To appear in *Language*

NP recursion over time: evidence from Indo-European

Manuel Widmer[1], Sandra Auderset[2], Johanna Nichols[3,4], Paul Widmer[1], and Balthasar Bickel[1]

[1]University of Zürich
[2]University of California, Santa Barbara
[3]University of California, Berkeley
[4]Higher School of Economics, Moscow

Abstract

Some languages constrain the recursive embedding of NPs to some specific morphosyntactic types, allowing it for example only with genitives but not with bare juxtaposition. In Indo-European, every type of NP embedding — genitives, adjectivizers, adpositions, head marking, or juxtaposition — is unavailable for syntactic recursion in at least one attested language. In addition, attested pathways of change show that NP types that allow recursion can emerge and disappear in less than 1000 years. This wide-ranging synchronic diversity and its high diachronic dynamics raises the possibility that at many hypothetical times in the history of the family recursive NP embedding could have been lost for all types simultaneously, parallel to what has occasionally been observed elsewhere (Everett 2005, Evans & Levinson 2009). Performing Bayesian phylogenetic analyses on a sample of 55 languages from all branches of Indo-European, we show however that it is extremely unlikely for such a complete loss to ever have occurred. When one or more morphosyntactic types become unavailable for syntactic recursion in an NP, an unconstrained alternative type is very likely to develop in the same language. This suggests that, while diachronic pathways away from NP recursion clearly exist, there is a tendency – perhaps a universal one – to maintain or develop syntactic recursion in NPs. A likely explanation for this evolutionary bias is that recursively embedded phrases are not just an option that languages have (Fitch et al. 2005), but that they are in fact preferred by our processing system.*

1 Introduction

It has often been noted that languages vary in the extent to which they allow syntactic recursion, i.e. the embedding of a phrase of some type within another phrase of the same type, such as [$_{\text{NP}}$ [$_{\text{NP}}$ *my mother*] *'s book*] (e.g. Givón 1979, Mithun 1984, Everett 2005, Heine & Kuteva 2007,

* *Author contributions:* Data analysis: MW (lead), SA, JN, PW, BB; theoretical background: BB; phylogenetic analysis: BB; research design: PW, BB; write-up: MW, SA, JN, PW, BB. *Corresponding authors:* MW (`manuel.widmer@uzh.ch`) and BB (`balthasar.bickel@uzh.ch`). *Acknowledgements:* This research was supported by a grant from the Dean's Office, Philosophical Faculty, University of Zurich. We are grateful to the many colleagues who advised us on data in the Supporting Material, and to the associate editor and the reviewers for helpful comments.

Evans & Levinson 2009, Mithun 2010, Karlsson 2010, Pullum & Scholz 2010, Viti 2015). For example, Russian allows recursive embedding of NPs if they are marked with the genitive but not if they are constructed with an adjectivizer (an affix that lets an embedded NP behave like an adjective morphosyntactically):[1]

(1) Russian

a. *kniga* *mam-y* (*Ivan-a*)
book(F) mother(F)-GEN.SG Ivan(M)-GEN.SG
'(Ivan's) mother's book'

b. (**Ivan-ov-a/Ivan-a*) *mam-in-a* *kniga*
Ivan(M)-ADJZ-SG.F/Ivan(M)-GEN.SG mother(F)-ADJZ-SG.F book(F)
'(Ivan's) mother's book'

Expansion by a second NP *Ivanova* or *Ivana* in 1b is not blocked by a universal principle that shields off adjectivized nouns from further modification, i.e. by some simple, surface-oriented version of a lexical integrity principle (see Ackerman et al. 2011 for discussion): other languages with adjectivizing morphology – indeed languages with morphology cognate with that in 1b – do allow recursive expansion with this kind of NP. This is the case for example in Upper Sorbian (Lötzsch 1965, Corbett 1987):

(2) Upper Sorbian
přez *Mar-in-eje* *maćer-n-u* *smjerć*
through Marja(F)-ADJZ-ACC.SG.F mother(F)-ADJZ-ACC.SG.F death(F).ACC.SG
'through Marja's mother's death'

Thus, the possibility of syntactic recursion is a parameter of variation which is independent of the specific morphological means by which the embedding relation is signaled. Also, recursion can be blocked with specific types of phrases, as in Russian, or it can be blocked across all phrase types, as has been found in the Amazonian language Pirahã (Everett 2005, Futrell et al. 2016).

The source of such variation is most likely diachronic: some structures evolved so as to freely allow syntactic recursion, for example by expanding the usage scope of a construction to any kind of attributive relation. Other structures evolved so as to block recursion, for example by becoming unproductive (e.g. unavailable for many nouns or restricted to two-member expressions) or by developing new dependency requirements (e.g. requiring an independent head as host). A survey of the literature suggests that unconstrained recursion (as in 1a and 2) is far more widespread in the world's languages than constrained recursion (as in 1b or generally in Pirahã). This suggests that languages in general prefer the evolution of structures that allow recursion over the evolution of structures that ban recursion. That is, there is an evolutionary bias that favors what we will speak of generically as MAINTENANCE of recursion, whether through

[1] *Abbreviations not covered by the Leipzig Glossing Rules:* ADJZ = adjectivizer, AI = ablative-instrumental, CO = common (gender in Hittite), DL = dative-locative, EZ = *ezāfe* marker, N = noun (head), N-H = head-marked NP, NP-A = adjectivizer-marked NP, NP-G = genitive-marked NP, NP-P = adposition-marked NP, NP-Ø = NP embedded by juxtaposition, NT = neuter, PTCL = particle.

innovation of a new recursive NP type, expansion of the range of recursion of an existing type, or retention of an existing alternative recursive type.

Here we test this hypothesis for one specific type of phrase, NPs, in the history of one specific family, Indo-European. Indo-European NPs provide an ideal test case for three reasons: First, Indo-European languages are very diverse with regard to their NP structure, which suggests that NP structures are highly dynamic and that there have been many different diachronic trials in their evolution. Second, the history of the language family is relatively well known, so that the evolution of syntactic structures can be explored both qualitatively and quantitatively. Third, individual Indo-European languages constrain recursion in different types of NPs, so that the possibilities of recursion are independent of the specific morphosyntactic form of the NP and independent of any constraints that might attach to this form.

In what follows we first introduce and elaborate the hypothesis, grounding it in theories on the relevance of recursion and hierarchical structure in syntax (Section 2). We then perform a test of the hypothesis (Section 3), using both qualitative and quantitative methods. In the final section (Section 4) we frame our findings in a more global perspective and suggest further research.

2 Theoretical background

2.1 Clearing the terminological thickets

There are few terms in linguistics that have created more confusion than the term 'recursion'. The key distinction is between (i) syntactic recursion in the sense of self-similar embedding, for example, of an NP inside an NP, and (ii) the notion of recursion in mathematics (Fitch 2010). Syntactic recursion relies on, but does not reduce to, mathematical recursion. Recursion in the mathematical sense is a necessary property of any intensionally specified finite system that can generate infinite sets (e.g. the natural numbers) by inductive definition (Fitch 2010, Watamull et al. 2014). This property may or may not characterize human grammar, or individual parts thereof, and it may or may not have evolved specifically for human language (see e.g. Hauser et al. 2002, Pinker & Jackendoff 2005, Perfors et al. 2010, Martins et al. 2015 for various positions here).

Syntactic recursion, as we understand it here, is present whenever a grammar can freely embed a phrase XP inside the same phrase XP, and – unlike recursion in the general mathematical sense – also assigns a structure of embedding with categories and relations that are repeated at each level to the resulting expression: an XP_n is recursively embedded into an XP_{n+1} iff the distributional properties of XP_n and the relationship between XP_n and XP_{n+1} are the same for all n, and there is no grammatical restriction on the value of n.[2] Not all structures are built this way. For example, combining a subject and a verb does not involve syntactic recursion in this sense because there is no identity of XPs across levels. Nor does example 1b result from syntac-

[2] Compare this with a recursive definition in mathematics, for example the common definition of the factorial function as $n! = n(n-1)!$ for $n > 0$, and $n! = 1$ for $n = 0$. This function does not assign any categories and relations, and it terminates in a simple number, not a phrase structure.

tic recursion: it is non-recursive because there is an arbitrary grammatical constraint that sets $n = 1$.

These statements are about what Martins (2012) calls the 'distinctive signature' of recursion, based on observable ('surface') category distributions (an XP is embedded in an XP and not in a YP), observable semantic relations (an embedded XP modifies an embedded XP which in turn modifies a head), and observable constraints on levels (unlimited *n*). As such, our definition of syntactic recursion (like that of, for example, Futrell et al. 2016) is fully orthogonal to the question of how one wishes to formally model syntax. It is always possible to model any syntactic structure with a recursive operation in the broader mathematical sense, combined with some category labeling mechanism (e.g. Everaert et al. 2015). Under such a conception of syntax, our notion of syntactic recursion would need to be defined in terms of label distributions and possible values of *n*, but the empirical issues (the ungrammaticality of 1b vs. the grammaticality of 2) remain the same.

Also, our notion of syntactic recursion is independent of the various kinds of markers that grammars employ when embedding phrases, such as genitives, adjectivizers, linkers, adpositions, complementizers etc. Some of these come with additional category properties, e.g. they assign an adjective or adposition property to the embedded NP: $[_{NP}$ $[_{AP}$ NP-ADJZ] N] or $[_{NP}$ $[_{PP}$ NP P] N]. However, to the extent that the embedded NP itself keeps the same distributional property as the higher NP, this still counts as recursion in our sense. As was illustrated in the discussion of the introductory examples in 1 and 2, the presence of such intervening elements, or indeed any other signal or marker of the embedding relation, is in principle independent of whether or not a phrase can be recursively elaborated. But we will take up this issue again in Section 3 below.

A notion that is sometimes taken as indicative of recursion (in either the syntactic or the mathematical sense) is hierarchical structure and embedding relations, but this is misleading. A structure is hierarchical if it satisfies the definition of an undirected acyclic graph with a distinct root, also known as a rooted tree (Fitch 2014). Grammars that generate hierarchical structure may or may not include recursion (Fitch 2010, Perfors et al. 2010, Tiede & Stout 2010, Martins 2012, Fitch & Friederici 2012, Martins et al. 2015). For example, a hierarchical structure may have limited depth, as is the case in our initial observation 1b. More generally, hierarchical structures can be stipulated by declaring one-off 'embedding' relations, for example in the form of fixed constructional schemas or templates. What syntactic recursion yields beyond such templates is unlimited expansion in embedding depths, and, equally important, recursion guarantees identity of categories across levels: while both the rule pair 'embed B in A' and 'embed C in B' and the single recursive rule 'embed α in α (with α =: {A,B,C})' can generate the hierarchical structure [A[B[C]]], only the recursive rule assigns the same category label α to all three constituents (Martins 2012).

2.2 Recursion as a processing preference

Despite this definitional independence, we contend that there is a natural link between the faculties for mathematical recursion and hierarchical structure on the one hand, and the presence of syntactic recursion in languages on the other hand. This link is established via the role of processing in language change. In brief, we propose that grammars with syntactic recursion are preferred (but not required) by the processing system, and that this preference establishes an evolutionary bias so that grammars with syntactic recursion are more likely to develop and persist over time than grammars without syntactic recursion. We elaborate the motivation for this proposal below, but we first clarify how it differs from other proposals.

As noted above, a prominent alternative proposal for linking the mathematical and the syntactic concepts of recursion is to build recursion in the mathematical sense directly into the very foundation of syntax (e.g. Hauser et al. 2002). This is usually motivated by assuming that syntactic structure must allow infinite — and thus recursively defined — counting of embedding levels in unbounded dependencies (Chomsky 1957, 1975).[3] Such a model of grammar entails that unconstrained syntactic recursion (in the sense defined above) is necessarily a universal *option* for human language (Fitch et al. 2005, Watamull et al. 2014). But the model makes no prediction on the *distribution* of syntactic recursion over time, over space or over individual phrase types (nor on the use of recursion in discourse, for that matter). One would expect random developments here that are only tied to the vagaries of lexical evolution, such as the various ways in which complementizers or genitives come and go. By contrast, our proposal predicts constraints on these developments, favoring phrase types that allow syntactic recursion over those that do not.

The motivation for our proposal is as follows. It has long been proposed and experimentally substantiated that the human processing system not only has a faculty for hierarchical structure, but the system actually prefers hierarchical over serial structures when confronted with a string of symbols (e.g. Miller 1967, Pallier et al. 2011, Fitch 2014, Ding et al. 2016, Christiansen & Chater 2016). In language, this human DENDROPHILIA (as Fitch 2014 terms it) is most efficiently satisfied by the use of syntactic recursion. In other words, increased use of syntactic recursion in grammar makes the processing of the favored type of structure, viz. hierarchical structure, more efficient. Evidence for this claim comes from two observations: First, syntactic recursion makes it straightforward to assign the same properties to different syntactic units: rather than stipulating by an extra rule that the higher NP in [$_{\text{NP}}$ [$_{\text{NP}}$ *my uncle*]*'s house*] has the same distributional ('N') properties as the lower NP, a recursive grammar provides this information about self-similarity for free, and indeed guarantees it (Martins 2012). Second, some amount of syntactic recursion seems to lead to the computationally best tradeoff between formal grammar complexity and data coverage in language acquisition (Perfors et al. 2010). Taken together, these two observations suggest that syntactic recursion makes a grammar computationally simpler, and it is likely that the processing system generally prefers simpler systems (e.g. Bornkessel-Schlesewsky & Schlesewsky 2009).[4]

[3] The alternative, which often goes unnoticed, is to dispute the claim that infinite counting in syntax is necessary for linguistic creativity (Pullum & Scholz 2010, Pullum 2013) and/or for sufficient information carrying capacity (Kornai 2014). This then suggests models of grammar that stipulate constraints on structures without presupposing anything about the finiteness of the set of structures or expressions in a language.

[4] A reviewer asks why one couldn't simply claim that expressing ideas like 'Ivan's mother's book' would be complicated without syntactic recursion. The problem is that, when used in some real context, there are many ways in which one can easily express such ideas without recursive NP embedding, for example: *Let's read this book. It was written by Ivan's mother.* Distributing information in this way is very common in actual discourse and is not

Independently of this, there is good reason to assume that the human brain indeed prefers to apply its general faculty for recursion when processing syntax. This means that syntactic recursion is likely to be preferred not only for its specific computational benefits when processing hierarchies, but also because of a much broader and more general bias. Initial support for this comes from the sheer ubiquity of recursion outside syntax: for example, in visual cognition (Pinker & Jackendoff 2005, Martins et al. 2015), where recursion is a key mechanism in pattern detection; in spatial navigation or kinship calculation, where recursion allows predicting the properties of unseen entities (as the same, or as bearing the same relation; Martins 2012); and in theory-of-mind cognition, where recursion enables us to think from the perspective of another person, place, or time (Corballis 2011) as well as the attribution of mental states about mental state attribution (Grice 1975, Tomasello 2008). In all these areas, it seems that recursion brings about massive cognitive benefits with limited resource pressure. Initial evidence for such benefits has recently been established through fMRI studies of visual processing (Fischmeister et al. In press). These findings make it likely that recursion is put to extensive use also in other areas of higher cognition, such as the construction of syntactic phrases.

