8b) and not with the O (Table 8: 9a). These results indicate a total dominance of nominative-accusative alignment at the proto-language state. However, the differences in the degrees of probability are interesting. We notice that nominative-accusative alignment is more likely in present progressive than in simple past, both for nouns and pronouns (2b vs. 3b, 4a vs. 5b), and nominative-accusative alignment is also more likely with pronouns as first argument than with nouns (4a vs. 2b, 5b vs. 3b, see further §3.3). The second most frequent type of alignment is neutral marking, followed by ergative (both with low probabilities).

Considering the reconstructions at internal nodes of the tree (Figure 3a–d) we see that the only noteworthy tendency is a change to neutral marking in Western sub-branches, such as Romance, Celtic and Germanic, and a change to tripartite and/or ergative marking in some of the Indo-Iranian sub-branches (Hindi-Urdu, Kurdish, Ashkun-Kati-Prasun). Again, the tendency is stronger in categories lower in grammatical hierarchies, i.e., in nouns and in simple past, and the tendency is almost absent in pronouns.

The active-stative or ergative alignment theories have low support in our reconstruction, which points to nominative-accusative dominance in nominal and pronominal case marking, in verbal conjugation (Table 8), and in the present/past distinction. Our result matches a canonical model of Indo-European alignment (cf. Meier-Brügger et al. 2010: 412). The only alternative theory with slight support in our data (for nouns only) is the neutral marking theory proposed by Hermann Hirt (1934: 76ff.), which never found support in the historical-comparative scholarly community. This is mainly because attested historical change shows that case marking was often lost in later Indo-European languages.

Table 8. Result overview of multistate characters (i.e., categorical features) (numbers 1–9) and traits (roman letters a–d) discussed in the text, including probability of presence of traits at the root (see §3.2). The most probable traits are marked by ▸

	Probability at root	Categorical features and trait labels	ID
1		**Present-Past: Marking difference**	
1a	0.034475	Present-Past: No marking difference	A1
▸ 1b	0.765625	Present-Past: A marking in Present progressive and Past	A2
1c	0.106325	Present-Past: Active-ergative in Present progressive and Past	A3
1d	0.093575	Present-Past: All systems	A4
2		**Noun: Present progressive**	
2a	0.03415	Noun, Present progressive: Tripartite	A5
▸ 2b	0.631175	Noun, Present progressive: Nominative-accusative	A6
2c	0.334675	Noun, Present progressive: Neutral marking	A7
3		**Noun: Simple past**	
3a	0.041225	Noun, Simple past: Tripartite	A8
▸ 3b	0.526275	Noun, Simple past: Nominative-accusative	A9
3c	0.122925	Noun, Simple past: Ergative	A10
3d	0.309575	Noun, Simple past: Neutral marking	A11
4		**Pronoun: Present progressive**	
▸ 4a	0.90245	Pronoun, Present progressive: Nominative-accusative	A13
4b	0.09755	Pronoun, Present progressive: Neutral marking	A14
5		**Pronoun: Simple past**	
5a	0.035225	Pronoun, Simple past: Tripartite	A15
▸ 5b	0.77255	Pronoun, Simple past: Nominative-accusative	A16
5c	0.11405	Pronoun, Simple past: Ergative	A18
5d	0.078175	Pronoun, Simple past: Neutral marking	A19
6		**Verb: Present progressive**	
6a	0.036825	Verb, Present progressive: Tripartite	A24
▸ 6b	0.8777	Verb, Present progressive: Nominative-Accusative	A25
6c	0.05835	Verb, Present progressive: Neutral marking	A26
7		**Verb: Simple past**	
7a	0.0794	Verb, Simple past: Tripartite	A27

Table 8. *(continued)*

	Probability at root	Categorical features and trait labels	ID
▸ 7b	0.676275	Verb, Simple past: Nominative-accusative	A28
7c	0.089625	Verb, Simple past: Ergative	A29
7d	0.114575	Verb, Simple past: Neutral marking	A30
8		**Reflexive with Agent**	
8a	0.04195	Reflexive not with Agent	A20
▸ 8b	0.95805	Reflexive with Agent	A21
9		**Reflexive with Object**	
▸ 9a	0.95445	Reflexive not with Object	A22
9b	0.04555	Reflexive with Object	A20

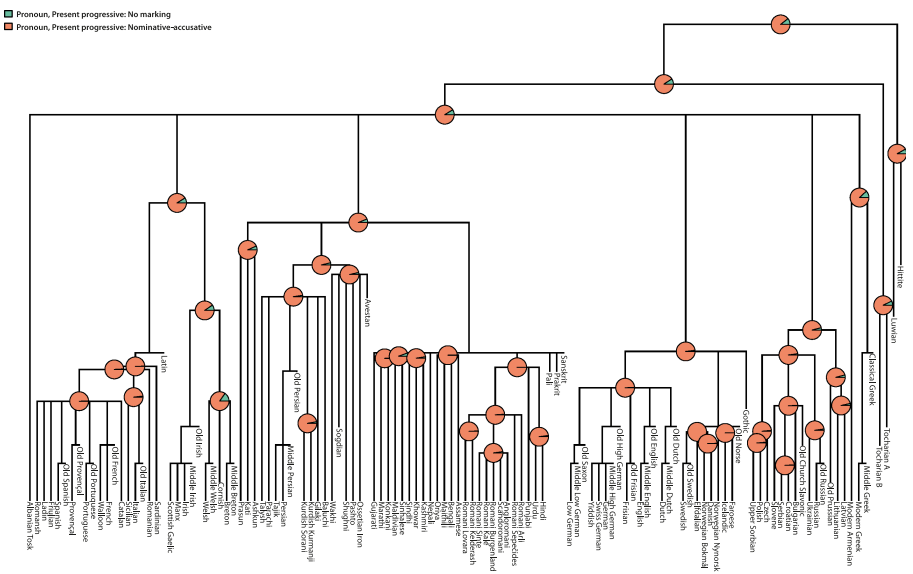


Figure 3a. Pie chart tree with reconstructions at hidden nodes for the categorical feature Pronoun: Present progressive

3.3 Probability levels and grammatical hierarchies of alignment

Here, we discuss the issue of grammatical or marking hierarchies in our results in the light of the received reconstructions for Proto-Indo-European, as well as the transition rates between states (§3.4). In the linguistic literature, typological generalizations about marking hierarchies typically refer to a synchronic variation or

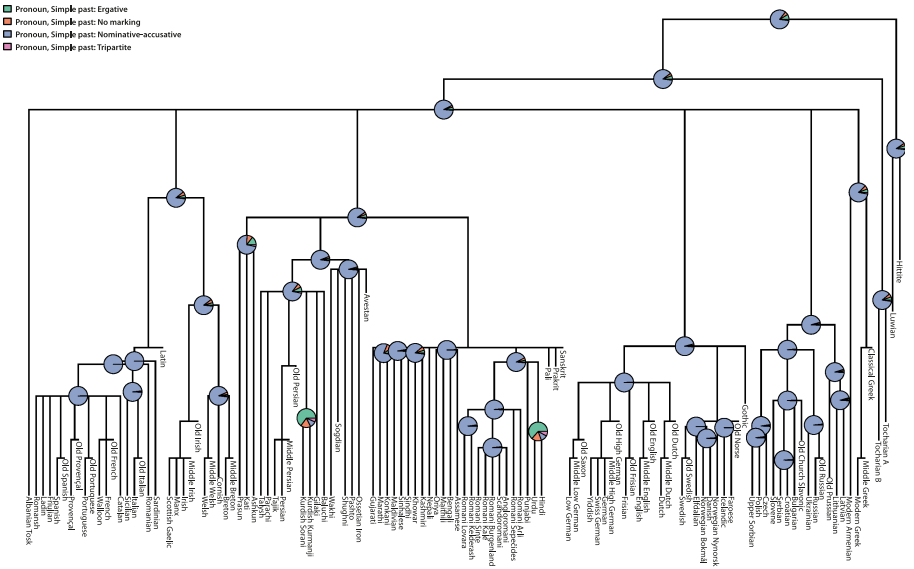


Figure 3b. Pie chart tree with reconstructions at hidden nodes for the categorical feature Pronoun: Simple past