Given these observations, we expect the following hypothesis to be true:

(3) Hypothesis:
All else being equal, for any kind of syntactic phrase (e.g. NPs), the human processing system prefers a grammar that includes recursion as a structure-building operation over a grammar that does not include recursion, as detectable through the distinctive signatures of syntactic recursion (i.e. identity of categories and relations across embedding levels and no grammatical constraint on the number of embedding levels)

As in other such cases, a processing-based hypothesis like the one in 3 makes a prediction about the probabilities with which syntactic structures evolve in language change (e.g. Hawkins 1994, 2014, Blevins 2004, Croft 2003, Christiansen & Chater 2008, Kemmerer 2012, Bickel 2015, Bickel et al. 2015, McDaniel et al. 2015): at any point in time, given the choice between a non-recursive, limited-depth grammar of, say, NP construction, and an unconstrained, recursive grammar, the processing system will (according to the hypothesis) tend to apply the recursive one. In many cases, there is no choice (i.e. not all else is equal): only one type of NP construction may be available for a given meaning, for many different reasons (e.g. it might not be possible to express possession with adjectivally marked NP embedding in a given language). But to the extent that the choice arises (e.g. *American territory* vs. *territory of America*), and there is a contrast in recursiveness, the system will favor the one allowing recursion. Given enough such occasions over time, and absent any sociolinguistic constraints, recursive grammars are then expected to gain ground. Such an evolutionary scenario is entirely parallel to established theories that predict for example the phonologization of final devoicing because of its energy-saving aspects (Hyman 1976, Lindblom et al. 1995, Boersma 1998, Blevins 2004): languages tend to develop and maintain final devoicing unless the change is blocked by some other process (e.g. sociolinguistic pressure against language change, contact effects favoring another structure, etc.).

particularly complicated – indeed probably even simpler. See Pullum & Scholz (2010) and Kornai (2014) for the general point that a lack of syntactic recursion does not in any way interfere with expressive power.

Given this, the hypothesis in 3 predicts for any surveyed lineage that, whenever a language develops constraints on recursion in some phrase, the possibility of recursion is likely to be restored over time or an unconstrained alternative type of the same phrase is likely to expand its range of use, i.e. apply to more contexts and become the preferred type over time. For example, recursion of NPs with genitives might become blocked for some reason: genitives might become strongly associated in frequency with two-word idioms, or they become unavailable for many nouns on phonological grounds, or the entire construction loses popularity in the wake of language contact, etc. If any such development takes place, we predict a high probability that the original type will be restored (e.g. by developing a new genitive that allows recursion) or that an alternative type, for example, a type involving adjectival morphology, will be expanded in its use and become fully available for recursion. As a result of this, we expect that for each phrase (e.g. NPs, or clauses), it is very likely that there will be at least one type that allows syntactic recursion at any given time.

The hypothesis will be falsified if recursion, when lost from some phrase type, is not restored either by renewing the original pattern or by extending another. In that case, the distribution of syntactic recursion will result from a pure chance process, perhaps coupled with factors from language contact or other local patterns, but without any systematic, universal bias. As a result, many languages are expected to develop in the way claimed for Pirahã (Everett 2005), and have no phrase type that freely allows recursion.

For the hypothesis to be testable, a variety of different phrase types needs to be able to develop, and there needs to be at least some overlap in what these types can express, i.e. the types should not be completely functionally distinct. If the types are fully distinct, the processing system cannot freely choose between them, and the system could not even start to favor one or the other. In our case study below, we therefore first establish that a range of different types developed, and that each of these types became the dominant or default structure for recursion in at least one language. This is only possible if the types share enough functional ground.

Another requirement for testing the hypothesis is diachronies that are sufficiently unstable, i.e. structures must come and go within known or at least reconstructable time at least once, and ideally many times.[5] Only then can we sample transitions and assess whether overall these transitions lead to syntactic recursion significantly more often than not. In our case study, we will therefore first establish that each type is sufficiently dynamic to provide a test case.

For a full test of the hypothesis, we need an extensive worldwide sample of the (reconstructed) diachrony of various phrases, so that we can compare the number of diachronies in line with our hypothesis and the number of diachronies in conflict with our hypothesis. No database of such breadth and scope is available at present. Instead, we focus on one phrase and one family: NPs in Indo-European. Here we can survey a substantial range of diachronies, i.e. we can test the hypothesis against a relatively large sample of individual diachronic transitions and thereby gain initial evidence for or against the hypothesis.

[5] The technically minimal requirement is that the recursive and non-recursive states are strongly connected, i.e. that no transition between these two states ever has the probability of exactly 1 or 0 (Greenberg 1995, Maslova 2000). But for sampling purposes, transition probabilities close to 0 are already problematic.

3 NPs in Indo-European as a test case

As noted above, there are two preconditions for our hypothesis to show any effects in a language family: first, several different morphosyntactic types of NPs (i.e. different ways of marking or establishing the embedding of an NP in an NP) need to be available and overlap in their functions. This makes it possible for speakers to choose one type and use it for recursion in at least some contexts. We therefore first (Section 3.1) survey the range of morphosyntactic types that is attested across Indo-European (IE) languages and show that each type has become the dominant choice for NP recursion in at least one language. This proves that there is no type that is intrinsically constrained to specific functions and could never be chosen as an alternative. We show this in detail in Section 3.2, providing additional evidence for functional overlap between in the types in the course of their history. The second precondition is that the range of types must be sufficiently dynamic over time. This enables us to trace transitions between types and test our prediction that there are significantly more transitions that make NP recursion available than transitions that make NP recursion unavailable. We demonstrate this in Section 3.3 through attested or reconstructable transitions.

Once this is established, we summarize the distribution of the types that are vs. are not available for recursion and report evolutionary biases based on Bayesian phylogenetic analyses (Section 3.4).

3.1 A survey of NP types

Across the family, there are at least five morphosyntactic types of how the embedding of an NP in an NP is marked or established. All five types are prominent means for syntactic recursion in one or more branches: genitives, adjectivizers, head marking, adpositions and juxtaposition. Each of these types has become the dominant or a particularly popular structure in at least one daughter language. We analyze an NP type as being available for syntactic recursion if the type can indicate the embedding of an NP that is in turn modified by another NP. For example, the genitive type is available for recursion if genitives can mark an NP that contains another NP (of any type). This is true of the genitive *mamy* 'mother's' in Russian (1a) or of English *'s*-genitives like *mother's* (cf. *John's mother's book*, *my mother's book* etc.), but not of adjectivizer-marked expressions like *mamina* 'mother's' in Russian (1b).

In the following we briefly survey the five types that are prominent in Indo-European. For detailed philological discussion of all data, and justification of our analyses, see Supporting Material 1.

Genitives. We define genitives as dependent markers that behave as case markers, phonologically hosted by stems or entire phrases (e.g. in the case of English *'s*). Unlike adpositions, genitives are not independent words with a distinct part of speech property and argument structure, and as a result they do not assign case. Unlike adjectivizers, genitives do not add an adjectival property to the embedded NP and therefore do not for example show any agreement. We notate embedding by means of genitives as [[NP-G] N]. Genitives are the only fully productive strategy for recursive NP embedding in Hittite, a long-extinct language of the Anatolian branch:[6]

(4) Hittite
^{D}*IŠTAR-aš lūli-aš* KÁ.GAL-*az*
Ishtar-GEN pond-GEN door-ABL
[[[NP-G] NP-G] N]
'from the door of Ishtar's pond' (KBo XVI 49 I 6) (Yoshida 1987:19)

Adjectivizers. Adjectivizers embed an NP by adding a morphosyntactic adjective property, such as agreement, to it. Often this is a lexically self-contained process that can only apply to noun stems, but, as we noted in the introductory example 2 from Upper Sorbian, adjectivization can be used for embedding full NPs as well, i.e. NPs with their own NP subconstituents. We notate adjectivization as [[NP-A] N]. This pattern has become the only fully productive way of NP embedding in a sister language of Hittite, Luwian. The following examples are from Hieroglyphic Luwian (henceforth 'H Luwian'), attested through inscriptions involving logograms.[7]

(5) H Luwian
a. *Tuwana-wanni-s*(URBS) |REX*ti-s*
Tuwana(CO)-ADJZ-NOM.SG.CO king(CO)-NOM.SG
[[NP-A] N]
'the king of Tuwana (city)' (BOR §1) (Bauer 2014:151)

b. [*a*]*wa=ta* |*z*[*ati*] *ámi*$_i$ *áláyaza*$_i$*-ss-an* *HÁ+LI-ass-an*
PTCL=PTCL this.DL.SG.NT 1SG.POSS.DL.NT Arrayaza(CO)-ADJZ-DL.SG Hattusili(CO)-ADJZ.DL.SG
[[NP-A] [NP-A] [[NP-A]
SERVUS*la-ya*$_i$ STATUA*rut-i* OVIS(ANIMAL)*ti* PRAE*i* (*69)*sasa-tu*
servant(CO)-ADJZ.DL.SG statue(NT)-DL.SG sheep(CO)-AI.SG ADV present-3PL.IMP
NP-A] N]
'Let them present(?) the statue of me, of Arrayazza, of the servant of Hattusili, with a sheep.' (MALPINAR §5) (Bauer 2014:148)

The basic construction is shown in 5a. The sentence in 5b contains an NP with recursive embedding: the head noun STATUA*ruti* 'statue' is modified by the conjoined adjectivizer-marked *áláyazassa* 'of Arrayazza' and SERVUS*laya* 'of the servant' (and both are in apposition to the initial possessive pronoun *ámi*). The NP SERVUS*laya* 'of the servant' is in turn recursively modified by another possessive adjective NP *HÁ+LI-assa* 'of Hattusili'.

[6] The Hittite writing system is a mixture of logograms and phonetic forms. Words can be written either by a logogram or a phonetic form, or as a logogram with a phonetic component. Logograms may also stand before phonetic words as a semantic marker dubbed 'determinative'. By convention, logograms are rendered in capitals if they are Sumerian and in italic capitals, if they are Akkadian. Determinatives are rendered as superscripts.

[7] Orthographic conventions are similar to those for Hittite, except that logograms are conventionally rendered as their capitalized Latin counterpart or their capitalized Anatolian spelling, if known. Determinatives are put in brackets. Special symbols include '|' for word divider and square brackets for broken signs (Bauer 2014:xiii).

A variant of adjectivizers assigns case to the embedded NP. This variant has become the dominant construction in Hindi and many other Indo-Aryan languages. The embedded NP is marked by a clitic adjectivizer *=k-*, which takes a further suffix agreeing in case, gender and number with the head noun and which additionally assigns oblique case to its host (although case is visible only in some noun classes):

(6) Hindi
Khannā=k-ī *bahin=k-e* *kutt-e=k-ā*
Khanna(M)=ADJZ-OBL.SG.F sister(F).OBL.SG=ADJZ-OBL.SG.M dog(M)-OBL.SG=ADJZ-NOM.SG.M
[[[[NP-A] NP-A] NP-A]
nām
name(M).NOM.SG
N
'Khanna's sister's dog's name' (Snell & Weightman 2003:66)

A similar construction is attested in Albanian, where an adjectivizing "particle of concord" assigns dative case to its host (Newmark et al. 1982:159-162). We do not analyze the Albanian and Hindi constructions as a type of their own, distinct from other adjectivizers. The key feature that marks the embedding relation is that the adjectivizer (*=k-* in Hindi) assigns a morphosyntactic adjective property to the embedded NP, thereby requiring agreement. The oblique case is only a side-effect that results from the etymology of the adjectivizer.

Adjectivizers are very frequent in Indo-European with embedded pronouns. Instead of genitives one often finds special possessive pronouns that agree in case, number and gender with their head (e.g. German *sein-en Brüder-n* his-DAT.PL.M brother(M)-DAT.PL). However, when this strategy is limited to pronouns (as it is for example in German), it is functionally specialized for this, and does not offer the open choice for NP embedding that our hypothesis seeks. We therefore exclude possessive pronouns from our survey.

Head marking. In this type, the embedding relation is marked on the head, unlike in all other types discussed so far. We notate this as [[NP] N-H]. The type is prominent for example in Persian, where the relevant marker is known as the *ezāfe* marker:

(7) Persian
ketāb-e pedar-e Hasan
book-EZ father-EZ Hasan
[N-H [N-H [NP]]]
'the book of Hasan's father' (Lazard 1992:67)

The head-marking type also dominates several Germanic languages, where it involves possessive pronouns or particles derived from this. An example is Afrikaans:

(8) Afrikaans
ons bur-e se vriend-e se seun
our neighbor-PL POSS friend-PL POSS son
[[[NP] N-H] N-H]
'our neighbors' friends' son' (Donaldson 1993:98)

A variant of this, popular for example in Swiss German, involves additional dative case marking on the dependent. This results in double marking and is illustrated by the dative case on the dependent NPs (*er Anna* 'Anna' and *irem Brueder* 'her brother') in the following example (for a parallel development in Ossetic, see Section S12.5 in Supporting Material 1).

(9) Swiss German (Bernese dialect)

er	*Anna*	*ir-em*	*Brueder*	*si-s*	*Huus*
ART.DAT.SG.F	Anna(F)	3SG.F.POSS-DAT.SG.M	brother(M)	3SG.M.POSS-NOM.SG.NT	house(NT)
[[[NP	]	N-H	]	N-H	]

'Anna's brother's house'

We do not distinguish this type of double marking from simple head marking (Nichols 1992, Nichols & Bickel 2005) for current purposes. What sets the Swiss German construction apart from the other types is the appearance of a head marker. The dependent marking component by means of case continues a standard pattern.

Adpositions. In Indo-European languages, adpositions are distinct syntactic words which are relatively independent of their host. As such, they typically govern case and can sometimes be stranded. English *of* is a prominent example and it is one of the dominant strategies for recursive embedding in the language. We notate adposition structures as [[NP-P] N]. Apart from English and other Germanic languages, the pattern has also become the strategy of choice in several Romance languages, e.g. in Spanish:

(10) Spanish

el	*perro*	*de-l*	*padre*	*de*	*mi*	*amigo*
ART.SG.M	dog(M)	of-ART.SG.M	father(M)	of	1SG.POSS	friend(M)
[	N	[NP-P		[NP-P	[NP-G]	]]]

'the dog of my friend's father' (Martha Mariani, p.c.)

As case only survives in pronouns in Spanish, the fact that *de* is an adposition and not a phrasal case affix is evident only when *de* occurs in combination with pronouns (cf. *Hablan de mí* 'They speak of me', where the first person singular pronouns stands in a "prepositional case", which only occurs in combination with adpositions).

Juxtaposition. Another frequent type of NP is formed by endocentric (also known as determinative) juxtaposition, without any dedicated marking. Juxtaposition is traditionally distinguished from compounding. However, this distinction is most often only made on the basis of phonological criteria and is as such independent of the syntactic structure, so we group the two together. Another distinction is sometimes made between two morphosyntactic types: juxtaposition of a fully-fledged NP (which for example admits further morphology, articles, and other modifiers) and a head noun, vs. compounding as a process of stringing together stems into words. While important on its own, the distinction is immaterial for our purposes as long as the resulting string is built up recursively: a modifying N or NP modifies an N or NP that again modifies an N or NP. We notate these structures here generally as [[NP-∅] N], assuming phrasal analyses throughout: $[_{NP}\ [_{NP}\ [_{NP}$ N] N] N]. Alternatively one could analyze these structures as involving stems and incorporation (or movement) of each embedded element into its next higher head [[[N-]N-]N] (where the hyphen indicates incorporation). As long as the incorporation analysis allows for recursive embedding and there is no marking of the embedding relation, the two analyses are equivalent for our purposes.

Juxtaposition is highly popular in Germanic, e.g. German or English:

(11) German
Auto- *reifen-* *wechsel*
car(NT) tire(M) replacement(M)
[[[NP-Ø] NP-Ø] N]
'car tire replacement'

Where it occurs, juxtaposition tends to favor embedding of the same unmarked type (here the complex NP [[*Auto-*]*reifen*]), but several languages also allow recursion with other types as well. In Vedic Sanskrit, for example, a juxtaposition member can consist of an NP with an embedded genitive-marked NP (a pattern also observed in Pāli and Avestan, cf. Sections S11.7 and and S12.1 in Supporting Material 1, respectively).

(12) Vedic Sanskrit
árvato *māṃsa-* *bhikṣā́m*
horse(F).GEN.SG meat(NT)- request(F).ACC.SG
[[[NP-G] NP-Ø] N]
'the request for the meat of the steed (s.c. the aforementioned horse which is being cooked during the horse sacrifice)' (Rig Veda 1.162.12c) (Wackernagel 1905:31)

The most flexible version of juxtaposition is one where fully-fledged NPs can be recursively stacked. This is the type that is popular in several modern Celtic languages, e.g. in Modern Breton:

(13) Modern Breton
pneuioù *marc'h-houarn* *glas* *ma* *mignon* *gwellañ*
tire(M).PL horse(M)-iron(M) blue 1SG.GEN friend(M) best
[N [NP-Ø [[NP-G] NP-Ø]]]
'the tires of the blue bicycle of my best friend' (Herve Le Bihan, p.c.)

Note that for juxtaposition to be fully comparable with the other NP types discussed here, it needs to involve recursion in the sense defined above: an embedded NP is embedded into another NP which in turn is embedded into another NP. The literature on juxtaposition, especially when focusing on 'compounding', often adopts a broader notion of recursion that applies to all cases where compounds are members of compounds, as when for example a noun modifies a compound noun as a whole (e.g. [*student* [*film award*]]). A recursive version of this in our sense requires that the embedded noun *film* or *student* is modified by another modifier noun, e.g. [[[*action*] *film*]] *award*] or [[[*bachelor*] *student*] [*film award*]].

Also, we exclude from our purview exocentric juxtaposition (*bahuvrīhi* compounds, cf. Vedic Sanskrit *marútas rúkma-vakṣasas* [Marut.NOM.PL decoration-chest.ADJZ.NOM.PL] 'Maruts (a class of gods) with decorations on their breasts'): in most cases it is the embedded juxtaposition as a whole that assumes the function of a modifier (e.g. [*marútas* [*rúkma-vakṣasas*]]). Furthermore these juxtapositions draw on a heterogeneous set of [X N] and [X V] constructions that admit a broad variety of incorporated modifiers, very often adverb-like in function (cf. Vedic Sanskrit *raghu-yā̆-man-* [rapid-go-NMLZ-] 'with a rapid course', English *white-washed wall*, *cross-sectional study*).

3.2 Functional overlap

The survey in the previous section shows that each of the five NP types is the dominant, or even the sole type that is available for recursion in at least one daughter language. This suggests that there is no family-wide limitation that intrinsically blocks any of the types from taking over the semantic domain that is covered by another type. For such take-overs to be possible in history, the functions of the types must overlap at least to some extent.

For some specific pairs of types, functional overlap can even be directly observed in synchronic data or in well-established reconstructions.

The overlap between genitives and adpositions, for example, can be observed in synchronic data from several Germanic languages. English is a case in point and has alternations like *the office of our administrator* and *our administrator's office* which differ only very minimally in meaning. The overlap between genitives and juxtaposition is well-established in several Celtic languages. The two types alternate freely in Middle Breton for example:

(14) Middle Breton (de la Haye & Gueguen 2002)

a. *oa cousquet en Sacristery a Lantreguer*
be.IMPF.3SG sleep.PTCP in sacristy(F) GEN Lantreguer
'he was sleeping in the sacristy of Lantreguer' (79.9–11)

b. *cousque en Sacristery Lantreguer*
sleep.IMPF.3SG in sacristy(F) Lantreguer
'he used to sleep in the sacristy of Lantreguer' (21.4–5)

The functional overlap between genitives and adjectivizers reconstructs for several nodes in the tree. One case is Anatolian, where in one daughter language (Hittite), the inherited Proto-Indo-European genitive (**-s* and its declension-class allomorphs; Fortson 2010:126) took over the entire functional domain of NPs, while in another daughter language (H Luwian), the inherited Proto-Indo-European adjectivizer (*-(i)yo*; Fortson 2010:134f) took over the same domain. This presupposes that the two types overlapped in this domain. A parallel case is Indo-Aryan, where some adjectivizers and genitives reconstruct to one and the same etymon: the Hindi adjectivizer *=k-*, illustrated in 6 above, derives from the same participial form of *kr̥-* 'do' as the Nepali genitive *-ko*, illustrated in 20 below. The etymology is well-established (Masica 1991:243), and we briefly present evidence for it in Section 3.3.

The overlap between genitives and adjectivizers is likely to reconstruct even to the Proto-Indo-European level since **-s*-genitives and *-(i)yo-*-adjectivizers seem to have been treated as equivalent in at least some contexts. Evidence for this comes from the *o*-stem genitive allomorph **-syo* and the adjectivizer suffixes Tocharian *-ṣṣe-* and Luwian/Hittite *-assa-*. If Anatolian is any indication of an inherited *o*-stem genitive **-s* (e.g. Hittite *atta-s* nom./gen. 'father'/'of the father'), the ending **-syo*, which is attested in Armenian, Italic, Indo-Iranian, and Greek, is likely to represent a combination of this genitive **-s* with **-(i)yo*, which presumably had a broad attributive function (Fortson 2010:134). Also, both the adjectivizer Tocharian B *-ṣṣe-* and Luwian/Hittite **-assa-* can be traced back to **-syo-*. The only difference between these adjectivizers and the genitive in **-syo* is that the adjectivizer agrees with the head of the NP, whereas the genitive does not. All of this suggests an intimate functional and formal relation between the *o*-stem genitives in **-s* and **-syo*, the adjectivizer **-(i)yo*, and also the linking particle/relative pronoun **yo* at least at some diachronic stage.

Finally, the functional overlap between adjectivizers and head markers is reconstructable in Iranian. Here, the Persian head marker (*ezāfe*) *-e* that was illustrated in 7 derives from the same etymon that gave rise to an adjectivizer in another daughter language, Northern Kurdish (Haig 2011:367):

(15) Northern Kurdish

Tu	*kijan*	*hesp-î*	*di-bîn-î?*	*– Yê*	*Soro.*
2SG	which	horse(M)-OBL	IND-see.PRS-2SG	EZ.M	Soro(M)

'Which horse did you see?' – 'Soro's'

The marker *yê* is cognate with the Persian *ezāfe*, but *yê* behaves like an adjectivizer: it agrees with the head noun (here in masculine gender), and it is a co-constituent of the dependent, as evidenced by the fact that [*yê Soro*] constitutes a full independent NP. In Persian, an answer to 'which horse' cannot be **e Soro*, but rather has to be *māl-e Soro*. This is because *-e* is a head marker and not an adjectivizer and, accordingly, always has to be attached to a head constituent. In the absence of an overt head noun, the *ezāfe* is thus attached to the semantically unspecific head constituent *māl* 'property' instead (Windfuhr & Perry 2009:435).

A parallel overlap between adjectivizers and head markers is likely to reconstruct also at some higher node in the family, predating at the least the split between Balto-Slavic and Indo-Iranian. Evidence comes from the fact that the Iranian *ezāfe* is partly cognate with the Balto-Slavic adjectivizing suffix *-ji/o-*. In Modern Lithuanian, the definite adjective is formed by suffixing the inflected marker *-ji-* to the fully inflected stem of the adjective as in 16a, where the adjective *ger-ą* as well as the definite marker *-jį* agree in gender, number and case with the head of the NP. In Old Lithuanian this marker was also used as an adjectivizer. In 16b for example the adjectivizing *-ia*, which agrees with the head *diewa*, serves to mark the locative *danguie* 'in heaven' as being embedded in an NP.

(16) a. Modern Lithuanian

gẽr-ą-jį	*šùn-į*
good-ACC.SG.M-DEF.ACC.SG.M	dog(M)-ACC.SG

'the good dog'

b. Old Lithuanian

pon-a	*diew-a*	*musu*	*dangu-ie-ia*
lord(M)-GEN.SG	god(M)-GEN.SG	1PL.GEN	heaven(M)-LOC.SG-ADJZ.GEN.SG.M

'of our lord, god in heaven' (Petit 2009:318)

3.3 The dynamics of NP types

None of the five types surveyed is very stable historically: for each type, we found at least one case in which it developed or declined in historical time, with attested or reconstructable stages. When specific types disappear, this is in most cases the result of morphological decay or of reanalysis. For example, a genitive decays by becoming unproductive or eroded; or adpositions are reanalyzed as new case markers. When types are innovated, this mostly results from expansion in use or again from reanalysis. For example, juxtaposition becomes more popular and unconstrained, or a pronominal element is a reanalyzed as a head marker. Table 1 lists one attested or reconstructable example for each transition. We discuss each examples in what follows.

Type	unavailable ≻ available	available ≻ unavailable
Genitives	Breton *a*	H Luwian GEN ≻ constrained
Adjectivizers	Hindi *kā*	Nepali ADJZ *kā* ≻ GEN
Head markers	Young Avestan *ezāfe*	Ossetic *ezāfe* ≻ constrained
Adpositions	Middle Persian (various)	Modern Persian (reanalyzed as case)
Juxtaposition	Icelandic	Nepali

Table 1: Attested or reconstructed transitions between availability and non-availability of an NP type for recursion, with one example each.

Genitives. Older stages of Insular Celtic have preserved the Proto-Indo-European genitive, but in Breton new genitival phrasal prefixes have developed from prepositions. The marker does not govern any case and is syntactically completely dependent on its host NP, from which it cannot be stranded:

(17) Modern Breton
ur plac'hig yaouank a-n oad a bemzek vloaz
a girl(F) young GEN-ART age(M) GEN 15 year(M)
[N [NP-G [NP-G]]]
'a young girl of the age of fifteen years' (Hemon 1976–1998:34)

The affixal status of the marker is particularly evident when it occurs with personal pronouns. In this context, *a* fused completely with a special form of the pronoun and neither can occur without the other, as in, for example, *anez-aff* [GEN-1SG] 'of me, mine'.

Loss and re-innovation of genitives is not the only option. In H Luwian for example, the genitive did not disappear but became limited to single-level constructions. There are attested cases of NPs containing a single embedded genitival NP, but no cases are attested with recursive embedding:

(18) H Luwian
VITIS-*si* FINES-*s*
vineyard(NT)-GEN.SG border(CO)-NOM.SG
'the border of the vineyard' (Babylon 1 §6, Hawkins 2000:392, Bauer 2014:143)

The H Luwian corpus is limited in size, but it is dominated by possessive relations, often very complex ones (Bauer 2014:132), and so one would expect recursively expanded genitival NPs to show up if they were possible. What H Luwian uses for recursive NP embedding is adjectivizers of the kind illustrated earlier by example 5.

This process of gradual reductions of genitives is frequent elsewhere in the family. When the genitive was lost in Western Europe it was typically replaced by adpositions; when it was lost in in Indo-Iranian it was typically replaced by adjectivizers (Indo-Aryan) and head-marking constructions (Iranian).

Adjectivizers. Wherever they are available for recursive embedding, adjectivizers are innovations. An example is Hindi *=k-*, which (as noted above) derives from a participle based on a root *kr̥-* 'do' (Masica 1991:243). Intermediate stages are attested (Bubenik 1998, Reinöhl 2016). Consider the following 12th century example where the *kr̥*-form has two possible analyses (separated by a pipe operator '|' in the interlinear gloss):

(19) a. Middle Indo-Aryan (12th century CE)

kesari jasu keraeṃ huṃkāraḍaeṃ muhahuṃ paḍanti
lion(M).NOM.SG REL.GEN.SG.M (do.PTCP|ADJZ).INS.SG roaring.INS.SG mouth(NT).GEN.PL fall.3PL
tr̥ṇāiṃ (Hemacandra 8.4.422; Bubenik 1998:75–76)
grass(NT).NOM.PL

b. Modern Hindi translation

vah śer jis=k-ī garaj-se tum-hār-e
DEM.NOM.SG lion(M).NOM.SG REL.OBL.SG=ADJZ-OBL.SG.F roar(F).OBL.SG-from 2SG-ADJZ-OBL.SG.M
mūh-se khānā gir gayā thā
mouth(M).OBL.SG-from food(M).NOM.SG fall go.PTCP.NOM.SG.M be.PST.SG.M
'The lion by whose roaring grass fell from your mouth.' (Reinöhl 2016)

The form *keraeṃ* is ambiguous between a literal translation as a participial form 'by whom the roaring was made/done' and a reanalysis as an adjectivizer that embeds the relative pronoun *jasu* ('whose roaring, roaring of whom'). In the Hindi translation (19b) the reanalysis is completed.

In several Indo-Aryan languages, adjectivization became unavailable again because the *k*-markers were further reanalyzed as plain genitives, shedding all agreement, government, or other adjective properties. The transition is still ongoing at present in Nepali, where more innovative dialects (especially in eastern Nepal, with many Tibeto-Burman L2-speakers), have completely lost the agreement options (see Section S10.7 in Supporting Material 1):

(20) Nepali

us-ko sāthi-ko didi-ko ghar
3SG-GEN friend-GEN elder.sister-GEN house
[[[[NP-G] NP-G] NP-G] N]
'his friend's elder sister's house' (fieldnotes, B. Bickel)

In more conservative varieties and in written Nepali, the *k*-marker would show agreement in gender (*sāthi-kī didi* 'friend's elder sister'). In all varieties, however, the *k*-marker has already lost one of its original adjectivizing properties: the marker is no longer able to embed adpositional phrases: while Hindi allows constructions like *mez-par-k-ī kitāb* [table-on-ADJZ-F.SG book] 'the book on the table' (Verma 1971:146), this use of the *k*-marker is not possible in Nepali (and so **ṭebul-mā-ko kitāb*, with *-mā* 'on, at, in', is ungrammatical; Narayan Gautam Sharma, p.c.).