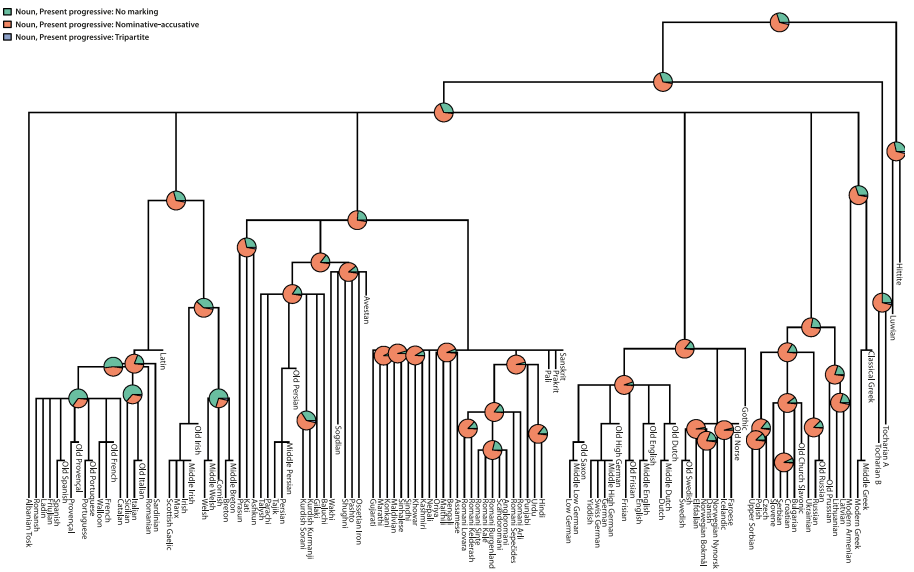


Figure 3c. Pie chart tree with reconstructions at hidden nodes for the categorical feature Noun: Present progressive

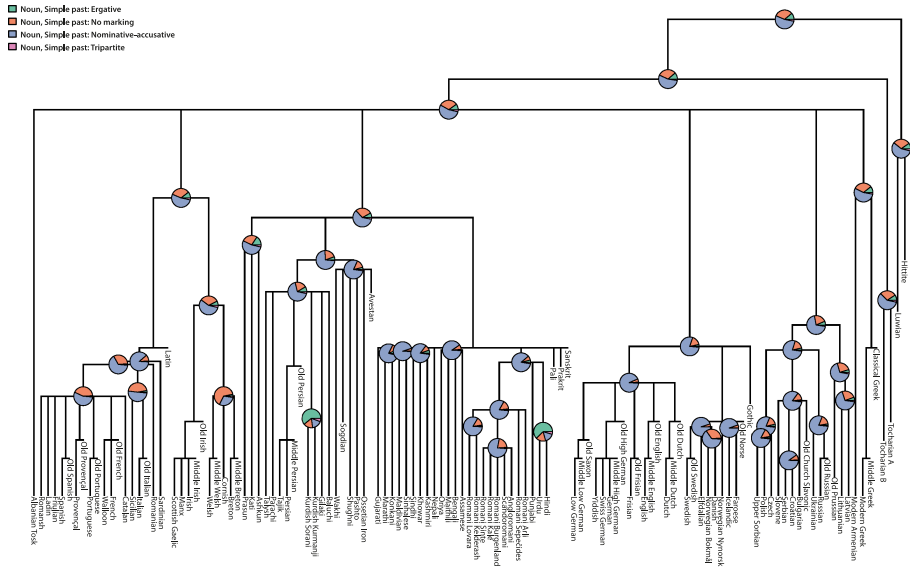


Figure 3d. Pie chart tree with reconstructions at hidden nodes for the categorical feature Noun: Simple past

deviation (Witzlack-Makarevich & Seržant 2018). Our data is based on the generalization of a default marking in a synchronic state (Carling et al. 2018) and for this reason, it is noteworthy that our result at the root mirrors a language-internal variation pattern. Of the marking hierarchies observed in the literature (Table 3), two categories are of relevance to our category definitions: present progressive < simple past, and pronoun < noun (additionally agent < object). We observe, as described in the previous section, a systematic difference between these categories in our result (Figure 4).

In what way are these results of relevance to marking hierarchies? The answer is complex. If we consider the distribution of probabilities at the Proto-Indo-European level (Table 8, also §3.2), we notice a tendency towards a clearer result, with a higher probability for the selected variant and a lower probability for other variants, in the grammatical categories of higher frequency and higher position in hierarchies (present, pronoun). On the contrary, we have a more unclear result, with lower probability for the selected variant and somewhat higher probabilities for other variants, in the categories of lower frequency and lower position in hierarchies (simple past, noun) (Table 3, Figure 4). In the present progressive by pronouns (Table 8: 4), we have a very high (0.9) probability of nominative-accusative marking, and a low probability of neutral marking (0.097). The other systems (ergative, tripartite), are absent. In present progressive by nouns (Table 8: 2), we have a medium probability, considerably lower than by pronouns, of nominative-

accusative marking (0.63). Besides, we find neutral marking (0.33) and tripartite marking (0.034). In simple past (Table 8: 3 and 5), we have a high probability of nominative-accusative marking (0.77), which is lower than in present progressive by pronouns, and besides, we have ergative (0.11), neutral marking (0.078), and tripartite (0.035). Second, we notice that the ergative appears (with a low probability), but in the simple past only.

The results indicate a clear pattern of hierarchical organization, which mirrors the marking hierarchy as observed in synchronic states of languages (§1.3).

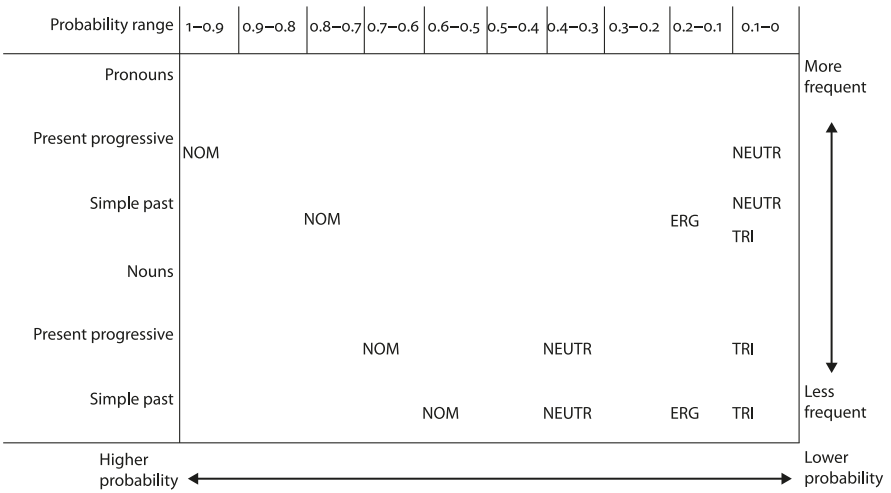


Figure 4. Overview of the ranges (divided into 10% segments) of the probabilities of the different tenses and word classes of alignment probabilities for Proto-Indo-European, contrasted to grammatical hierarchies of categories observed in languages. NOM = nominative-accusative, NEUTR = neutral marking, ERG = ergative, TRI = tripartite

3.4 Phylogenetic strength

We carry out the test for phylogenetic stability described in §2.3 on our data. We test each character (i.e., categorical feature) according to the method of Borges et al. (2018), which provides an estimation of the uncertainty of reconstructed states at internal nodes of the tree. The method gives a statistic called δ for evaluating the degree of phylogenetic signal between a phylogeny and categorical traits. The higher the δ value, the higher the degree of phylogenetic signal between a given tree and a trait, i.e., the higher phylogenetic strength (Figure 5). In general, we find that δ values show the same sensitivity to the grammatical hierarchy as the results of the proto-language root, discussed in the previous section. Characters involving pronominal marking show higher stability and lower phylogenetic vari-

ability than nouns. Within person categories such as pronoun and noun, the hierarchically higher tense category (i.e., present < past) shows higher phylogenetic stability as well. Taken together, pronoun is stronger than verb, which is stronger than noun. Agent is stronger than object. The present – past distinction character is between the strong and the weak group. Noun characters are the weakest of all, far below both pronouns and verbs.

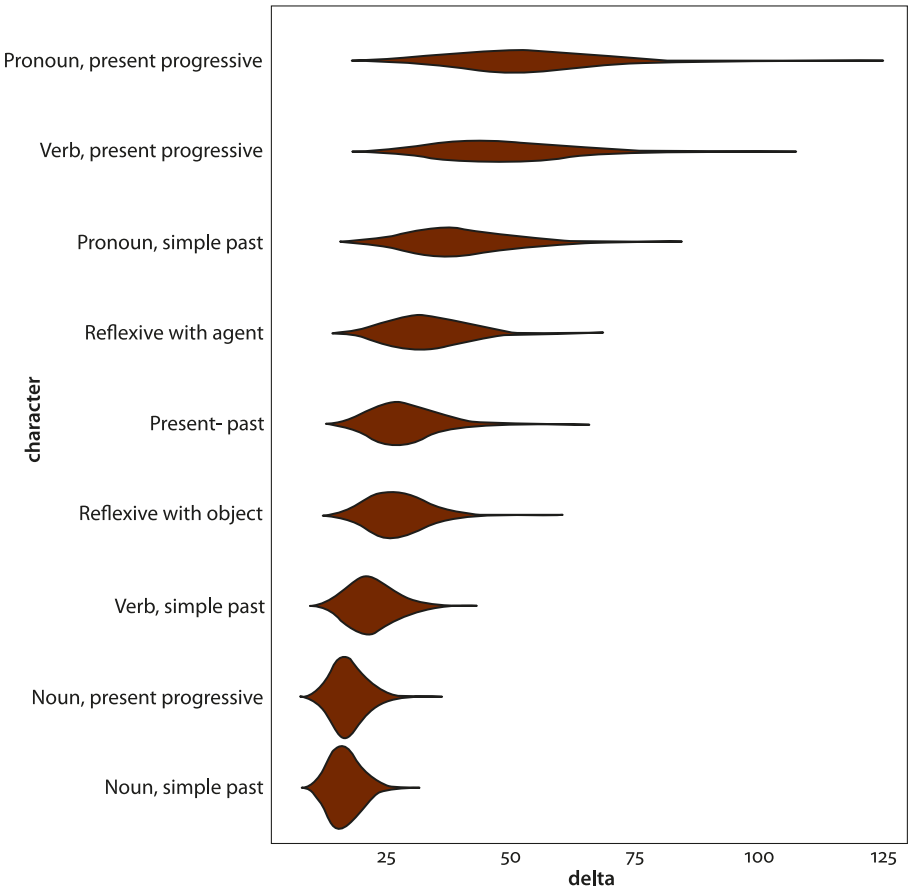


Figure 5. Delta distributions for each multistate character (i.e., categorical feature) of alignment marking, from highest to lowest phylogenetic stability

3.5 Summary of results

We have reconstructed the alignment system at the root of the Indo-European family tree. Alignment properties in our data are coded as multistate characters, which we define as categorical features with variants, called traits. We reconstruct

the evolution of Indo-European alignment by means of a model, which estimates gains and losses of states of traits against a phylogenetic reference tree. The outcome is a probability of the presence of traits at the root of the tree, which we consider equal to Proto-Indo-European as reconstructed by the comparative method. The result of our analysis is a high probability of a nominative-accusative system, which shows tendencies of neutral marking and ergative (and tripartite to a very low degree) in the less frequent categories of grammar, lower in the marking hierarchies (present < past, pronoun < noun).

Second, we have estimated the overall stability of phylogenetic states within characters (i.e., categorical features) as a whole using a technique which estimates the uncertainty of state reconstructions at all internal nodes of the tree. We find that character stability appears to follow grammatical hierarchies as well, where features higher in grammatical hierarchies (more frequent, unmarked) have a higher stability and features lower in hierarchies (less frequent, marked) have a lower stability. The result, which organizes our categorical features from highest to lowest phylogenetic stability, confirms the observed grammatical hierarchies of pronoun < noun, present < past, as well as agent < object. Interestingly enough, we also find that the verb is consistently weaker than the pronoun, but stronger than the noun.

4. Concluding discussion: Reconstructing the evolution of alignment

The current study employs a Bayesian comparative phylogenetic model for reconstructing grammar, i.e., morphology and syntax. The data consist of comparative concepts, i.e., abstractions of grammar properties in attested languages, which are not connected to any morphological material that can be reconstructed by the comparative method. In this aspect, the model complements models of reconstruction, which use morphological reconstruction in combination with syntactic meaning reconstruction and diachronic typology (see §1.3). The computational phylogenetic comparative method gives us valuable information about the general principles of grammar change of a family. The method takes attested grammar traits of all (or a representative amount of) languages of a family, calculates the transitions (gains and losses) for each trait of a categorical feature and returns an estimation of a probability of the trait at a proto-language state. The model is strictly phylogenetic and does not explicitly account for possible contact-induced change. A computational phylogenetic comparative model can be enhanced with so-called phylogeographic methods, where aspects of nearness in space and time and possible contact scenarios are taken into consideration (Cathcart et al. 2018). This is not done here: our model only accounts for the reconstruction and the

phylogenetic strength of traits within the family. Further, our data is confined to one family, Indo-European. The disadvantage of using data from one family is that the generality of claims may be connected to uncertainty. In addition, any contact-induced changes involving languages from other families cannot be accounted for. However, Indo-European is provided with a higher number of ancient languages than most other families and is therefore very useful for reconstruction.

The methodology relies on a uniformitarian model in the sense that grammar processes of unattested periods are supposed to reflect changes of attested periods (Labov 1972; Walkden 2019). A precondition to the approach is to adapt a uniformity-of-state model, just as a uniformity-of-rate model, to language change. This means that we expect rules that govern language structure to be similar in the present and in the past, and all languages to reflect some basic universal principles (Croft 2003: 233; Walkden 2019). Accordingly, we also expect basic universal principles to impact the processes of change (Croft 2003: 272–279). In addition, the methodology has the shortcoming (or advantage) that it cannot reconstruct novel states. The estimations of probability at hidden nodes are based entirely on features that are attested in the data.

Considering these facts, it is remarkable how close our result approaches a canonical reconstruction model of Indo-European alignment, as presented already by the Neogrammarians (Delbrück 1893, 1897, 1900; Krahe et al. 1972; see Meier-Brügger et al. 2010). Despite the variation and change found at various states over the Indo-European family tree, the model reconstructs a synthetic, nominative-accusative alignment system of the proto-language. A fundamental aspect of evaluation is to consider how this result relates to syntactic reconstruction by the linguistic comparative method. Since our results are based on a computational phylogenetic comparative reconstruction of typological comparative concepts, they should be seen as complementary to results received by the linguistic comparative method, using morphological reconstruction and diachronic typology. However, the two reconstructions should not be considered as completely unrelated: if the two models of reconstruction (the linguistic comparative versus computational phylogenetic comparative model) give disparate results, something is fundamentally wrong with one of the reconstructions.

The alternative (ergative, active-stative) theories of Proto-Indo-European (see §1.1) assume profound typological changes from Early to Late Proto-Indo-European, which are concluded by internal reconstruction based on Proto-Indo-European paradigmatic correlations and a typological comparison with other, unrelated families (e.g., Caucasian). Let us scrutinize the arguments of the alternative theories in the light of the results achieved by the comparative phylogenetic reconstruction. The issue has several aspects. First, the chronological aspect of the dating and extension in time of the Indo-European proto-state (Bouckaert et al.

2012; Chang et al. 2015; Meid 1975; Schlerath 1981). Second, the cyclic (or spiral) behavior of typological features (nominative-accusative → ergative or neutral; ergative or neutral → nominative-accusative) (Dixon 1997: 182–206) of the type that we investigate here, including the timespan of these changes (Croft 2003: 252; Haspelmath 2018; Hock & Joseph 1996: 183–184).

In the alternative models, Indo-European developed from an active-stative or ergative system into a nominative-accusative system during the period from early to late Proto-Indo-European or from Proto-Indo-European and continuing into the sub-branches (see §1.3). As we have seen, the nominative-accusative alignment type is dominant in many contemporary and most historical states of the family (Figure 1) and is also reconstructed with the highest probability for the proto-language state (Table 8). Neutral marking is reconstructed for the proto-language with a lower probability (nouns only, 0.3). The system of reconstruction inferred by our model is compatible with a trend where a nominative-accusative system is either eroding into a neutral system or occasionally developing into an ergative system. In both cases, the trend appears more pronounced in categories lower in grammatical hierarchies (past tense, noun, object), an interpretation supported by the fact that categories higher in grammatical hierarchies (present tense, pronoun, agent) show higher overall phylogenetic stability. Emergence of other systems are either marginal (tripartite) or non-existent (active-stative). The problem of alternative theories thus becomes somewhat of an Ockham's razor problem: why assume a system at the proto-language, which is concluded by exceptions within the attested languages (e.g., tendency to ergativity in Hittite, residues of animacy marking in Anatolian, non-canonical case marking in Latin)? The system is not attested as such in any of the surviving Indo-European languages (ergativity in Modern Indo-Aryan is clearly an innovation) and it is not supported by the general typological trend (keeping the nominative-accusative structure) of the past 6000–7000 years in the family.

An interesting factor here is the confirmation of grammatical hierarchies at the proto-language state as well as in the evolutionary dynamics in the tree. Grammatical hierarchies have been extensively discussed in the linguistic literature, and the underlying causes for hierarchies, which often come out as a markedness distinction in languages, relate to animacy criteria as well as economy and frequency (Comrie 1981; Greenberg 1963, 1966; Hawkins 2004; Jaeger 2010). As we have seen in §3.3, the distribution of probabilities at the root in higher categories (pronoun < noun, present < past) is clearer, with a higher probability of the preferred variant over the non-preferred variant (Figure 1). In addition, alternative systems (ergative, tripartite) appear only in the categories lower in the hierarchy (noun, past). The distribution of phylogenetic strength confirms this tendency, where categories higher in grammatical hierarchies consistently have higher phylogenetic

strength, whereas categories lower in grammatical hierarchies have a lower phylogenetic strength (Figure 4) (Haiman 1983; Haspelmath 2008; cf. Jaeger 2010). In this respect, our results are more in line with the theoretical implications by, e.g., Rumsey (1987), Villar (1984), who argue, following Silverstein (1976), that the ergative (and active-stative) theories are untenable "for strictly comparative reasons as for reasons drawn from the normal behavior of ergative, accusative, and split systems" (Villar 1984: 175).