Head markers. All head-marking NP constructions in Indo-European are innovated. They all draw on pronominal elements, either via possessive pronouns (e.g. Afrikaans and other Germanic languages) or via anaphoric and/or relative pronouns (e.g. the *ezāfe* in Iranian). Head marking via possessive pronouns had a relatively brief life in the history of English. It developed in Middle English in the form of a non-agreeing, invariant form *(h)ys* (syntactically comparable to Afrikaans *se*):

(21) Middle English
to fortefy *hys* *brethren* *ys* *sayyngys*
to strengthen 3SG.GEN brethren ys comments
[[[NP-G] NP] N-H]
'to strenghthen his brethren's comments' (Allen 2008:247)

This construction competed with the *s*-genitive, which developed phrasal affix status at about the same time (Allen 2008) and eventually completely replaced the head-marking construction (conceivably by support from the similar-sounding marker *ys*).

The Iranian *ezāfe* construction developed early on. An 8th century BC example is the following.

(22) Young Avestan (ca. 8th c. BCE)
puθr-əm *yaṯ* *pourušasp-ahe*
son(M)-ACC.SG EZ Pourušaspa(M)-GEN.SG
[N-H [NP-G]]
'the son of Pourušaspa' (Yašt 5.18)

The particle *yaṯ* is formally the neuter form of the relative pronoun *ya-* but it no longer shows agreement; instead it serves as a linker of a genitival attribute to its head. This type of construction quickly developed into the productive and multi-faceted *ezāfe*-construction characteristic of many Iranian languages, such as Persian, illustrated in example 7 above. The *ezāfe* was lost again in Ossetic. Evidence that it was present in earlier stages comes from remnant usages like the following:[8]

(23) Ossetic, Iron dialect
mæ *fid-ɨ* *zærond*
1SG.GEN father-EZ old
'my old father' (Thordarson 2009:109)

[8] The relevant marker (*-ɨ*) happens to be formally identical to the genitive case, resulting in the apparent synchronic puzzle of a genitive marking a head.

In Ossetic modifiers normally precede the head, and according to Thordarson (2009:109) constructions like 23 are semantically specialized for certain expressions of physical and mental properties and states such as 'old', 'good', 'stupid', etc. Critically for our current interest, in modern Ossetic, ɨ-linked postnominal modifiers occur in a few fixed expressions only (Belyayev 2010:298) and do not support any recursive NP embedding (David Erschler, p. c.).

Adpositions. Adpositions are not attested in adnominal use in Sanskrit, Avestan or Old Persian, and such constructions are unlikely to have existed in Proto-Indo-Iranian, the common ancestor of these languages. Adnominal adpositions developed as a new type in some of the daughter languages, however. A case in point is Middle Persian, where *pad* 'in, against' is an adposition that governs the oblique case (which is visible however only in a few contexts, such as pronouns; see Section S10.5 in Supporting Material 1):

(24) Middle Persian (Zādspram, 9th century CE)
tār-kirb-ān pad čihr ud dēs ī Azdahāg
darkness-body-PL in shape and form EZ Azdahāg
[N [NP-P N-H [NP]]]
'creatures of darkness in the shape and appearance of (the dragon) Azdahāg' (Gignoux & Tafazzoli 1993:36)

The construction disappeared again in Modern Persian since adpositions were reanalyzed as phrasal case prefixes that no longer assign case.

Adpositions also had a relatively fast turnover in Celtic: adnominal adpositions were reanalyzed as genitives in Middle Welsh and Modern Breton (cf. 17 above). In Modern Welsh, however, the phrasal genitive prefix newly acquired adpositional properties again (under influence from English). As such, it can be stranded:

(25) Modern Welsh
Lle 'dach chi 'n dod o?
where be.PRS.2PL 2PL PROG come.INF from
'Where do you come from?' (Borsley et al. 2007:116)

This reanalysis has not affected the possibilities for recursion:

(26) Modern Welsh
disgrifaid o-r rhes o dai
description(M) from-ART row(F) from house(M).[PL]
[N [NP-P [NP-P]]]
'the description of a row of houses' (Borsley et al. 2007:72)

Juxtaposition. In several ancient IE languages, endocentric juxtaposition was not very prominent and became fully recursive only later. A case in point is Old Norse, where juxtaposition with more than two members is rarely attested (Carr 1939:200-201). The few cases that do occur either involve a genitive (27a) or are not recursive (27b):

(27) a. *guðs- reiðis- verk*
God(M).GEN.SG- anger(F).GEN.SG- act(NT).NOM.SG
[[[NP-G] NP-G] N]
'act rousing the anger of God' (Carr 1939:200)

b. *họfuð- rað- gjafi*
head(NT)- advice(NT)- giver(M).NOM.SG
[NP-Ø [[NP-Ø] N]]
'chief counsellor' (Carr 1939:201)

In 27b, the first element (*họfuð* 'head') is not recursively embedded into an embedded element but modifies the unit *rað-gjafi* 'counsellor' as a whole. Juxtaposition became freely available for recursion only later in the history of the branch. In modern Icelandic for example, juxtaposition allows recursive interpretations (Harðarson 2016):[9]

(28) Icelandic
járn- stál- fótur
iron(NT) chair(M) leg(M)
[[[NP-Ø] NP-Ø] N]
'leg of an iron chair'

A similar development can be seen in Irish: whereas Old Irish did not allow juxtaposition of more than two nominal elements, this has become a popular strategy in Modern Irish (see Section S8.4 in Supporting Material 1). Similarly, while Latin blocks recursive juxtaposition (see Section S13.3), Italian – unlike many other modern Romance languages – now allows it (as in for example *programma riciclo materiali* 'material recycling program'; see Section S13.2).

Indo-Aryan illustrates the opposite process, where recursive juxtaposition became unavailable for NP recursion over time. As observed above (12), juxtaposition freely allowed NP recursion in Sanskrit and became particularly popular in Late Sanskrit and Pāli. Despite this prominence, several modern daughter languages have come to disallow juxtaposition and now require case or other markers instead. This is so for example in Nepali, where one cannot juxtapose *māsu* 'meat' and *tarkāri* 'curry' to form **māsu-tarkāri* 'meat curry'. Instead, one needs to use a genitive affix: *māsu-ko tarkāri* (cf. example 20 above). This contrasts with Sanskrit borrowings in Nepali, where juxtaposition is still abundant, e.g. *sthāna-nāma-koś* 'place-name-dictionary' (the title of a publication by the Nepal Academy). A similar fate of juxtaposition is found in most other Indo-Aryan languages in our sample, except in Oriya and Sinhala (see Sections S10.6 and S10.8 in Supporting Material 1, respectively).

3.4 Phylogenetic distribution

The survey in the preceding section shows that each NP type has become newly available for recursion at least once in historical time and has become unavailable at least once. This means that the system was sufficiently dynamic for our hypothesis to be testable: some of the types

[9] The expression also allows a non-recursive interpretation as 'iron [chair leg]', i.e. 'iron leg of a chair' (Harðarson 2016).

developed and disappeared again within less than 500 years (e.g. head marking for recursion in English, new genitives in Middle Welsh) and many within less than 2,000 years (e.g. head marking for recursion in Ossetic, or juxtaposition in Nepali). Therefore, there were many situations in which a choice arose between keeping and not keeping a specific type for recursion. Indeed, the high dynamics of the types suggests that it would have been perfectly possible for an Indo-European language to completely lose all types simultaneously. This would then block NP recursion entirely, mirroring what appears to have happened in Pirahã. If this happened many times, it would falsify our hypothesis.

To test this, we compiled a systematic sample of Indo-European languages (Supporting Material 1). With this sample, we first assess below whether any historically attested or extant language lacks NP recursion across all types (Section 3.4.1). However, synchronically attested distributions can be deceptive (Maslova 2000, Cysouw 2011). For example, even if some feature or state dominates the attested languages of a family, it is entirely possible that it was dispreferred in the unknown past. The problem is illustrated schematically in Figure 1.

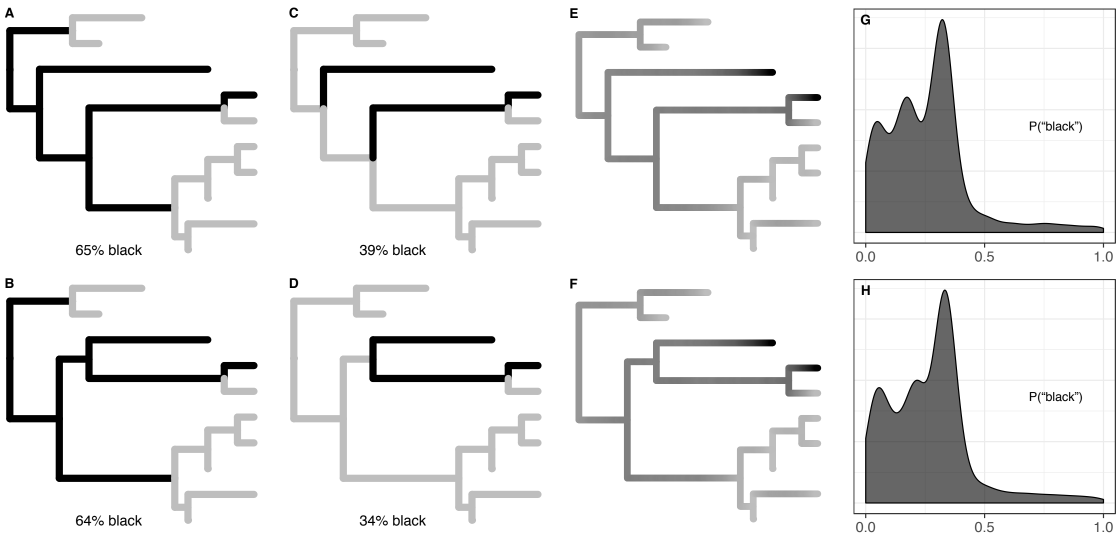

FIGURE 1: Different historical scenarios compatible with the same synchronic data. *A-D:* contrasting possible histories A vs. C and B vs. D that are compatible with the same synchronic distribution of "black" (20%) vs. "gray" (80%) languages (= tips of the trees), assuming different tree topologies across the two rows (A and C vs. B and D, respectively). *E-H:* posterior probability distributions of states estimated by the methods explained in Section 3.4.2, visualized as densities over time in a gray-to-black gradient (E and F) and as density plots for the probability of "black" across samples of 30-year intervals in each tree (G and H).

Assume that 80% (8 out of 10) attested languages in a family are in state 'gray' (which could stand for 'allows syntactic recursion'), as indicated by the colors at the tips of the tree. One is tempted to conclude that 'gray' is the preferred, dominant state. However, keeping the same synchronic frequencies, it is entirely possible that the family was dominated in fact by 'black', regardless of what exact tree structures one assumes (A and B in Figure 1). Of course, it is also possible that the family was indeed dominated by 'gray' (C and D). The problem is that we cannot tell by inspecting synchronic frequencies alone. Sometimes, assuming maximum parsimony for the number of transitions might favor a certain scenario — for example scenario D assumes one change less than B — but this is not always possible: scenarios A and C involve the exact same number of state transitions (namely three), and there is no good reason to assume maximum parsimony to begin with, especially for changes that seem as rapidly reversible as the ones we observed for NP types and their availability for recursion. Qualitative reconstruction of proto-syntax could in principle resolve the problem, but such reconstruction is exceedingly difficult (or perhaps completely impossible) because abstract syntactic properties of the kind we assess here form no natural cognate sets that would be rich enough for deciding between alternative reconstructions.

In response to these problems, rather than debating possible histories, we turn to probability estimates in Section 3.4.2 below. We use Bayesian phylogenetic methods to estimate the posterior probabilities of each type throughout the history of the family. We explain the method below, but panels E-H in Figure 1 show the posterior probability distributions that the method would estimate in the schematic example (assuming branch lengths with realistic time depths). The distribution is shown as a gray-to-black gradient in panels E and F and as a density plot across sampled time intervals in panels G and H. Here, results suggest that it was always slightly less probable for a language to be in state 'black' than in state 'gray', under either of the two tree structures.

For both the qualitative and the quantitative study we sampled languages so as to cover one representative of each branch that was separated from all other branches for at least about a thousand years. So for example, we include both English and Afrikaans, but not both Afrikaans and Dutch.[10] In addition we covered as densely as possible all earlier and intermediate stages of languages. This generates sufficient resolution for phylogenetic methods, while it keeps data acquisition within reasonable limits.

3.4.1 Attested languages

The sample is summarized in Figure 2. The tree is a Maximum Clade Credibility tree taken from Chang et al.'s (2015) ancestry-constrained phylogeny (see Supporting Material 2 for details on the mapping of languages between the two datasets and Figure S2 for a visualization of the data mapped to an alternative tree from Bouckaert et al. 2012). The data in Figure 2 shows that in every language, one or more types of NP embedding are unavailable for recursion. However, no language developed in the direction of Pirahã: in each language of our sample, there is at least one NP type that allows recursion. This supports our hypothesis that languages prefer developing and maintaining structure-building operations that include syntactic recursion (see Section 2). But, as noted above, synchronic distribution can be deceptive, and we turn to probabilistic methods to estimate the most likely pattern of how NP recursion evolved over time.

10 For this we rely on the tree topology and date estimates from Chang et al. (2015), except for Slavic and Romance where their approach underestimates the age of speciation.

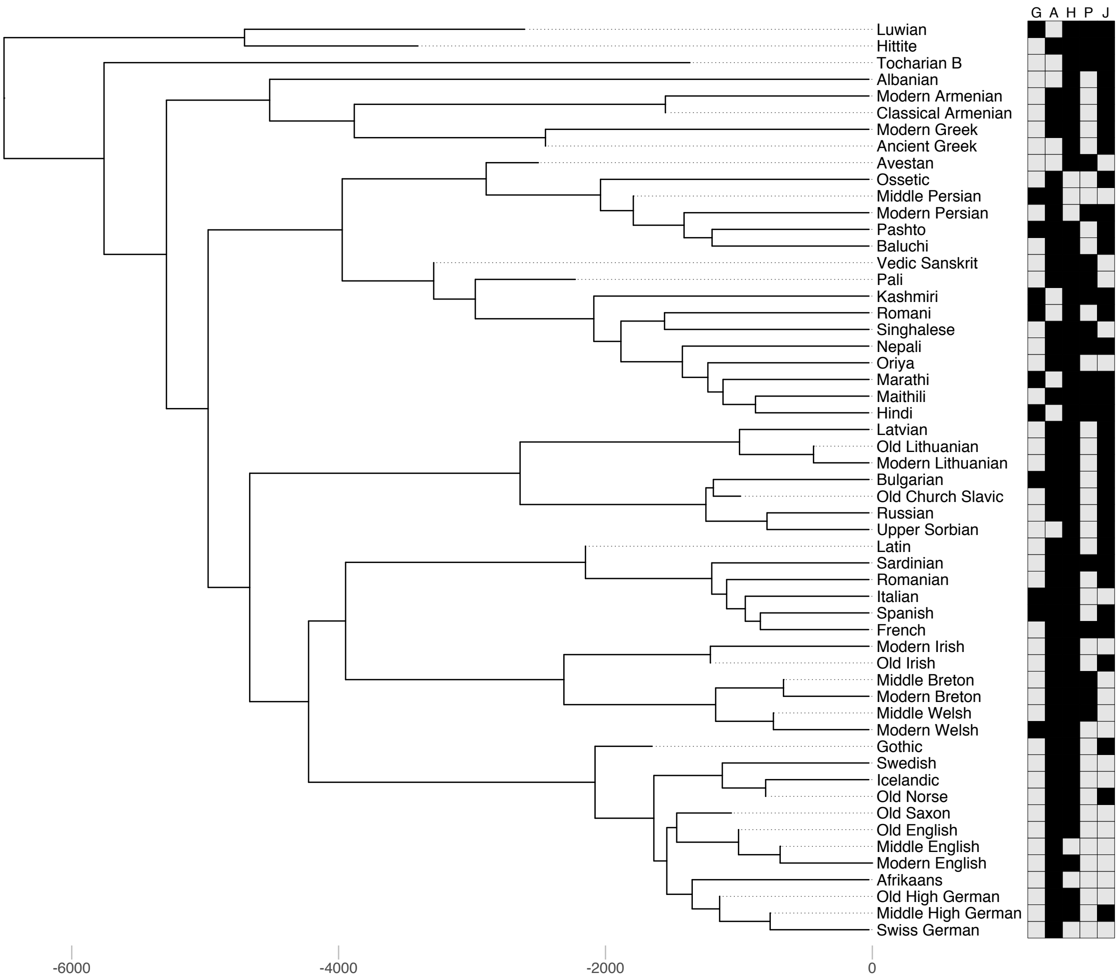

FIGURE 2: Summary of the analyses in Supporting Material 1. NP types: G = genitives, A = adjectivizers, H = head marking, P = adpositions, J = juxtaposition. gray = type is available for recursion, Black = type is not available for recursion. The tree is a Maximum Clade Credibility tree estimated by Chang et al. (2015), with our additions.