In sum, our computational phylogenetic comparative reconstruction confirms Proto-Indo-European as a nominative-accusative language with a slight tendency to neutral marking and ergativity in the categories lower in grammatical hierarchies (past, noun). This goes for alignment marked both by cases as well as by the verb. The nominative-accusative marking is consistent with respect to tense (present, past) and reflexives align with the agent. Phylogenetically, pronoun is stronger than verb, which is stronger than noun. Agent is stronger than object, and present is stronger than past. In all, noun categories are the weakest of all characters. The result confirms that the dynamics of language change is both individual and variating, as well as subjugated to universal principles of typology, such as frequency and economy. The result also confirms that a proto-language may be just as varying as any attested or spoken language.

Funding

The research was supported by the project CHRONOS – Chronology of Roots and Nodes of Family Trees (MAW 2017.0050) from the Marcus and Amalia Wallenberg Foundation.

Acknowledgements

We thank our colleagues Arthur Holmer, Filip Larsson, and Niklas Erben Johansson for input on data and coding. We thank the anonymous reviewers for their suggestions. We also thank Simon Devylder and Johannes Dellert for proofreading (French/German translations).

References

Barðdal, Jóhanna. 2014. Syntax and syntactic reconstruction. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 343–373. London/New York: Routledge.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2009. The origin of the oblique subject construction: An Indo-European comparison. In Vit Bubenik, John Hewson & Sarah Rose (eds.), *Grammatical change in Indo-European languages*, 179–193. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.305.19bar

Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. Reconstructing syntax: Construction grammar and the Comparative Method. In Hans C. Boas & Ivan A. Sag (eds.), Sign-based construction grammar, 257–308. Stanford: Center for the Study of Language and Information.

Bauer, Brigitte. 2000. *Archaic syntax in Indo-European: The spread of transitivity in Latin and French* (Trends in linguistics. Studies and monographs 25). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110825992

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1). 239–251. https://doi.org/10.1515/LINGTY.2007.018

Bickel, Balthasar. 2008. On the scope of the referential hierarchy in the typology of grammatical relations. In Greville G. Corbett & Michael Noonan (eds.), *Case and grammatical relations. Studies in honor of Bernard Comrie*, 191–210. Amsterdam/ Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.81.09ont

Bickel, Balthasar. 2010. Capturing particulars and universals in clause linkage: A multivariate analysis. In Isabelle Bril (ed.), *Clause linking and clause hierarchy: Syntax and pragmatics*, 51–101. Amsterdam – Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.121.03bic

Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 399–444. Oxford: Oxford University Press.

Bickel, Balthasar & Johanna Nichols. 2002. Autotypologizing databases and their use in fieldwork. In Peter Austin, H. Dry & P. Wittenburg (eds.), *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics, Las Palmas*. Nijmegen: ISLE and DOBES.

Bollback, J. P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7, 88 (2006). https://doi.org/10.1186/1471-2105-7-88

Borges, Rui, João Paulo Machado, Cidália Gomes, Ana Paula Rocha & Agostinho Antunes. 2018. Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics* 2018. 1–8.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097). 957–960. https://doi.org/10.1126/science.1219669

Bowern, Claire. 2018. Computational phylogenetics. *Annual Review of Linguistics* 4(1). 281–296. https://doi.org/10.1146/annurev-linguistics-011516-034142

Brugman, Claudia & Anne David. 2014. *Descriptive grammar of Pashto and its dialects*. Berlin/New York: Mouton De Gruyter.

Bubenik, Vit. 2016. On the establishment of ergative alignment during the Late Middle Indo-Aryan period. In Eystein Dahl & Krzysztof Stronski (eds.), *Indo-Aryan ergativity in typological and diachronic perspective*, 109–132. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.112.04bub

Calude, Andreea S. & Annemarie Verkerk. 2016. The typology and diachrony of higher numerals in Indo-European: A phylogenetic comparative study. *Journal of Language Evolution* 1(2). 91–108. https://doi.org/10.1093/jole/lzw003

Campbell, Lyle & Alice C. Harris. 2002. Syntactic reconstruction and demythologizing 'Myths and the Prehistory of Grammars'. *Journal of Linguistics* 38(3). 599–618. https://doi.org/10.1017/S0022226702001706

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. 2017. *DiACL – Diachronic Atlas of Comparative Linguistics Online*. https://diacl.ht.lu.se/

Carling, Gerd. 2019. *Mouton atlas of languages and cultures. Vol. 1: Europe and West, Central and South Asia*. Berlin/Boston: Mouton de Gruyter. https://doi.org/10.1515/9783110367416

Carling, Gerd & Chundra Cathcart. 2021. Reconstructing the evolution of Indo-European grammar. *Language*.

Carling, Gerd, Filip Larsson, Chundra A. Cathcart, Niklas Johansson, Arthur Holmer, Erich Round & Rob Verhoeven. 2018. Diachronic Atlas of Comparative Linguistics (DiACL) – A Database for Ancient Language Typology. *PLOS ONE* 13(10). https://doi.org/10.1371/journal.pone.0205313

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, & Allen Riddell. 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76(1). 1–32. https://doi.org/10.18637/jss.v076.i01

Cathcart, Chundra, Gerd Carling, Filip Larsson, Niklas Johansson & Erich Round. 2018. Areal pressure in grammatical evolution. *Diachronica* 35(1). 1–34. https://doi.org/10.1075/dia.16035.cat

Cennamo, Michaela. 2009. Argument structure and alignment variations and changes in Late Latin. In Jóhanna Barðdal & Shobhana Lakshmi Chelliah (eds.), *The rope of semantic, pragmatic, and discourse factors in the development of case*, 307–346. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.108.17cen

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

Clackson, James. 2007. *Indo-European linguistics: An introduction* (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808616

Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.

Croft, William. 2003. *Typology and universals* (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press.

Dahl, Eystein. 2016. The origin and development of the Old Indo-Aryan predicated *-tá* construction. In Eystein Dahl & Krzysztof Stronski (eds.), *Indo-Aryan ergativity in typological and diachronic perspective*, 61–108. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.112.03dah

Dahl, Eystein & Krzysztof Stroński. 2016a. Ergativity in Indo-Aryan and beyond. In Eystein Dahl & Krzysztof Strónski (eds.), *Indo-Aryan Ergativity in typological and diachronic perspective*, 1–37. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.112.01dah

Dahl, Eystein & Krzysztof Stroński. 2016b. *Indo-Aryan ergativity in typological and diachronic perspective* (Typological Studies in Language). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.112

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. https://doi.org/10.1515/9783111600550

Dench, Alan. 1982. The development of an accusative case marking pattern in the Ngayarda languages of Western Australia. *Australian Journal of Linguistics*, 2(1). 43–59. https://doi.org/10.1080/07268608208599281

Dixon, R. M. W. 1979. Ergativity. *Language* 55(1). 59–138. https://doi.org/10.2307/412519

Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511611896

Dixon, R. M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612060

Dixon, R. M. W. 2010. *Basic linguistic theory. Vol. 2, Grammatical topics*. Oxford: Oxford University Press.

Drinka, Bridget. 1999. Alignment in Early Proto-Indo-European. In Carol F. Justus & Edgar C. Polomé (eds.), *Language change and typological variation: In honor of Winfred P. Lehmann on the occasion of his 83rd birthday* (Journal of Indo-European Studies, Monograph Series, II), 464–500. Washington D.C.: Institute for the Study of Man.

Dryer, Matthew S. & Martin Haspelmath. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info

Dunn, Michael. 2014. Language phylogenies. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 190–211. Florence: Routledge.

Felsenstein, Joseph. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17(6). 368–376. https://doi.org/10.1007/BF01734359

Felsenstein, Joseph. 2004. *Inferring phylogenies*. Sunderland, Mass.: Sinauer.

Friedrich, Paul. 1975. *Proto-Indo-European syntax: The order of meaningful elements* (Journal of Indo-European studies. Monograph 4). Washington: Institute for the Study of Man.

Gamkrelidze, Tamaz Valerianovič, Vjačeslav Vsevolodovič Ivanov & Werner Winter. 1995. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture* (Trends in Linguistics. Studies and Monographs 80). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110815030

Garrett, Andrew. 1990. The origin of NP split ergativity. *Language* 66(2). 261–96. https://doi.org/10.2307/414887

Garrett, Andrew. 1996. Wackernagel's Law and unaccusativity in Hittite. In Aaron Halpern & Arnold M. Zwicky (eds.), *Approaching second. Second position clitics and related phenomena*, 85–133. Stanford: CSLI Publications.