3.4.2 Probability estimates

Methods. We model the availability vs. non-availability of an NP type for recursion as discrete states in a Continuous Time Markov Chain evolving over the phylogeny and estimate the rates of transitions between these states in a Bayesian framework (using the package `BayesTraitsV2`; Pagel & Meade 2014). For each type, we fit models assuming equal rates, i.e. the same rates for gaining and losing a type for recursion, and models assuming unequal rates, i.e. evolution that is biased toward one of the two options. We then compare the likelihood of these models with Bayes Factors.[11] The best-fitting transition rate estimates are then used for Stochastic Character Mapping (Nielsen 2002, Huelsenbeck et al. 2003, Revell 2012). Stochastic Character Mapping simulates histories of state change in a tree, given transition rates and the states in the tips, e.g. an NP type might be available for 500 years, then be unavailable for 1000 years, then emerge again etc. There are many different ways in which such a history (technically known as a character map) is compatible with the transition rates and the data. The solution to this problem is to estimate the posterior probability distribution of character maps through Monte-Carlo Markov Chain sampling. The posterior character maps are then aggregated into density estimates over time by binning the branches in the tree into time intervals of about 30 years. For this, we use a procedure introduced by Revell (2013) for visualizing stochastic character maps on a tree (as illustrated by panels E and F in Figure 1). Finally we combine the estimated density distributions from each NP type and compute for each time interval (bin) the probability that at least one type is available for recursion. The time interval can also be thought of as as a DIACHRONIC TRIAL, in which we assess the posterior probability of a type being available or not. This corresponds to the visualizations in panels G and H in Figure 1.

Phylogenies. Since the topologies and branch lengths of trees are themselves uncertain, we estimate the posterior probabilities of stochastic character maps not on any one consensus or summary tree, but on a large sample of posterior trees. For this we used the posterior sample of Indo-European trees estimated by Chang et al. (2015). In order to assess whether our results are robust against the assumptions of this model, we furthermore replicated all analyses on the tree sample estimated by Bouckaert et al. (2012).

Our dataset includes several extinct languages which are not covered by either of these trees because the lexical data was not sufficiently worked up or insufficient for reliably inferring phylogenies (Chang et al. 2015:219-212): Old Saxon, Old English, Middle English, Middle High German, Middle Welsh, Middle Breton, Middle Persian, Old Lithuanian, Pāli, and (in the case of Chang et al.'s trees only) Luwian. In order to include these languages in our estimates, we grafted them on the tree sample. The age of each language was randomly sampled from a uniform distribution bounded by the earliest and latest attestation dates (see Supporting Material 2, Section 1). The resulting tree sample is visualized in Figure 3 as a `DensiTree` which

[11] Bayes Factors (*BF*) are reported on a log scale and are defined as double the difference between the log marginal likelihood of the more complex model (assuming unequal rates) and the log marginal likelihood of the simpler model (assuming equal rates). *BF*s smaller than 2 are conventionally interpreted as only weak or no evidence for the more complex model, *BF*s higher than 2 as positive, higher than 5 as strong and higher than 10 as very strong evidence for the more complex model (e.g. Pagel & Meade 2014, Dunn et al. 2011, Cysouw 2011).

gives an impression of the amount and loci of phylogenetic uncertainty (Bouckaert & Heled 2014). (For a similar representation of Bouckaert et al.'s (2012) tree estimates, see Figure S4 in Supporting Material 2.)

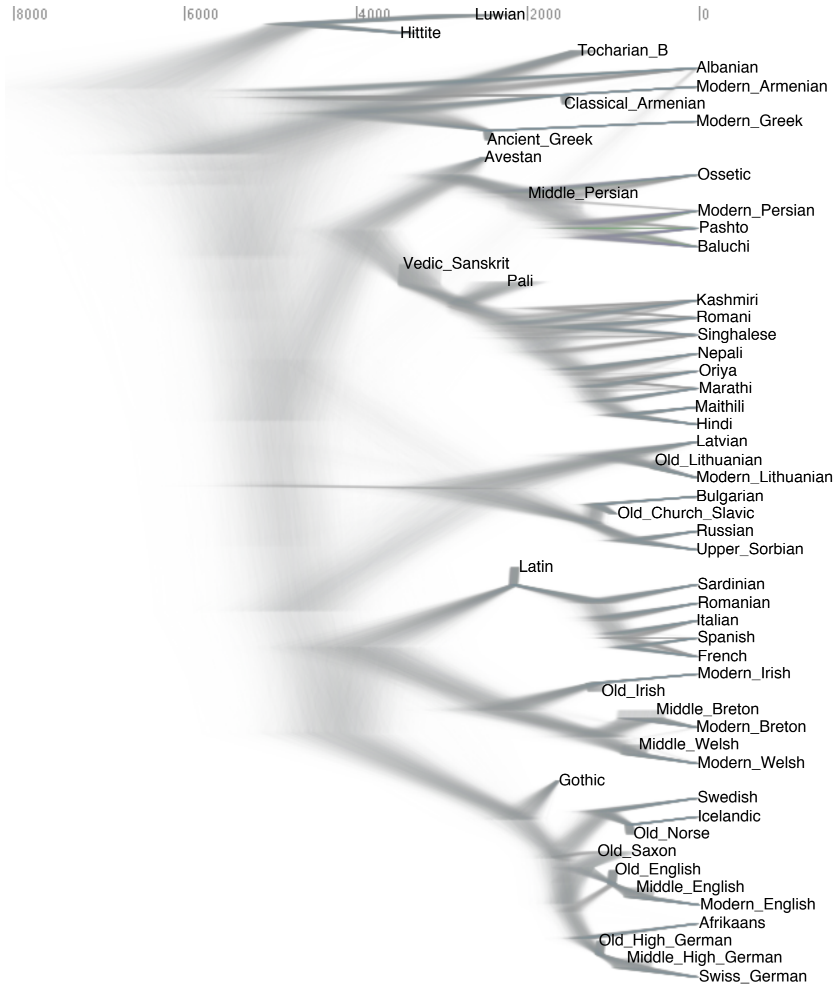

FIGURE 3: DensiTree representation of the posterior tree sample in Chang et al. (2015). In order to match our dataset, we removed some and added other languages, as described in the main text and with further detail in Supporting Material 2.

We estimated transition rates on the full posterior tree samples.[12] But for Stochastic Character Mapping we used a random subset of 1000 trees only because larger samples are computationally extremely expensive, with no apparent gain in estimation quality.

Results and discussion. Results across 10 replications of the transition rate estimates suggest that for adpositions and juxtaposition, there is no evidence for biased evolution, i.e. models

[12] See Supporting Material 2 for further details, including a discussion of our assumptions about priors.

assuming equal rates fit the data better (adpositions, *BF* = 3.53 ± .39) or just as well (juxtaposition, *BF* = .81 ± .13) as models assuming different rates. By contrast, genitives, adjectivizers and head marking show strong evidence of biased evolution. Genitives favor evolution toward being available for recursion (*BF* = 13.22 ± .18), while adjectivizers and head marking favor evolution away from being available for recursion (*BF* = 7.03 ± .49 and *BF* = 19.58 ± 2.25, respectively) (for a visualization of the rate differences, see Figure S7 in Supporting Material 2). Results were very similar when we replicated the analysis on Bouckaert et al.'s (2012) trees, except that the evidence for a bias against recursively used adposition structures was slightly weaker (*BF* = 5.36 ± .62) (see Table S1 in Supporting Material 2).

The positive bias for recursion with genitives is expected to reach stationarity within a period of about 5,700 years (and about 4,000 years when assuming Bouckaert et al.'s (2012) trees, see Section S2.6 in Supporting Material 2). At this point, which has been reached by now, there will always be an estimated 79% of Indo-European languages that have recursive genitives and 21% that do not (or 77% vs. 23% when assuming Bouckaert et al.'s (2012) trees). Comparing the stationary distribution with the synchronic distribution (where 80% languages have recursive genitives) suggests that the synchronic distribution reflects the bias toward genitives reliably in Chang et al.'s (2015) trees but slightly overestimates the bias in Bouckaert et al.'s (2012).[13]

Our hypothesis predicts that languages prefer NPs with syntactic recursion. The bias in genitives supports this, as there will always be more than almost four times as many languages where at least genitives are available for recursion. This still leaves a bit over 20% of occasions where genitives are unavailable as well. In order to assess the impact of this over diachronic trials, we turn to the results from Stochastic Character Mapping. These are summarized in Figure 4. The figure plots the posterior probabilities of a type being available over 2.8 million time intervals of about 30 years. These intervals are sampled from the character maps across trees. The results based on Bouckaert et al.'s (2012) trees are very similar (see Figure S23 in Supporting Material 2).[14]

The distribution of probabilities varies greatly across types and none is guaranteed to be available in each generation of speakers. However, the combined probabilities of at least one type (leftmost column) reaches an estimated mean of 98%. Results based Bouckaert et al.'s (2012) are very similar (see Section S3.6 of Supporting Material 2).

These results are strong support for our hypothesis: while on about 20% occasions, the preferred type in IE, genitives, is unavailable for recursion, in these occasions chances are close to 100% that at least one other type will be available instead.

[13] Note that if we knew only the currently extant languages, with 72% recursive genitives, we would underestimate the strength of the bias. This should caution us further against quick conclusions based on synchronic samples.

[14] Sections S3.3-4 of Supporting Material 2 include visualizations of Stochastic Character Maps on the Maximum Clade Credibility trees (corresponding to panels E and F in Figure 1) in order to give a sense of the estimated dynamics, both for each type separately and for the combination of types. Such a visualization is not possible when Stochastic Character Mapping is performed on entire posterior tree samples, as we do here for the main analysis.

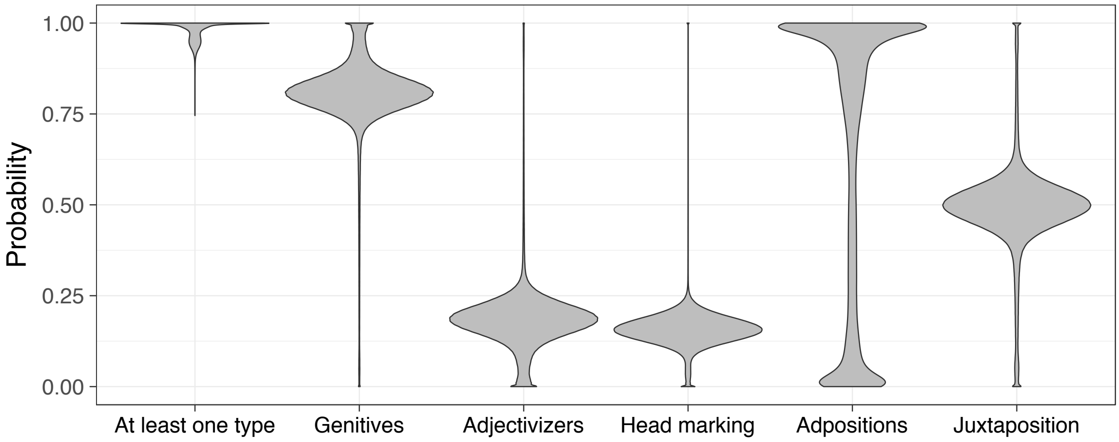

FIGURE 4: Posterior probabilities of types being available for recursion in each speaker generation, i.e. diachronic trial. The probabilities are plotted as density distributions mirrored on the vertical axis ('violin plots'; Hintze & Nelson 1998): the wider the shapes, the greater the probability values across all sampled trials

4 Discussion and conclusions

Our hypothesis predicts that within any given phrase, there tends to be at least one type that allows syntactic recursion. This is confirmed for Indo-European NPs both qualitatively (all attested languages in our sample have at least one such type) and quantitatively (the probability of having at least one type in any given time interval is estimated to be close to 100%). Some NP types are disfavored to various degrees, i.e. they are likely to be lost or not to develop at all for recursion. But we find that whenever these types become or remain unavailable in a language, there is a very strong bias for developing or retaining an 'escape' type that allows recursion.

Our hypothesis maintains that this evolutionary bias is caused by a processing principle that favors syntactic recursion. Alternatively, one might attribute our findings to a preference for having genitives, since this is the dominating escape type in Indo-European. However, this alternative has two shortcomings. First, it does not does explain why in those cases where genitives are unavailable for recursion, this is compensated by alternative strategies for NP recursion (e.g. by adjectivizers in Luwian and in several modern Indo-Aryan languages). Second, a preference for genitives lacks a natural explanation, i.e. there is no intrinsic reason for them to be preferred. We know from other families that for example head-marking strategies can just as well be preferred, as is generally the case around the Pacific (Nichols 1992, Nichols & Bickel 2005). And among the dependent marking types that are prevalent in Eurasia, genitives are only one possibility, along with adpositions and adjectivizers.

Given this, we submit that the best explanation for the evolutionary bias we find is indeed a processing principle that favors not a specific NP type, but NP recursion in at least one type, regardless of which type that is. If this explanation is on the right track, it challenges the idea that syntactic recursion in NPs is a mere option (Fitch et al. 2005, Watamull et al. 2014), with no implications on cross-linguistic distributions. Given the evolutionary bias we detect, there must be another mechanism beyond this, and given the preference of recursive operations across many different domains of human cognition (Section 2), a processing principle seems the most likely candidate for this mechanism. In this light, it seems that the link between syntactic recursion and recursion in the general, mathematical sense is more fruitfully explored through research on processing and its effect on language evolution than through controversial (Pullum & Scholz 2010, Kornai 2014) assumptions about the role of infinite counting in the description of syntax or linguistic creativity.