Gelman, Andrew & Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* (4). 457–511. https://doi.org/10.1214/ss/1177011136

Goedegebuure, Petra. 2013. Split-ergativity in Hittite. *Zeitschrift für Assyriologie und verderasiatische Archäologie* 102. 207–303. https://doi.org/10.1515/za-2012-0015

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426 (6965). 435–439. https://doi.org/10.1038/nature02029

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* (Janua linguarum: Series minor 59). The Hague: Mouton.

Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade & Russell D. Gray. 2010. The shape and tempo of language evolution. *Proceedings of the Royal Society: Biological Sciences* 277 (1693). https://doi.org/10.1098/rspb.2010.0051

Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819. https://doi.org/10.2307/413373

Harris, Alice C. 2008. Reconstruction in syntax. Reconstruction of patterns. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 73–95. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.302.05har

Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective* (Cambridge studies in linguistics 74). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620553

Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33. https://doi.org/10.1515/COG.2008.001

Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15(3). 535–567. https://doi.org/10.1515/LITY.2011.035

Haspelmath, Martin. 2018. Revisiting the anasynthetic spiral. In Heiko Harrog & Berndt Heine (eds.), *Grammaticalization from a typological perspective*, 97–115. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198795841.003.0006

Hawkins, John A. 2004. *Efficiency and complexity in grammars* (Oxford linguistics). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Hendery, Rachel. 2012. *Relative clauses in time and space: A case study in the methods of diachronic typology* (Typological Studies in Language). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.101

Hirt, Hermann Alfred. 1934. *Indogermanische Grammatik*. Cambridge: Cambridge University Press.

Hock, Hans Henrich. 2013. Proto-Indo-European verb-finality reconstruction, typology, validation. *Journal of Historical Linguistics* 3(1). 49–76. https://doi.org/10.1075/jhl.3.1.04hoc

Hock, Hans Henrich & Brian D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics* (Trends in Linguistics. Studies and Monographs 93). Berlin: Mouton de Gruyter.

Hopper, Paul & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165525

Hrozný, Bedřich. 1915. Die Lösung des hethitischen Problems. *Mitteilungen der Deutschen Orient-Gesellschaft* 56. 17–50.

Huelsenbeck, John P., Rasmus Nielsen & Jonathan P. Bollback. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52(2). 131–158. https://doi.org/10.1080/10635150390192780

Jaeger, Florian T. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Jasanoff, Jay H. 1978. *Stative and middle in Indo-European* (Innsbrucker Beiträge zur Sprachwissenschaft 23). Innsbruck: Institut für Sprachwissenschaft.

Jasanoff, Jay H. 2003. *Hittite and the Indo-European verb* (Oxford scholarship online). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249053.001.0001

Jäger, Gerhard. 2019. Computational historical linguistics. *Theoretical Linguistics* 45(3/4). 151–182. https://doi.org/10.1515/tl-2019-0011

Klimov, Georgij Andreevich. 1973a. *Očerk obščej teorii èrgativnosti*. Moskva: Izdateľstvo Nauka.

Klimov, Georgij Andreevich. 1973b. Tipologija jazykov aktivnogo stroja i rekonstrukcija protoin-doevropejskogo. *Izvestija Akademii Nauk SSSR, Serija literatury i jazyka* 32. 442–447.

Klimov, Georgij Andreevich. 1974. On the character of languages of active typology. *Linguistics* (12), 11–25. https://doi.org/10.1515/ling.1974.12.131.11

Kortlandt, Frederik. 1983. Proto-Indo-European verbal syntax. *Journal of Indo-European Studies* 11. 307–324.

Krahe, Hans, Hans Schmeja & Wolfgang Meid. 1972. *Grundzüge der vergleichenden Syntax der indogermanischen Sprachen* (Innsbrucker Beiträge zur Sprachwissenschaft 8). Innsbruck: Institut für vergleichende Sprachwissenschaft.

Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* (1). 97–120. https://doi.org/10.1017/S0047404500006576

Lehmann, Winfred P. 1973. A structural principle of language and its implications. *Language* 49(1). 47–66. https://doi.org/10.2307/412102

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

Lehmann, Winfred P. 1989. Problems of Proto-Indo-European grammar – Residues from Pre-Indo-European active structure. *General Linguistics* 29. 228–246.

Lopuhaä-Zwakenberg, Milan. 2019. The Anatolian "ergative". In Alwin Kloekhorst & Tijmen Pronk (eds.), *The precursors of Proto-Indo-European. The Indo-Anatolian and Indo-Uralic Hypotheses* (Leiden Studies in Indo-European), 131–150. Rodopi: Brill. https://doi.org/10.1163/9789004409354_011

Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45(2). 435–464. https://doi.org/10.1515/flin.2011.016

Luraghi, Silvia. 2012. Basic valency orientation and the middle voice in Hittite. *Studies in Language* 36(1). 1–32. https://doi.org/10.1075/sl.36.1.01lur

Malchukov, Andrej L. 2015. Towards a typology of split ergativity: A TAM-hierarchy for alignment splits. In Ina Bornkessel-Schlesewsky, Andrej Malchukov & Marc D. Richards (eds.), *Scales and hierarchies: A cross-disciplinary perspective*, 275–296. Berlin – New York: Mouton de Gruyter. https://doi.org/10.1515/9783110344134.275

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. 2011. Clause alignment in Indo-European. *Manuscript*. Zagreb: Zagreb University. http://mudrac.ffzg.unizg.hr/~rmatasov/ClauseAlignmentInPIE.pdf.

Matasović, Ranko. 2013. Latin paenitet me, miseret me, pudet me and active clause alignment in Proto-Indo-European. *Indogermanische Forschungen* (118). 93–110. https://doi.org/10.1515/indo.2013.118.2013.93

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

Maurits, Luke & Thomas L. Griffiths. 2014. Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences* 111(37). 13576–13581. https://doi.org/10.1073/pnas.1319042111

McGregor, W. B. 2009. Typology of ergativity. *Linguistics and Language Compass* 3(1). 480–508. https://doi.org/10.1111/j.1749-818X.2008.00118.x

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

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

Meiser, Gerhard. 2009. Zur Typologie des urindogermanischen Mediums. In Rosemarie Lühr & Sabine Ziegler (eds.), *Protolanguage and Prehistory. Akten der XII. Fachtagung der Indogermanischen Gesellschaft, vom 11. bis 15. Oktober 2004 in Krakau*, 318–334. Wiesbaden: Reichert.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226580593.001.0001

Nichols, Johanna. 1995. Diachronically stable structural features. In Henning Andersen (ed.), *Historical Linguistics 1993. Selected Papers from the 11th International Conference on Historical Linguistics. Los Angeles 16–20 August 1993*, 337–355. Amsterdam – Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.124.27nic

Nichols, Johanna. 1998. The Eurasian spread zone and the Indo-European dispersal. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language II. Archaeological data and linguistic hypotheses*, 220–266. New York: Routledge. https://doi.org/10.4324/9780203202913_chapter_10

Nielsen, Rasmus. 2002. Mapping mutations on phylogenies. *Systematic Biology* (5). 729–739. https://doi.org/10.1080/10635150290102393

Pooth, Roland, Peter Alexander Kerkhof, Leonid Kulikov & Jóhanna Barðdal. 2019. The origin of non-canonical case marking of subjects in Proto-Indo-European. *Indogermanische Forschungen* 124. 245–263. https://doi.org/10.1515/if-2019-0009

Roberts, Ian G. 2007. *Diachronic syntax* (Oxford textbooks in linguistics 2). Oxford: Oxford University Press.

Ronan, Patricia. 2011. The Celtic languages. In Bernd Kortmann & Johan van der Auwera (eds.), *The languages and linguistics of Europe. A comprehensive guide*, 31–46. Berlin/New York: Mouton De Gruyter. https://doi.org/10.1515/9783110220261.31

Rumsey, A. 1987. The chimera of Proto-Indo-European ergativity. Lessons for historical syntax. *Lingua* 71(1–4). 297–318. https://doi.org/10.1016/0024-3841(87)90076-3

Schlerath, Bernfried. 1981. Ist ein Raum/Zeit-Modell für eine rekonstruierte Sprache möglich?. *Zeitschrift für vergleichende Sprachforschung* 95(2). 175–202.

Schmalsteig, William R. 1981. Ergativity in Indo-European. In Yoël L. Arbeitman & Allan R. Bomhard (eds.), *Bono Homini Donum: Essays in Historical Linguistics in Memory of J. Alexander Kerns*, 243–258. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.16.26sch

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

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

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian National University.

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

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.