To fully establish our theory, however, at least three issues need to be worked out. First, our predictions need to be tested in a larger sample of diachronic transitions from other families and continents. We are confident that results will replicate. One reason is that the typological record suggests that there are many more languages that allow NP recursion than languages that do not allow NP recursion. As we noted, synchronic generalizations of this kind can be deceptive and some previous generalizations have indeed been challenged by phylogenetic and other diachronic analysis (Dunn et al. 2011, Bickel et al. 2014), but when effects are as strong as in our case here, we would not expect this to happen. Another reason for confidence is that the Indo-European case is relatively strong evidence by itself. Because we relied on posterior samples of trees and stochastic character maps, our results are not based on a handful of surveyed time intervals (generations), but on a sample of about about 2.8 million such intervals. Also, given the geographic spread of the family, these transitions occurred in very different contact situations and sociolinguistic environments. This makes it plausible that our results reflect principles of diachronic change and are not just specific to Indo-European.

The second issue that needs further research is the scope of the theory. Here, we focused on NPs only. With regard to sentence-level syntax, it remains an open question whether syntactic recursion or simple conjunction is preferred. Again, for this a larger sample of data would be needed. Similarly, with regard to within-word syntax (morphotactics), it is unclear to what extent recursive operations dominate. Many word structures appear to be built by simple string concatenation although explicitly recursive operations are also attested, e.g. in derivations like *anti-anti-establishment* and in some languages also in verb compounding, which in the Tibeto-Burman language Chintang involves recursive additions of partially inflected stems (Bickel et al. 2007).

A third issue that remains open concerns the biological basis of our theory. So far, we have left it open whether the preference for syntactic recursion is caused by general efficiency gains when using the brain's broader faculty for recursion or whether the preference is mediated by the specific advantages of recursion for building hierarchical structures (DENDROPHILIA). Further insight here will depend on neurobiological research that systematically disentangles recursive from hierarchical operations and compares each of these across cognitive domains (Martins et al. 2014, Fischmeister et al. In press). At present we cannot locate the source of syntactic recursion sufficiently well.

References

Ackerman, Farrell, Gregory T. Stump & Gert Webelhuth. 2011. Lexicalism, periphrasis, and implicative morphology. In Robert D. Borsley & Kersti Börjars (eds.), *Non-transformational syntax: formal and explicit models of grammar*, 325 – 357. London: Blackwell.

Allen, Cynthia. 2008. *Genitives in early English: typology and evidence.* Oxford: Oxford University Press.

Bauer, Anna. 2014. *Morphosyntax of the Noun Phrase in Hieroglyphic Luwian.* Leiden: Brill.

Belyayev, Oleg. 2010. Evolution of case in Ossetic. *Iran and the Caucasus* 14. 287–322.

Bickel, Balthasar. 2015. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis, 2nd edition*, 901 – 923. Oxford: Oxford University Press.

Bickel, Balthasar, Goma Banjade, Martin Gaenszle, Elena Lieven, Netra Paudyal, Ichchha P. Rai, Manoj Rai, Novel K. Rai & Sabine Stoll. 2007. Free prefix ordering in Chintang. *Language* 83. 43–73.

Bickel, Balthasar, Alena Witzlack-Makarevich, Kamal K Choudhary, Matthias Schlesewsky & Ina Bornkessel-Schlesewsky. 2015. The neurophysiology of language processing shapes the evolution of grammar: evidence from case marking. *PLoS ONE* 10. e0132819.

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko. 2014. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej Malchukov & Marc Richards (eds.), *Scales: a cross-disciplinary perspective on referential hierarchies*, 7–43. Berlin: De Gruyter Mouton.

Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns.* New York: Cambridge University Press.

Boersma, Paul. 1998. *Functional phonology: formalizing the interactions between articulatory and perceptual drives.* The Hague: Holland Academic Graphics.

Bornkessel-Schlesewsky, Ina & Matthias Schlesewsky. 2009. *Processing syntax and morphology: a neurocognitive perspective.* Oxford: Oxford University Press.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The Syntax of Welsh.* Cambridge: Cambridge University Press.

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. 957–960.

Bouckaert, Remco R & Joseph Heled. 2014. `DensiTree` 2: seeing trees through the forest. BioRxiv [`http://dx.doi.org/10.1101/012401`].

Bubenik, Vit. 1998. *A historical syntax of late Middle Indo-Aryan (Apabhramsa).* Amsterdam, Philadelphia: John Benjamins.

Carr, Charles Telford. 1939. *Nominal compounds in Germanic.* London: St. Andrews University.

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

Chomsky, Noam. 1957. *Syntactic structures.* The Hague: Mouton.

Chomsky, Noam. 1975. *The logical structure of linguistic theory.* New York: Plenum Press [first circulated in 1955].

Christiansen, Morten H & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31. 489–509.

Christiansen, Morten H. & Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences* 39. doi:10.1017/S0140525X1500031X.

Corballis, Michael C. 2011. *The Recursive Mind: The Origins of Human Language, Thought, and Civilization.* Princeton: Princeton University Press.

Corbett, Greville G. 1987. The morphology/syntax interface: evidence from possessive adjectives in Slavonic. *Language* 63. 299–345.

Croft, William. 2003. *Typology and universals.* Cambridge: Cambridge University Press [2nd edition].

Cysouw, Michael. 2011. Understanding transition probabilities. *Linguistic Typology* 15. 415–431.

de la Haye, Pierre & Tanguy Gueguen. 2002. *Vie de saint Yves – Buhez sant Euzen*, vol. 1. Saint-Brieuc: Skol.

Ding, Nai, Lucia Melloni, Hang Zhang, Xing Tian & David Poeppel. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* 19. 158–164.

Donaldson, Bruce C. 1993. *A grammar of Afrikaans.* Berlin, New York: Mouton De Gruyter.

Dunn, Michael J., Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.

Evans, Nicholas & Stephen C. Levinson. 2009. The Myth of Language Universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32. 429–448.

Everaert, Martin B.H., Marinus A.C. Huybregts, Noam Chomsky, Robert C. Berwick & Johan J. Bolhuis. 2015. Structures, not strings: linguistics as part of the Cognitive Sciences. *Trends in Cognitive Sciences* 19. 729 – 743.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Current Anthropology* 46. 621–646.

Fischmeister, Florian P., Mauricio J.D. Martins, Roland Beisteiner & W. Tecumseh Fitch. In press. Self-similarity and recursion as default modes in human cognition. *Cortex* .

Fitch, W. Tecumseh. 2010. Three meanings of "recursion": key distinctions for biolinguistics. In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *The evolution of human language: biolinguistic perspectives*', 73–90. Cambridge: Cambridge University Press.

Fitch, W. Tecumseh. 2014. Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews* 11. 329–364.

Fitch, W. Tecumseh & Angela D. Friederici. 2012. Artificial grammar learning meets formal language theory: an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367. 1933–1955.

Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97. 179–210.

Fortson, Benjamin W. 2010. *Indo-European Language and Culture. An Introduction.* Oxford: Blackwell 2nd edn.

Futrell, Richard, Laura Stearns, Daniel L. Everett, Steven T. Piantadosi & Edward Gibson. 2016. A corpus investigation of syntactic embedding in Pirahã. *PLoS ONE* 11. e0145289.

Gignoux, Philippe & A. Tafazzoli (eds.). 1993. *Anthologie de Zādspram. Édition critique du texte pehlevi.* Paris: Association pour l'Avancement des Etudes Iraniennes.

Givón, Talmy. 1979. *On understanding grammar.* New York: Academic Press.

Greenberg, Joseph H. 1995. The diachronic typological approach to language. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 143–166. Oxford: Clarendon.

Grice, H. P. 1975. Logic and conversation. In Peter Cole & Jerry Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, 41–58. New York: Academic Press.

Haig, Geoffrey. 2011. Linker, relativizer, nominalizer, tense particle: on the ezafe in West Iranian. In Ha Yap Foong, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages*, 363–389. Amsterdam: Benjamins.

Harðarson, Gísli Rúnar. 2016. Peeling away the layers of the onion: on layers, inflection and domains in Icelandic compounds. *Journal of Comparative German Linguistics* 19. 1–47.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: what it is, who has it, and how did it evolve? *Science* 298. 1569–1579.

Hawkins, J. David. 2000. *Corpus of Hieroglyphic Luwian Inscriptions. Vol. 1 Inscriptions of the Iron Age.* Berlin/New York: de Gruyter.

Hawkins, John A. 1994. *A performance theory of order and constituency.* Cambridge: Cambridge University Press.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency.* Oxford: Oxford University Press.

Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar: a reconstruction.* Oxford: Oxford University Press.

Hemon, Roparz. 1976–1998. *Geriadur istorel ar brezhoneg.* Plomelin: Preder 2nd edn.

Hintze, J.L. & R.D. Nelson. 1998. Violin plots: a box plot - density trace synergism. *The American Statistician* 52. 181 – 184.

Huelsenbeck, John P., Rasmus Nielsen & Jonathan P. Bollback. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52. 131–158.

Hyman, Larry M. 1976. Phonologization. In Alphonse Juilland (ed.), *Linguistic studies presented to Joseph H. Greenberg*, 407–418. Saratoga: Alma Libri.

Karlsson, Fred. 2010. Syntactic recursion and iteration. In Harry van der Hulst (ed.), *Recursion and human language*, 43–67. Berlin: De Gruyter Mouton.

Kemmerer, David. 2012. The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistics Compass* 6. 50–66.

Kornai, András. 2014. Resolving the infinitude controversy. *Journal of Logic, Language and Information* 23. 481–492.

Lazard, Gilbert. 1992. *A grammar of contemporary Persian.* Costa Mesa, CA: Mazda Publishers.

Lindblom, Björn, Susan Guion, Susan Hura, Seeung-Jae Moon & Raquel Willerman. 1995. Is sound change adaptive? *Rivista di Linguistica* 7. 5 – 37.

Lötzsch, Ronald. 1965. Das sog. Possessivadjektiv im Slawischen, speziell im Sorbischen, und seine Stellung im System der Redeteile. *Forschungen und Fortschritte* 39. 377–379.

Martins, Maurício D. 2012. Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367. 2055–2064.

Martins, Maurício D., F.P. Fischmeier, E. Puig-Waldmüller, J. Oh, A. Geissler, S. Robinson, W. Tecumseh Fitch & R. Beisteiner. 2014. Fractal image perception provides novel insights into hierarchical cognition. *NeuroImage 96* 300–308.

Martins, Maurício D., Isabel Pavão Martins & W. Tecumseh Fitch. 2015. A novel approach to investigate recursion and iteration in visual hierarchical processing. *Behavior Research Methods* 47. doi:10.3758/s13428-015-0657-1.

Masica, Colin P. 1991. *The Indo-Aryan languages.* Cambridge: Cambridge University Press.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4. 307–333.

McDaniel, Dana, Cecile McKee, Wayne Cowart & Merrill F. Garrett. 2015. The role of the language production system in shaping grammars. *Language* 91. 415–441.

Miller, George A. 1967. *The psychology of communication.* New York: Basic Books.

Mithun, Marianne. 1984. How to avoid subordination. *Proceedings of the 10th Annual Meeting of the Berkeley Linguistics Society* 493–509.

Mithun, Marianne. 2010. The fluidity of recursion and its implications. In Harry G. van der Hulst (ed.), *Recursion and human language*, 17–41. Berlin: De Gruyter Mouton.

Newmark, Leonard, Philip Hubbard & Peter Prifti. 1982. *Standard Albanian: a reference grammar for students.* Stanford, CA: Stanford University Press.

Nichols, Johanna. 1992. *Linguistic diversity in space and time.* Chicago: The University of Chicago Press.

Nichols, Johanna & Balthasar Bickel. 2005. Locus of marking (in the clause; in possessive noun phrases; and whole-language typology). In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 98–109. Oxford: Oxford University Press.

Nielsen, Rasmus. 2002. Mapping mutations on phylogenies. *Systematic Biology* 51. 729–732.

Pagel, Mark & Andrew Meade. 2014. `BayesTraits V2.0`. Software and manual, `http://www.evolution.rdg.ac.uk/BayesTraits.html`.

Pallier, Christophe, Anne-Dominique Devauchelle & Stanislas Dehaene. 2011. Cortical representation of the constituent structure of sentences. *Proceedings of National Academy of Science (USA)* 108. 2522–2527.

Perfors, Amy, Joshua B Tenenbaum, Edward Gibson & Terry Regier. 2010. How recursive is language? a Bayesian exploration. In Harry van der Hulst (ed.), *Recursion and human language*, 159–177. Berlin: De Gruyter Mouton.

Petit, Daniel. 2009. La préhistoire des adjectifs détérminés du baltique et du slave. *Bulletin de la Société de Linguistique de Paris* 104. 311–360.

Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: what's special about it? *Cognition* 95. 201–236.

Pullum, Geoffrey K. 2013. The central question in comparative syntactic metatheory. *Mind & Language* 28. 492–521.

Pullum, Geoffrey K. & Barbara C. Scholz. 2010. Recursion and the infinitude claim. In Harry van der Hulst (ed.), *Recursion and human language*, 113–137. Berlin: De Gruyter Mouton.

Reinöhl, Uta. 2016. *Grammaticalization and the Rise of Configurationality in Indo-Aryan.* Oxford: Oxford University Press.

Revell, Liam J. 2012. `phytools`: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3. 217–223.

Revell, Liam J. 2013. Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution* 4. 754–759.

Snell, Rupert & Simon Weightman. 2003. *Teach yourself Hindi.* London: Hodder and Stoughton Educational.

Thordarson, Fridrik. 2009. *Ossetic grammatical studies.* Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Tiede, Hans-Jörg & Lawrence Neff Stout. 2010. Recursion, infinity, and modeling. In Harry G. van der Hulst (ed.), *Recursion and human language*, 147–158. Berlin: De Gruyter Mouton.

Tomasello, Michael. 2008. *Origins of human communication.* Cambridge, MA: MIT Press.

Verma, Manindra K. 1971. *The structure of the noun phrase in English and Hindi.* Delhi: Motilal Banarsidass.

Viti, Carlotta. 2015. *Variation und Wandel in der Syntax der alten indogermanischen Sprachen.* Tübingen: Narr.

Wackernagel, Jacob. 1905. *Altindische Grammatik: Einleitung zur Wortlehre. Nominalkomposition*, vol. 2 1. Göttingen: Vandenhoeck & Ruprecht.

Watamull, Jeffrey, Marc D. Hauser, Ian G. Roberts & Norbert Hornstein. 2014. On recursion. *Frontiers in Psychology* 4. `doi:10.3389/fpsyg.2013.01017`.

Windfuhr, Gernot & John R Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian Languages*, 416–544. London, New York: Routledge.

Yoshida, Daisuke. 1987. *Die Syntax des althethitischen substantivischen Genitivs.* Heidelberg: Winter.