Walkden, George. 2013. The correspondence problem in syntactic reconstruction. *Diachronica*, 30(1). 95–122. https://doi.org/10.1075/dia.30.1.04wal

Walkden, George. 2019. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics* 4(1), 1–17. https://doi.org/10.5334/gjgl.888

Watkins, Calvert. 1976. Towards Proto-Indo-European syntax: Problems and pseudoproblems. In Sanford Steever, Carol A. Walker & Salikoko S. Mufwene (eds.), *Papers from the Parasession on Diachronic Syntax*, 305–326. Chicago: Chicago Linguistic Society.

Verbeke, Saartje & De Cuypere, Ludovic. 2009. The rise of ergativity in Hindi: Assessing the role of grammaticalization. *Folia Linguistica Historica* (43). 367–389. https://doi.org/10.1515/FLIH.2009.006

Verbeke, Saartje & De Clercq, Eva. 2016. Looking for ergativity in Indo-Aryan. In Eystein Dahl & Krzysztof Stronski (eds.), *Indo-Aryan ergativity in typological and diachronic perspective*, 39–60. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.112.02ver

Villar, Francisco. 1984. Ergativity and animate/inanimate gender in Indo-European. *Zeitschrift für vergleichende Sprachforschung* 97(2). 167–196.

Viti, Carlotta. 2015. Historical syntax: Problems, materials, methods, hypotheses. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 3–34. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.169.01vit

Winter, Werner. 1984. Reconstructional comparative linguistics and the reconstruction of undocumented stages in the development of languages and language families. In Jacek Fisiak (ed.), *Historical syntax*, 613–625. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110824032.613

Witzlack-Makarevich, Alena & Seržant, Ilja A. 2018. Differential argument marking: Patterns of variation. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 1–40. Berlin: Language Science Press.

Yang, Z. 2014. *Molecular evolution: A statistical approach*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199602605.001.0001

Zhou, Kevin & Claire Bowern. 2015. Quantifying uncertainty in the phylogenetics of Australian numeral systems. *Proceedings of the Royal Society B*, 282(1815). https://doi.org/10.1098/rspb.2015.1278

Appendices: Data and results

Appendix 1.

Languages, including latitude and longitude, used in the current study

Languages	Latitude	Longitude
Albanian (Tosk)	40,44695	19,98825
Angloromani	53,77952	−2,35527
Ashkun	35,25591	70,79106
Assamese	26,14354	91,79022
Avestan	31,70708	55,9499
Baluchi	26,27794	65,03622
Bengali	23,78057	90,27924
Breton	47,99562	−4,10286
Bulgarian	42,69859	23,3535
Catalan	41,37923	2,179642
Classical Greek	37,69686	23,99921
Cornish	50,19182	−5,56752
Croatian	45,81349	15,99266
Czech	50,07124	14,46814
Danish	55,67989	12,58318
Dutch	52,37437	4,898126
Elfdalian	61,22594	14,04049
English	51,49773	−0,10006
Faroese	62,00975	−6,77332
French	48,85445	2,347857
Frisian	53,20394	5,795597
Friulian	46,12703	13,485
German	52,52082	13,40909
Gilaki	37,52606	49,284
Gothic	53,04287	20,19932
Gujarati	22,44456	72,13987

Languages	Latitude	Longitude
Hindi	28,62525	77,225
Hittite	40,01395	34,62116
Icelandic	64,14424	−21,9397
Irish	53,24443	−9,3011
Italian	43,11596	12,38515
Kashmiri	34,323	75,93918
Kati	35,79999	71,31457
Khowar	36,45796	72,51658
Konkani	19,26208	72,87107
Kurdish (Kurmanji)	37,78808	43,82275
Kurdish (Sorani)	35,82116	45,72851
Ladin	46,65738	11,92376
Latin	41,89019	12,4923
Latvian	56,915	24,11327
Lithuanian	54,68018	25,25585
Low German	53,93199	9,488653
Luwian	36,6206	36,78827
Maithili	26,0754	86,15545
Maldivian	4,17191	73,51194
Manx	54,15199	−4,48466
Marathi	19,38701	75,49471
Middle Breton	47,20785	−1,53247
Middle Dutch	52,21307	5,959192
Middle English	51,31361	−0,75862
Middle Greek	41,05864	28,99756
Middle High German	48,74272	9,717403
Middle Irish	52,65892	−8,63441
Middle Low German	53,86778	10,68504
Middle Persian	32,64703	51,67588
Middle Welsh	52,24036	−3,38002
Modern Armenian	40,14739	44,52716
Modern Greek	38,02862	23,69433
Nepali	26,6597	89,24363
Norwegian (Bokmål)	59,91279	10,7408
Norwegian (Nynorsk)	60,33098	5,075361

Languages	Latitude	Longitude
Old Church Slavonic	43,95597	22,87177
Old Dutch	52,09095	5,123259
Old English	51,0607	−1,31416
Old French	48,85445	2,347857
Old Frisian	53,2157	6,567738
Old High German	49,79548	9,962616
Old Irish	53,72679	−6,87383
Old Italian	43,77628	11,24925
Old Norse	63,42698	10,39752
Old Persian	29,9355	52,8912
Old Portuguese	41,15818	−8,62914
Old Provençal	43,42683	6,222519
Old Prussian	54,43629	19,9008
Old Russian	58,52143	31,27535
Old Saxon	53,07585	8,806043
Old Spanish	42,386	−3,894
Old Swedish	58,48108	16,32194
Oriya	20,60855	86,27841
Ossetian (Iron)	43,0438	44,67569
Pali	27,6179	83,00038
Parachi	34,83289	69,7026
Pashto	31,43293	68,1888
Persian	32,73531	55,74804
Polish	52,23789	20,99402
Portuguese	38,70804	−9,13513
Prakrit	25,3791	84,72798
Prasun	35,33661	70,74162
Provençal	43,42683	6,222519
Punjabi	31,5366	74,34941
Romani (Arli)	42,97035	19,88117
Romani (Burgenland)	47,19316	16,37753
Romani (Kale)	60,97209	21,47097
Romani (Kelderash)	47,18768	22,49509
Romani (Lovara)	47,40154	23,0902
Romani (Sepečides)	38,3928	27,13167

Languages	Latitude	Longitude
Romani (Sinte)	51,4992	10,20342
Romanian	44,42751	26,08637
Romansh	46,49676	9,838341
Russian	55,67758	37,64843
Sanskrit	23,51142	76,28612
Sardinian	40,31746	9,32737
Scandoromani	59,48131	13,15862
Scottish Gaelic	57,16276	−7,36897
Serbian	44,80133	20,47216
Shughni	38,38964	71,50304
Sicilian	38,11514	13,36835
Sindhi	23,25348	69,6748
Sinhalese	6,935115	79,85635
Slovene	46,05059	14,50823
Sogdian	39,64658	66,96268
Spanish	40,41688	−3,70348
Swedish	59,73823	17,43872
Swiss German	47,36888	8,538354
Tajik	38,66346	70,34262
Talysh	38,85417	48,73002
Tocharian A	41,76034	86,15479
Tocharian B	41,6455	81,51444
Ukrainian	50,44947	30,52541
Upper Sorbian	51,17991	14,42544
Urdu	33,68114	73,01842
Wakhi	37,21948	72,76006
Walloon	50,46448	4,865382
Welsh	52,92641	−4,38518
Yiddish	49,85303	24,03117

Appendix 2.

List of typological features for alignment (extracted from DiACL database, https://diacl.ht.lu.se/) used in current study. Grid = topmost organizational unit in database, corresponding to linguistic domain, Feature = second organizational unit in database, Feature description = Description of Feature in database, Variant = lowest organizational unit in database, Variant description = description of variant in database, ID = unique database ID of Variant.

Grid	Feature	Feature description	Variant	Variant description	ID
Alignment	Noun: Simple Past	In simple past: how is the marking of subject and object of nouns realized?	N:PST:A=O?	In simple past: Is the noun form for A the same as for O? Ie: Does the noun look the same when it is subject of a transitive clause than when it is object of a transitive clause?	302
Alignment	Noun: Simple Past	In simple past: how is the marking of subject and object of nouns realized?	N:PST:A=Sa?	In simple past:Is the noun form for A the same as for Sa? Ie: Does the noun look the same when it is subject of a transitive clause as when it is subject of an agentive intransitive verb such as "work" or "dance"?	303
Alignment	Noun: Simple Past	In simple past: how is the marking of subject and object of nouns realized?	N:PST: O=So?	In simple past: Is the noun form for O the same as for So? Ie: Does the noun look the same when it is object of a transitive clause as when it is subject of an unaccusative	304

Grid	Feature	Feature description	Variant	Variant description	ID
				verb such as "fall" or "die"?	
Alignment	Noun: Simple Past	In simple past: how is the marking of subject and object of nouns realized?	N:PST:Sa=So?	In simple past: Does a noun bear the same case form when it is Sa (subject of e.g. work) or So (subject of e.g. fall or die)? Ie: There does not exist a split into stative and active intransitive verbs.	305
Alignment	Noun: Present Progressive	In present progressive: how is the marking of subject and object of nouns realized?	N:PROG:A=O?	In present progressive: Is the noun form for A the same as for O? I.e.: Does the noun look the same when it is subject of a transitive clause and when it is object of a transitive clause?	306
Alignment	Noun: Present Progressive	In present progressive: how is the marking of subject and object of nouns realized?	N:PROG:A=Sa?	In present progressive:Is the noun form for A the same as for Sa? I.e.: Does the noun look the same when it is subject of a transitive clause and when it is subject of an agentive intransitive verb such as "work" or "dance"?	307

Grid	Feature	Feature description	Variant	Variant description	ID
Alignment	Noun: Present Progressive	In present progressive: how is the marking of subject and object of nouns realized?	N:PROG:O=So?	In present progressive: Is the noun form for O the same as for So? I.e.: Does the noun look the same when it is object of a transitive clause and when it is subject of an unaccusative verb such as "fall" or "die"?	308
Alignment	Noun: Present Progressive	In present progressive: how is the marking of subject and object of nouns realized?	N:PROG:Sa=So?	In present progressive: does a noun bear the same case form when it is Sa (subject of e.g., "work") or So (subject of e.g., "fall" or "die")? I.e.: The language does not have a split between stative and active intransitive verbs.	309
Alignment	Pronoun: Simple Past	In present progressive: how is the marking of subject and object of pronouns realized?	P:PST:A=O?	In simple past: Is the pronoun form for A the same as for O? I.e.: Does the pronoun look the same when it is subject of a transitive clause than when it is object of a transitive clause?	310

Grid	Feature	Feature description	Variant	Variant description	ID
Alignment	Pronoun: Simple Past	In present progressive: how is the marking of subject and object of pronouns realized?	P:PST:A=Sa?	In simple past: Is the pronoun form for A the same as for Sa? I.e.: Does the pronoun look the same when it is subject of a transitive clause than when it is subject of an agentive intransitive verb such as "work" or "dance"?	311
Alignment	Pronoun: Simple Past	In present progressive: how is the marking of subject and object of pronouns realized?	P:PST: O=So?	In simple past: Is the pronoun form for O the same as for So? I.e.: Does the pronoun look the same when it is object of a transitive clause than when it is subject of an unaccusative verb such as "fall" or "die"?	312
Alignment	Pronoun: Simple Past	In present progressive: how is the marking of subject and object of pronouns realized?	P:PST:Sa=So?	In simple past: Does the pronoun bear the same case form when it is Sa (subject of e.g. work) or So (subject of e.g. fall or die)? I.e.: There does not exist a split into stative and active intransitive verbs.	313

Grid	Feature	Feature description	Variant	Variant description	ID
Alignment	Pronoun: Present Progressive	In present progressive: how is the marking of subject and object of pronouns realized?	P:PROG:A=O?	In present progressive: Is the pronoun form for A the same as for O? I.e.: Does the pronoun look the same when it is subject of a transitive clause than when it is object of a transitive clause?	314
Alignment	Pronoun: Present Progressive	In present progressive: how is the marking of subject and object of pronouns realized?	P:PROG:A=Sa?	In present progressive:Is the pronoun form for A the same as for Sa? I.e.: Does the pronoun look the same when it is subject of a transitive clause as when it is subject of an agentive intransitive verb such as "work" or "dance"?	315
Alignment	Pronoun: Present Progressive	In present progressive: how is the marking of subject and object of pronouns realized?	P:PROG: O=So?	In present progressive: Is the pronoun form for O the same as for So? I.e.: Does the pronoun look the same when it is object of a transitive clause than when it is subject of an unaccusative verb	316

Grid	Feature	Feature description	Variant	Variant description	ID
				such as "die" or "fall"?	
Alignment	Pronoun: Present Progressive	In present progressive: how is the marking of subject and object of pronouns realized?	P:PROG:Sa=So?	In present progressive: Does a pronoun bear the same case form when it is Sa (subject of e.g. work) or So (subject of e.g. fall or die)? Ie: There does not exist a split into stative and active intransitive verbs.	317
Alignment	Verb: Simple Past	In simple past, how is alignment realized on the verb?	V:PST:A=O?	In simple past: Is the verb affix for A the same as for O? I.e.: Does the verb look the same when it refers to the subject of a transitive clause than when it refers to the object of a transitive clause? If there is no O-marking on the verb, but there is an S-marking, the answer would be no, they do not look the same. (e.g., German, Russian) If there is neither an O, nor an A marking, like in Swedish, the answer	318

Grid	Feature	Feature description	Variant	Variant description	ID
				would be yes, they look the same!	
Alignment	Verb: Simple Past	In simple past, how is alignment realized on the verb?	V:PST:A=Sa?	In simple past: Is the verb affix for A the same as for Sa? I.e.: Does the verb look the same when it refers to subject of a transitive clause than when it refers to subject of an agentive intransitive verb like "work" or "dance"?	319
Alignment	Verb: Simple Past	In simple past, how is alignment realized on the verb?	V:PST:O=So?	In simple past: Is the verb affix for O the same as for So? I.e.: Does the verb look the same when it refers to the object of a transitive clause as when it refers to the subject of an unaccusative verb (such as "fall" or "die")?	320
Alignment	Verb: Simple Past	In simple past, how is alignment realized on the verb?	V:PST:Sa=So?	In simple past: Is the verb affix the same for Sa (subject of e.g. work) as or So (subject of e.g. fall or die)? I.e., does the verb agreement affix look the same regardless of whether the verb is "work" or "die" (as in German:	321

Grid	Feature	Feature description	Variant	Variant description	ID
				"arbeitete-st", "starb-st"). I.e.: There does not exist a split into unaccusative and agentive intransitive verbs.	
Alignment	Verb: Present Progressive	In present progressive, how is alignment realized on the verb?	V:PROG:A=O?	In present progressive: Is the verb affix for A the same as for O? Ie: Does the verb look the same when it refers to the subject of a transitive clause than when it refers to the object of a transitive clause? If there is no O-marking on the verb, but there is an S-marking, the answer would be no, they do not look the same. (e.g. German, Russian) If there is neither an O, nor an A marking like in Swedish, the answer would be yes, they look the same!	322
Alignment	Verb: Present Progressive	In present progressive, how is alignment realized on the verb?	V:PROG:A=Sa?	In present progressive:Is the verb affix for A the same as for Sa? Ie: Does the verb look the same when it	323

Grid	Feature	Feature description	Variant	Variant description	ID
				refers to subject of a transitive clause as when it refers to subject of an agentive intransitive verb such as "work"?	
Alignment	Verb: Present Progressive	In present progressive, how is alignment realized on the verb?	V:PROG: O=So?	In present progressive: Is the verb affix the same for O as for So? Ie: Does the verb look the same when it refers to the object of a transitive clause as when it refers to the subject of an unaccusative verb (such as "fall" or "die")?	324
Alignment	Verb: Present Progressive	In present progressive, how is alignment realized on the verb?	V:PROG:Sa=So?	In present progressive: Is the verb affix the same for Sa (subject of e.g. "work") as or So (subject of e.g. "fall" or "die")? I.e. does the verb agreement affix look the same regardless of whether the verb is "work" or "die" (as in German: arbeite-t, stirb-t). I.e.: There does not exist a split into unaccusative and	325

Grid	Feature	Feature description	Variant	Variant description	ID
				agentive intransitive verbs.	
Alignment	Compare PROG-PAST	What is the marking relation between subject and object in present progressive and simple past?	PROG_So=PAST_So	Does the subject of e.g. die or fall bear the same case in both progressive present and simple past? (the answer for e.g., Megrelian would be no)	326
Alignment	Compare PROG-PAST	What is the marking relation between subject and object in present progressive and simple past?	PROG_A=PAST_O	Does the subject of a transitive verb in the present progressive bear the same case form as the object of a transitive verb in the simple past? (e.g. as in Georgian)	327
Alignment	Compare PROG-PAST	What is the marking relation between subject and object in present progressive and simple past?	PAST_A=PROG_O	Does the subject of a transitive verb in the simple past bear the same case form as the object of a verb in the present progressive? (e.g., Kurdish)	328
Alignment	Reflexive pronoun in transitive clause	What is the alignment of reflexive pronouns?	REFL-ref-A	In a transitive clause, can O be a reflexive which refers back to A (as in English “herself”, Swedish “sig”)?	329

Grid	Feature	Feature description	Variant	Variant description	ID
Alignment	Reflexive pronoun in transitive clause	What is the alignment of reflexive pronouns?	REFL-ref-O	In a transitive clause, can A be a reflexive which refers back to O (as appears to be the case in some Caucasian languages)?	330

Appendix 3.