NP recursion over time: evidence from Indo-European – Supporting Material 1: Data Analysis

Manuel Widmer[1], Sandra Auderset[2], Johanna Nichols[3,4], Paul Widmer[1], and Balthasar Bickel[1]

[1]University of Zürich
[2]University of California, Santa Barbara
[3]University of California, Berkeley
[4]Higher School of Economics, Moscow

Contents

1 Acknowledgments

We are very grateful for the generous help offered by the following colleagues: Riaz Ahmed (Balochi), Tim Aufderheide (Indo-Aryan), Damián Blasi (Spanish), Elias Bounatirou (Old Russian), Laura Canedo (Spanish), Kamal Kumar Choudhary (Hindi), Sadananda Das (Oriya), Stefan Dedio (Welsh), Michael Erlach (Anatolian), David Erschler (Ossetic), Victor Friedman (Albanian), Mícheál Hoyne (Modern Irish), Carina Jahani (Balochi), Andra Kalnača (Latvian), Raikhan Kerner (Modern Russian, Old Russian), Agnes Korn (Balochi), Herve Le Bihan (Modern Breton), Martha Mariani (Spanish), Jurgis Pakerys (Lithuanian and Latvian), Elisa Papathanassiou (Modern Greek), Maneedipa Patnaik (Oriya), Netra P. Paudyal (Nepali), Salvatore Scarlata (Indo-Iranian), Narayan Sharma Gautam (Nepali), Sascha Völlmin (Modern Persian), Max Wahlström (Albanian), Alena Witzlack-Makarevich (Afrikaans), Sonja Wölke (Upper Sorbian), Guðrún Þórhallsdóttir (Icelandic).

2 Notational conventions and abbreviations

Where no reference is indicated, the examples and descriptions are based on native or near-native knowledge of the language by team members. Both the branches and the individual languages within the branches are ordered alphabetically. In examples, we only mark gender on nouns if there is gender agreement between a given noun and one of its dependents. In all other contexts, gender is not indicated.

Glosses

1	first person	EZ	*ezāfe* marker
2	second person	F	feminine
3	third person	FUT	future
ABL	ablative	GD	genitive-dative
ACC	accusative	GEN	genitive
ADJ	adjective	IMP	imperative
ADJZ	adjectivizer	INDF	indefinite
ADV	adverbial	INF	infinitive
AI	ablative-instrumental	INS	instrumental
AOR	aorist	INST	instrumental case
ART	article	LNK	linker
CO	common (gender)	LOC	locative
DAT	dative	M	masculine
DEF	definite	MD	middle
DEM	demonstrative	N	head
DISJ	disjunctive pronoun	N	neuter
DL	dative-locative	N-H	head-marked NP
DU	dual	NA	nominative-accusative

NEG	negative	POSS	possessive
NMLZ	nominalizer	PREP	prepositional case
NO	nominative-oblique	PROG	progressive
NOM	nominative	PRS	present
NP-A	adjectivizer-marked NP	PST	past
NP-G	genitive-marked NP	PTCL	particle
NP-P	adposition-marked NP	PTCP	participle
NP-Ø	embedding by juxtaposition	QUOT	quotative
NT	neuter	RCT	rectus (direct) case
OBL	oblique	REL	relative
PASS	passive	SG	singular
PL	plural		

3 Coding Conventions

For each language, we investigate the five types of signaling the dependency relation discussed in the main paper, determining the presence of the type and whether or not it allows recursive expansion. For a type to be coded as allowing recursion, it must be fully productive and allow recursive embedding of an NP into an NP (or, if one analyzes compounding as morphological incorporation, of an N into an N). We limit our attention to embedded nouns and do not include pronouns, numerals and other referential expressions. We also generally exclude from our purview constructions which only exist as borrowings and calques and have no currency of their own, such as the many Sanskrit *tadbhava*s and *tatsama*s that are commonly used in literary genres in most modern Indo-Aryan languages.

Our working definitions of the five types and their formal notations (disregarding linear order) in the data are as follows:

Genitives: **[[NP-G] N].** The embedded NP is marked as a dependent by a case affix, attached to the entire phrase or to all or any constitutive stems of the phrase, typically the main noun stem. Genitives do not assign any further distributional properties nor do they act as independent syntactic entities.

Adpositions: **[[NP-P] N].** The embedded NP is marked as a dependent by a syntactically independent unit that has an argument structure and belongs to a distinct part of speech. Typical effects of the marker's syntactic independence include case government (reflecting their argument structure) and stranding (reflecting their status as syntactic words).

Adjectivizers: **[[NP-A] N].** The embedded NP is assigned a property that makes it behave distributionally like a modifying adjective, resulting in whatever effects adjectives may have in a language, e.g. agreement with the head noun. Adjectivizing markers can be phrasal or stem affixes.

Juxtaposition: **[[NP-∅] N].** The embedded NP is placed next to the head without any further marking (apart from secondary effects in morphophonology, stem alternations, sandhi, stress shifts etc.). We do not distinguish between incorporated/compounded vs. formations (i.e. between [[[...] N-] N] vs. [[$_{\text{NP}}$ [...] N] N], respectively) nor between phonologically bound vs. phonologically free combinations, and annotate them invariably as [[NP-∅] N].

Head-markers: [[NP] N-H]. The NP is embedded into an NP by means of a marker on the head of the higher NP. The marker can be a phrasal affix, a stem affix, or a syntactic co-constituent of the head. For justification, see the main paper, Section 3.1

In discussing juxtaposition phenomena, especially under the rubric of 'compounding', the literature sometimes talks of recursion when the operation of compounding is applied to a pre-established compound. Our interest here is more specifically in recursive embedding. Thus, a structure like [[*student*] *film award*], where *student* modifies the constituent *film award* as a whole and not the embedded constituent *film*, does not count as recursion according to our coding scheme.

Similarly, in exocentric structures (*bahuvrīhi*s) such as Vedic Sanskrit *marútas rúkma-vakṣasas* [Marut.NOM.PL decoration-chest.ADJZ.NOM.PL] 'Maruts (a class of gods) with decorations on their breasts', it is in most cases the composite constituent as a whole that assumes the function of a modifier ([*marútas* [*rúkma-vakṣasas*]], [[*model-theoretical*] *syntax*]). Moreover, such expressions draw on constructional patterns where the second level of embedding admits various incorporated modifiers, often adverb-like in function (cf. Vedic Sanskrit *raghu-yā́-man-* [rapid-go-NMLZ-] 'with a rapid course', English *white-washed wall*, *cross-sectional study*), and does not define the specific [[[NP] N] N] pattern that we are interested in. For these reasons we discard exocentric structures from our analysis.

4 Albanian

4.1 Albanian

We have sampled Gheg Albanian instead of Tosk Albanian as in Chang et al. (2015), because we had better access to these data. The branch lengths of the two varieties can be assumed to be very similar.

Albanian exhibits three of the surveyed NP structures, but their productivity and recursive abilities vary: genitives, adpositions and adjectivizers are both productive and allow NP recursion, while compounds are constrained with regard to recursion. There is no head-marking construction.

What we analyze as a genitive construction involves a case marker sometimes referred to as the "attributive ablative". This construction is only used with an indefinite head, e.g. *zër-a fëmij-ësh* [voice-INDF.NA.PL child-INDF.ABL.PL] 'children's voices' (Buchholz & Fiedler 1987:415). It can embed other NPs:

(1)

potpuri	*këng-ësh*	*dashuri-e*	
potpourri.INDF.NA.SG	song-INDF.ABL.PL	love-INDF.ABL.SG	
[N	[NP-G	[NP-G	]]]

'potpourri of love songs' (Max Wahlström, p.c.)

Albanian prepositions govern case. The choice is lexically determined, compare e.g. *prej Malasej-i* [from Malasej-ABL.SG] 'from Malasej' and *me shpatë* [with sword.ACC.SG] 'with the sword' (Newmark et al. 1982:290-296). Prepositions can embed other constituents, such as another preposition-marked NP in Example 2.

(2)

libr-i	*mbi*	*tavolin-ën*	*në*	*dhom-ën*	*e*	*dit-ës*	
book-DEF.NOM.SG	on	table-DEF.ACC.SG	in	room(F)-DEF.ACC.SG	ADJZ.DEF.NA.F	day-DEF.DAT.SG	
[N	[NP-P		[NP-P		[NP-A		]]]]

'the book on the table in the living room' (Plator Gashi, p.c.)

What is traditionally referred to as the genitive case in Albanian (on behalf of its function) is by closer inspection an adjectivizer construction. The key feature is a "particle of concord", which agrees with its head noun in case, gender and number. This particle, which probably derives from demonstrative pronouns (Bubenik 2009:97-99), allows the embedding of NPs, certain adjectives and other parts of speech and it also indicates third person possession with a few kinship terms. The particle assigns the dative case to the embedded NP (Newmark et al. 1982:136,159-162). The following illustrates NP recursion in an adjectivized NP:

(3) a.

shtëpi-a	*e*	*shok-ut*	*të*	*baba-it*
house(F)-DEF.NOM.SG	ADJZ.DEF.NA.F	friend(M)-DEF.DAT.SG	ADJZ.DEF.DAT.SG.M	father(M)-DEF.DAT.SG
[N	[NP-A		[NP-A	

tim	
1SG.POSS.DAT.SG.M	
[NP-A	]]]]

'the house of my father's friend' (Plator Gashi, p.c.)

b. *burim-i* *i* *mirëqenie-s* *së* *popuj-ve*
source(M)-DEF.NOM.SG ADJZ.DEF.NOM.SG.M prosperity(F)-DEF.DAT.SG ADJZ.DEF.DAT.SG.F people(M)-DEF.DAT.PL
[N [NP-A [NP-A
të *vende-ve* *socialiste*
ADJZ.DEF.DAT country(M)-DEF.DAT.PL socialist
[NP-A [NP-A [NP-A]]]]]]
'the source of the prosperity of the peoples of the socialist countries' (Buchholz & Fiedler 1987:414)

There are three further adjectivizers: *-ar*, *-tar* (which has developed from *-ar*) and *-or*. They appear to have roughly the same function and are limited in their productivity. Examples are *bregdet-ar* [coast-ADJZ] 'coastal', *am-tar* [mother-ADJZ] 'maternal', and *diell-or* [sun-ADJZ] 'solar' (Newmark et al. 1982:200-202). They agree with the top level head of the NP in gender and number. Such adjectivizers can also embed other constituents, as illustrated in Examples 4a and 4b.

(4) a. *rezidenc-a* *presidencial-e* *ship-tar-e*
residence(F)-DEF.NOM.SG president.ADJZ-SG.F Albania-ADJZ-SG.F
[N [NP-A [NP-A]]]
'the residence of the Albanian president'

b. *shtëpi-a* *ime* *atëror-e*
house(F)-DEF.NOM.SG 1SG.POSS.NOM.SG.F father(M).ADJZ-SG.F
[N [[NP-A] NP-A]]
'the house of my father' (Plator Gashi, p.c.)

Albanian also has determinative compounds, in which the second member modifies the first, though the reverse is also possible (Newmark et al. 1982:176). Determinative compounds, however, cannot have more than two members (Max Wahlström, p.c.). As there are no other constructions involving juxtaposition, we conclude that juxtaposition is not available for NP recursion.

5 Anatolian

5.1 Hittite

Hittite is attested from about 1600-1100 BCE on clay tablets found in today's central Turkey. The tablets were used for administrative purposes of the Hittite kingdom (Hoffner & Melchert 2008:1-3). Hittite is written in a cuneiform script adapted from Mesopotamia, although in some instances hieroglyphs are used as well (Hoffner & Melchert 2008:9).[1]

[1] Orthographic conventions: The Hittite writing system is a mixture of logograms and phonetic forms. Words can be written either by a logogram or a phonetic form or as a logogram with a phonetic component. Logograms may also stand before phonetic words as a semantic marker dubbed 'determinative'. By convention, logograms are rendered in capitals if they are Sumerian and in italic capitals if they are Akkadian. Determinatives are rendered as superscripts. The writing system can only indicate V, CV, VC and CVC syllables (Hoffner & Melchert 2008:11-14). The examples provided here represent the phonological forms as closely as possible.

The genitive is the only construction that is fully productive and allows recursive NP embedding. Two constructions are only marginally attested: adjectivizers and juxtaposition. Adpositions exist, but they are not used adnominally. There are no head-marking strategies for NP building.

The genitive case is marked by the suffix *-aš* in the singular and *-an* in the plural. After the Old Hittite period, the suffix *-aš* was used in the plural as well. It is unclear whether this is the result of a merger with the genitive singular or the dative-locative plural, which was also *-aš* (Hoffner & Melchert 2008:73). Genitives are productive and allow NP recursion:

(5) ^{D}*IŠTAR-aš* *lūli-aš* KÁ.GAL-*az*
Istar-GEN pond-GEN door-ABL
[[[NP-G] NP-G] N]
'from the door of Ištar's pond' (KBo XVI 49 I 6) (Yoshida 1987:19)

Hittite does not have adnominal adpositions. There are words expressing spatial relations, which can be used as adverbs, preverbs and postpositions (Hoffner & Melchert 2008:294), but they always modify verbs, cf. Example 6.

(6) KUR-*e* *anda uet*
land-DL.SG into come.3SG.PST
'he came into the land' (Hoffner & Melchert 2008:298)

There are a few suffixes that function as adjectivizers. One of them is *-ašša/i-*, which is related to the Luwian adjectivizer *-assa/i-* (see Section 5.2). However, adjectivized forms are not numerous and they do not seem to be productive. Forms with the suffix *-ašša/i-* are mostly nominalized, e.g. ^{D}*wašdul-ašši-* [sin-ADJZ] '(demon) concerned with sin' (Hoffner & Melchert 2008:56). An example for a denominal adjective is *išpant-iya-* [night-ADJZ-] 'nocturnal' (Hoffner & Melchert 2008:58). There is no evidence that adjectivizers allow the recursive embedding of further NPs.

Hittite possesses determinative compounds, e.g. *zaškar-aiš* [excrement-mouth] 'anus' (Hoffner & Melchert 2008:63). However, we have not found any instances of recursive nominal compounding and since there are no other cases of embedding by juxtaposition, we conclude that this type is not available for NP recursion in Hittite.

5.2 Luwian

Luwian was spoken in central and western Anatolia and northern Syria. The language is attested in two forms: ritual texts written in cuneiform script (Cuneiform Luwian) and inscriptions written in hieroglyphs (Hieroglyphic Luwian). The relationship between the two varieties is not entirely clear, but the linguistic differences appear to be marginal (Melchert 2008:31-32). The following description is based on Hieroglyphic Luwian (abbreviated as H Luwian) material. Most attestations of H Luwian are from the 10th to 7th century BCE, but a few date back to the 13th and 14th century BCE (Melchert 2008:31).[2]

[2] Cuneiform Luwian is only attested in ritual texts of the 16th-13th century BCE and in loanwords in Hittite texts. Consequently, not much is known about this form of Luwian (Melchert 2008:31).