Multistate characters and reconstructed probabilities of traits at the probability of presence at the root of the tree. Block = ID of multistate character block (for reference in text), Label = descriptive property label, ID = unique trait ID (A = alignment, NM = nominal morphology, T = tense, VM = verbal morphology, WO = word order), Variant (1–4) = Variant of multistate character from DiACL, given as Grid|Feature|Variant (see S2a), Result = reconstructed probability of presence for earhc trait at the protolanguage state.

Block	Label	ID	Variant (1)	Variant (2)	Variant (3)	Variant (4)	Result
1	Present-Past		ALIGNMENT|Compare PROG-PAST|PAST_A=PROG_O	ALIGNMENT|Compare PROG-PAST|PROG_A=PST_O	ALIGNMENT|Compare PROG-PAST|PROG_So=PST_So		
	Present-Past: No marking difference	A1	0	0	0		0.034475
	Present-Past: A marking in Present Progressive and Past	A2	0	0	1		0.765625
	Present-Past: Active-ergative in Present Prog and Past	A3	0	1	1		0.106325
	Present-Past: All systems	A4	1	1	1		
2	Alignment|Noun: Present Progressive		ALIGNMENT|Noun: Present Progressive|N:PROG: O=So?	ALIGNMENT|Noun: Present Progressive|N:PROG:A=O?	ALIGNMENT|Noun: Present Progressive|N:PROG:A=Sa?	ALIGNMENT|Noun: Present Progressive|N:PROG:Sa=So?	0.093575
	Noun, Present progressive: Tripartite	A5	0	0	0	1	0.03415
	Noun, Present progressive:	A6	0	0	1	1	0.631175

Block	Label	ID	Variant (1)	Variant (2)	Variant (3)	Variant (4)	Result
	Nominative-accusative						
	Noun, Present progressive: No marking	A7	1	1	1	1	0.334675
3	Alignment| Noun: Simple Past		ALIGNMENT| Noun: Simple Past| N:PST: O=So?	ALIGNMENT| Noun: Simple Past| N:PST:A=O?	ALIGNMENT| Noun: Simple Past| N:PST:A=Sa?	ALIGNMENT| Noun: Simple Past| N:PST:Sa=So?	
	Noun, Simple past: Tripartite	A8	0	0	0	1	0.041225
	Noun, Simple past: Nominative-accusative	A9	0	0	1	1	0.526275
	Noun, Simple past: Ergative	A10	1	0	0	1	0.122925
	Noun, Simple past: No marking	A11	1	1	1	1	0.309575
4	Alignment| Pronoun: Present Progressive		ALIGNMENT| Pronoun: Present Progressive| P:PROG: O=So?	ALIGNMENT| Pronoun: Present Progressive| P:PROG:A=O?	ALIGNMENT| Pronoun: Present Progressive| P:PROG:A=Sa?	ALIGNMENT| Pronoun: Present Progressive| P:PROG:Sa=So?	
	Pronoun, Present progressive: Nominative-accusative	A13	0	0	1	1	0.90245
	Pronoun, Present progressive: No marking	A14	1	1	1	1	0.09755
5	Alignment|Pronoun: Simple Past		ALIGNMENT| Pronoun: Simple Past| P:PST: O=So?	ALIGNMENT| Pronoun: Simple Past| P:PST:A=O?	ALIGNMENT| Pronoun: Simple Past| P:PST:A=Sa?	ALIGNMENT| Pronoun: Simple Past| P:PST:Sa=So?	
	Pronoun, Simple past: Tripartite	A15	0	0	0	1	0.035225
	Pronoun, Simple past: Nominative-accusative	A16	0	0	1	1	0.77255
	Pronoun, Simple past: Ergative	A18	1	0	0	1	0.11405
	Pronoun, Simple past: No marking	A19	1	1	1	1	0.078175
6	Alignment|Reflexive pronoun in transitive clause, A		ALIGNMENT|Reflexive Pronoun in trans. Clause|REFL-ref-A				

Block	Label	ID	Variant (1)	Variant (2)	Variant (3)	Variant (4)	Result
	Reflexive not with Agent	A20	0				0.04195
	Reflexive with Agent	A21	1				0.95805
7	Alignment|Reflexive pronoun in transitive clause, O		ALIGNMENT|Reflexive Pronoun in trans. Clause|REFL-ref-O				
	Reflexive not with Object	A22	0				0.95445
	Reflexive with Object	A23	1				0.04555
8	Alignment|Verb: Present Progressive		ALIGNMENT| Verb: Present Progressive| V:PROG: O=So?	ALIGNMENT| Verb: Present Progressive| V:PROG:A=O?	ALIGNMENT| Verb: Present Progressive| V:PROG:A=Sa?	ALIGNMENT| Verb: Present Progressive| V:PROG:Sa=So?	
	Verb, Present progressive: Tripartite	A24	0	0	0	1	0.036825
	Verb, Present progressive: Nominative-Accusative	A25	0	0	1	1	0.8777
	Verb, Present progressive: No marking	A26	1	1	1	1	0.05835
9	Alignment|Verb: Simple Past		ALIGNMENT| Verb: Simple Past| V:PST: O=So?	ALIGNMENT| Verb: Simple Past| V:PST:A=O?	ALIGNMENT| Verb: Simple Past| V:PST:A=Sa?	ALIGNMENT| Verb: Simple Past| V:PST:Sa=So?	
	Verb, Simple past: Tripartite	A27	0	0	0	1	0.0794
	Verb, Simple past: Nominative-accusative	A28	0	0	1	1	0.676275
	Verb, Simple past: Ergative	A29	1	0	0	1	0.089625
	Verb, Simple past: No marking	A30	1	1	1	1	0.114575

Résumé

Cet article utilise un model phylogénétique pour reconstruire le système d'alignement de l'indo-européen. Nous utilisons des propriétés morphosyntaxiques catégoriques, qui prennent des états tels que « nominatif-accusatif », « actif-statif » ou « ergatif ». Par une méthode phylogénétique comparative bayésienne, nous déduisons les taux de transition entre les caractères sur la base d'une représentation phylogénétique des langues dans les données. En utilisant ces taux,

nous reconstruisons la probabilité de présence de traits à la racine et aux branches de l'indo-européen. Nous constatons que le système le plus probable pour le proto-indo-européen est un système nominatif-accusatif, avec de faibles probabilités d'ergativité dans les catégories inférieures dans les hiérarchies grammaticales (noms, temps passé). En utilisant un test de signal phylogénétique, nous constatons que les caractères appartenant à des catégories plus élevées dans les hiérarchies grammaticales montrent une plus grande stabilité phylogénétique que les catégories inférieures dans les hiérarchies. Nous examinons nos résultats par rapport aux théories antérieures sur l'alignement proto-indo-européen ainsi qu'à la typologie générale.

Zusammenfassung

In diesem Artikel wird ein phylogenetisches Modell verwendet, um das Alinierungssystem des Indogermanischen zu rekonstruieren. Die Daten liegen dabei in Form morphosyntaktischer Merkmale vor, die Zustände wie „Nominativ-Akkusativ", „Aktiv-Stativ" oder „Ergativ" annehmen. Mittels einer Bayesschen Standardmethode zum phylogenetischen Vergleich schätzen wir zunächst auf der Grundlage einer phylogenetischen Darstellung der in den Daten vorkommenden Sprachen Übergangsraten zwischen den Merkmalswerten ab. Anhand dieser Raten rekonstruieren wir anschließend Wahrscheinlichkeiten für das Vorhandensein der verschiedenen Merkmale sowohl am Wurzelknoten als auch den intermediären Knoten des indogermanischen Stammbaums. Im Ergebnis ergibt sich als das wahrscheinlichste System für das Proto-Indogermanische ein Nominativ-Akkusativ-System, während der neutralen Markierung und der Ergativität geringere Wahrscheinlichkeiten in denjenigen Kategorien zugeordnet werden (Substantive, Vergangenheitstempus), die in üblichen grammatischen Hierarchien niedriger platziert sind. Bei einem Test des phylogenetischen Signals stellen wir fest, dass Merkmale, die sich auf in grammatikalischen Hierarchien höher platzierte Kategorien beziehen, eine größere phylogenetische Stabilität aufweisen als niedriger stehende Kategorien. Diese Ergebnisse werden abschließend in Bezug zu früheren Theorien zur Alinierung im Indogermanischen sowie zur allgemeinen Typologie gestellt.

Address for correspondence

Gerd Carling
Department of Linguistic and Phonetics
Centre for Languages and Literature
Lund University
Box 201
221 00 Lund
Sweden
gerd.carling@ling.lu.se

Co-author information

Chundra Cathcart
Department of Comparative Language Science
University of Zurich
chundra.cathcart@uzh.ch