There are two constructions for complex NPs: the genitive and adjectivizers (Bauer 2014). Only adjectivizers allow NP recursion, though. Compounds exist, but they are not productive. There is no head-marking construction in H Luwian and adpositions are not used adnominally.

The H Luwian genitive is marked by the suffix *-s(i)* (Bauer 2014:27). We are not aware of any NP constructions in which a genitive is modified by another genitive or an adjectivized NP. To be sure, there are a small number of examples in which a genitive seems to embed an additional NP constituent Bauer (2014:192-193,199-201). However, all of the relevant examples are ambiguous. In this context, it is important to bear in mind that the H Luwian corpus is dominated by expressions for possessive relations, often very complex ones (Bauer 2014:132). If genitives could embed additional NPs, one would expect to encounter such constructions in the corpus. Since we have not found any instances of such constructions, we conclude that genitives cannot control NP constituents in H Luwian. Example 7a illustrates a simple genitive construction and Example 7b the apposition of a genitive and an adjective.[3]

(7) a. VITIS-*si* FINES-*s*
vineyard-gen.sg border-nom.sg
'the border of the vineyard' (Babylon 1 §6) (Hawkins 2000:392, Bauer 2014:143)

b. *zaya=wa* (*262)*sasaliya* *mariti-s*$_i$ *zúwarimi-s* |FILIUS*muw-iya-ya*$_i$
dem.na.pl.nt=ptcl shooting(nt).na.pl Mariti-gen.sg Zuwarimi-gen.sg son-adjz-na.pl.nt
[n [np-g] [[np-g] np-a]]
'These shootings(?) (are) of Mariti$_i$, of the son$_i$ of Zuwarimi' (Malatya §1)
(Bauer 2014:144)

The language has several adjectivizing suffixes, which agree in case, number and gender with their head noun. The most common suffixes are *-assa-*, *-iya-* and *-alla-*, but there are also others (Bauer 2014:151-155), as in Example 8.

(8) *Tuwana-wanni-s*(URBS) |REX*ti-s*
Tuwana-adjz-nom.sg.co king(co)-nom.sg
'the king of Tuwana (lit.: the Tuwanean king)' (Bor §1) (Bauer 2014:151)

It is conceivable that the formal syncretism of the nominative singular, the genitive singular, and the genitive plural, all ending in (*)-*as* in some of the major inflectional classes, favored the use of adjectivizers.

[3] Orthographic conventions: The H Luwian writing system is a mixture of logograms and phonetic forms. Words can be written either by a logogram or a phonetic form or as a logogram with a phonetic component. Logograms may also stand before or after phonetic words as a semantic marker dubbed 'determinative'. By convention, logograms are rendered as their capitalized Latin counterpart or their capitalized Luwian spelling (if known). The writing system does not represent voicing contrasts in the case of most consonants or distinguish simple from geminate consonants. Only V and CV sequences can be written so that there is no way of directly indicating consonant clusters and word-final consonants. As a consequence, the phonological interpretation of the form of a H Luwian word is often based on Cuneiform Luwian data (Melchert 2008:33-34). The examples provided here are rendered in Bauer (2014)'s transcription, which represent the phonological forms as closely as possible. The symbols used in the Examples represent: ( ) 'determinative', || 'line break', | 'word divider', [ ] 'broken sign', < > 'emended sign' (Bauer 2014:xiii).

Unlike genitives, adjectivizers are entirely unrestricted in their ability for NP recursion. In these constructions, the adjective always agrees with the head of the entire phrase and not with its immediate head:

(9) [*a*]*wa=ta* |*z*[*ati*] *ámi*$_i$ *áláyaza*$_i$*-ss-an* *HÁ+LI-ass-an* SERVUS*la-ya*$_i$
PTCL=PTCL this.DL.SG.NT 1SG.POSS.DL.NT Arrayaza-ADJZ-DL.SG Hattusili-ADJZ-DL.SG servant-ADJZ.DL.SG
[[NP-A] [NP-A] [[NP-A] NP-A
STATUA*rut-i* OVIS(ANIMAL)*ti* PRAE*i* (*69)*sasa-tu*
statue(NT)-DL.SG sheep-AI.SG ADV present-3PL.IMP
] N]
'Let them present(?) the statue of me, of Arrayazza, of the servant of Hattusili, with a sheep.' (MALPINAR §5) (Bauer 2014:148)

In Example 9, the head noun STATUA*rui* 'statue' is modified by the conjoined adjectives *áláyazassan* 'of Arrayazza' and SERVUS*laya* 'of the servant', the latter of which is in turn modified by the adjective *HÁ+LI-assa* 'of Hattusili'. Note that the possessive pronoun *ámi* 'my' refers to the same person as the adjectives *áláyazassa* 'Arrayazza' and SERVUS*laya* 'servant', i.e. the speaker's name is Arrayazza and he is a servant of Hattusili.

Compounds, especially determinative ones, are scarce and not productive in Luwian and the other Anatolian languages (cf. Section 5.1) (Payne 2004:20). The few attested cases, such as REGIO*ni* DOMINUS [country lord] 'country-lord', are all confined to two members. We thus conclude that juxtaposition is not available for recursive NP embedding in the language.

6 Armenian

6.1 Armenian, Classical

Classical Armenian dates back to the 5th century CE, though many of the earlier works are translations from Greek. The manuscripts, however, are from later centuries, the most ancient one, the Moscow Gospel, was written down in 887 CE (Godel 1975:1-2). The Classical Armenian period ends around the 12th century CE.

Classical Armenian uses genitives, adpositions, adjectivizers, and juxtaposition for building complex NPs. Of these, genitives and adpositions allow the embedding of further constituents. There is no head-marking construction.

In Classical Armenian nouns are inflected for case, with several syncretisms. In the singular, there is one form covering the genitive and the dative case (and in some declension classes also the locative case), the plural form covers the ablative as well (Godel 1975:104,106).

(10) a. *jayn-i* *ordw-oy-n* *Astowac-oy*
voice-GD.SG son-GD.SG-ART God-GD.SG
[N [NP-G [NP-G]]]
'the voice of the son of God' (John 5.25)

b. *p'aṙ-aw-k'* *hawr* *iwr-oy*
glory-INS-PL father.GD.SG 3SG.POSS-GD.SG
[N [NP-G [NP-A]]]
'in the glory of his father' (Mark 8.38)

Classical Armenian has a number of prepositions which govern case. Like genitives, they support NP recursion, as in Example 11, in which the prepositional phrase contains a genitive-marked NP:

(11) *Yovsēp‘* *ï* *tan-ē* *Dawt‘-i*
Joseph.NOM.SG in house-ABL.SG David-GD.SG
[N [NP-P [NP-G]]]
'Joseph of the house of David' (Luke 1.27)

There are a number of suffixes that derive adjectives from nouns (Godel 1975:59), e.g. *-(a)kan* as in *tare-kan* [year-ADJZ] 'yearly'. However, we have not found any evidence that derived adjectives are able to embed additional NP constituents and thus conclude that this is not possible in Classical Armenian.

The language also possesses determinative compounds, e.g. *cov-ezr* [sea-side] 'coast'. While many compounds are calques from Greek, there is also a substantial number of genuine Armenian formations, which demonstrates that nominal compounding is a productive process in the language (Godel 1975:59–60). However, we have not found evidence for compounds with more than two members and thus conclude that juxtaposition is not available as a strategy for recursive NP embedding.

6.2 Armenian, Modern

Our description is based on the Eastern dialect of Modern Armenian, because this is the variety included in Chang et al. (2015).

Modern Eastern Armenian relies on the same constructions that were already used in Classical Armenian: genitives, adjectivizers, adpositions, and juxtaposition ('compounding'). Like in Classical Armenian, adjectivizers and juxtaposition are not able to embed another constituent. There is no head-marking construction.

Due to case syncretism that was already present in Classical Armenian (cf. Section 6.1), there is only one marker covering both the dative and the genitive functions (Dum-Tragut 2009:83). In Modern Armenian it is generally referred to as a dative and also glossed as such. As mentioned above, the dative and other cases allow the recursive embedding of an NP:

(12) *usucič’-ner-i* *ašxatavarj-i* *ač-i* *č’ap-ě*
teacher-PL-DAT salary-DAT growth-DAT rate.NOM-DEF
[[[[NP-G] NP-G] NP-G] N]
'the expansion rate of the teachers' salary' (Dum-Tragut 2009:235)

(13) *derviš-i* *hagust-ov* *mi* *mard*
dervish-DAT clothe-INST INDF man.NOM
[[[NP-G] NP-G] N]
'a man in dervish clothes' (lit. 'a man with the clothes of a dervish') (Dum-Tragut 2009:574)

Modern Armenian displays both pre- and postpositions, but postpositions are more frequent (Dum-Tragut 2009:294). Both pre- and postpositions govern a number of different cases

(Dum-Tragut 2009:295-306). Adpositions can be found in adnominal position and they are able to embed additional NPs, as the following examples illustrate:

(14) *erku kron-ner-i mijew p'oxagorcakc'ut'yun-ě*
two religion-PL-DAT between cooperation.NOM-DEF
[[NP-P] N]
'the cooperation between the two religions' (Dum-Tragut 2009:268)

(15) *t'urk'-er-i ew hay-er-i patmut'y-an veraberyal banaveč*
Turk-PL-DAT and Armenian-PL-DAT history-DAT about discussion.NOM
[[[NP-G] [NP-G] NP-P] N]
'a discussion about the history of Turks and Armenians' (Dum-Tragut 2009:551)

Modern Eastern Armenian displays a number of suffixes that derive adjectives from nouns, e.g. *-ayin* as in *leṛn-ayin* [mountain-ADJZ] 'mountainous' (Dum-Tragut 2009:663-665). However, we have found no evidence that adjectivized NPs can embed additional NP constituents and thus conclude that such constructions are not grammatical.

Compounding is a productive process, cf. *cov-ap'* [sea-shore] 'sea shore'. Dum-Tragut (2009:671-673) does not mention any instances of recursive compounding . We thus conclude that juxtaposition is not available as a strategy for recursive NP embedding.

7 Baltic

7.1 Latvian

Latvian has two constructions that allow the recursive embedding of NPs: genitives and adpositions. The language further displays adjectivizers and juxtaposition, but these constructions do not allow the recursive expansion of NPs. There is no head-marking construction for the embedding of NPs.

Latvian nouns are inflected for case, one of which is the genitive. It is marked by a case suffix on the noun. The phonological shape of the genitive suffix varies depending on the declension class of its head noun (Prauliņš 2012:27-35). Genitive-marked nouns can embed additional NPs:

(16) *man-as mās-as draug-a sun-s*
1SG.POSS-F.GEN.SG sister(F)-GEN.SG boyfriend-GEN.SG dog-NOM.SG
[[[[NP-A] NP-G] NP-G] N]
'my sister's boyfriend's dog' (Jurgis Pakerys p.c.)

There are a range of different prepositions, which govern the accusative, dative, or genitive (Prauliņš 2012:169-173). As mentioned above, prepositions allow NP recursion:

(17) *zēn-s no ciem-a pie upe-s*
boy-NOM.SG from village-GEN.SG at river-GEN.SG
[N [NP-P [NP-P]]]
'the boy from the village by the river' (Jurgis Pakerys p.c.)

Latvian has an inventory of suffixes that derive adjectives from nouns, e.g. *-isk-* (*zinātn-isk-s* [science-ADJZ-NOM.SG.M] 'scientific' < *zinātne* 'science') or *-īg-* (*priec-īg-s* [happiness-ADJZ-NOM.SG.M] 'happy' < *prieks* 'happiness') (Prauliņš 2012:74-75). These suffixes cannot be used to build recursive NP constructions (Andra Kalnača p.c.).

Finally, Latvian makes use of nominal compounding, e.g. *viesistaba* 'living room' < *vies(i)-* 'guest' + < *istaba* 'room' (Prauliņš 2012:50). Recursive nominal compounding, however, is not common in the language (Andra Kalnača p.c.), which makes juxtaposition unavailable as a strategy for recursive NP embedding.

7.2 Lithuanian, Modern

Lithuanian has four of the five constructions we surveyed: genitives, adjectivizers, adpositions, and juxtaposition. Of these, genitives and adpositions allow NP recursion, while adjectivizers and juxtaposition do not. Head marking is absent altogether.

The genitive case is expressed by a suffix. In the singular, its form varies depending on the declension class of the noun: *-(i)o/-(i)aus/-(i)os/-(i)es/-s*. In the plural, the genitive is marked by *-(i)ų* throughout (Ambrazas 1997:122-123). Genitives are not restricted with regard to NP recursion and are able to embed other constituents such as other genitives:

(18) *mano* *seser-s* *vaikin-o* *šuo*
1SG.POSS.GEN.SG sister-GEN.SG boyfriend-GEN.SG dog.NOM.SG
[[[[NP-G] NP-G] NP-G] N]
'my sister's boyfriend's dog' (Jurgis Pakerys, p.c.)

There are several prepositions which assign either genitive, accusative or instrumental case to the noun they govern. Some of them may also be used as postpositions (Ambrazas 1997:404-407). As mentioned above, they allow NP recursion:

(19) *berniuka-s* *iš* *kaim-o* *prie* *upė-s*
boy-NOM.SG from village-GEN.SG at river-GEN.SG
[N [NP-P [NP-P]]]
'the boy from the village by the river' (Jurgis Pakerys, p.c.)

Lithuanian has a number of adjectivizers, e.g. *-išk-* as in *vyr-išk-as* [man-ADJZ-NOM.SG.M] 'male, masculine' (Ambrazas 1997:160). We have found no evidence for denominal adjectives that embed further NP constituents and thus conclude that adjectivizers cannot be used to build recursive NP constructions.

There are determinative compounds, such as *savait-galis* [week-end] 'weekend' (Ambrazas 1997:129). Compounds appear to be restricted to two members, and so juxtaposition is not available as a strategy for recursive NP embedding.

7.3 Lithuanian, Old

Old Lithuanian refers to the Lithuanian spoken between 1500-1700 CE in Prussia and the Polish-Lithuanian Commonwealth. The language is attested in Bible translations and other religious texts, most of which are translated from German or Polish.

In Old Lithuanian, there are genitives, adjectivizers, adpositions, and juxtaposition (Stang 1966, Schmalstieg 1987). Genitives and adpositions allow the recursive embedding of NPs, while this is not the case for adjectivizers and juxtaposition. Head marking is absent altogether.

Nouns are inflected for case, one of which is the genitive. It is marked by a suffix, which varies depending on the number and declension class of the noun. The genitive allows NP recursion, as illustrated in Example 20.

(20) *bals-ą* *sunaus* *diev-o*
voice-ACC.SG son.GEN.SG god-GEN.SG
[N [NP-G [NP-G]]]
'the voice of the son of god' (John 5.25, Bitner 1701)

Old Lithuanian has a number of adpositions that govern case (Schmalstieg 1987:273-298). Like genitives, they can embed further NPs:

(21) *krisl-ą* *iš* *tavo* *brolio* *akes*
splinter-ACC.SG out 2SG.GEN brother.GEN.SG eye.GEN.SG
[N [NP-P [[NP-G] NP-G]]]
'a splinter out of your brother's eye' (Matthew 7.5, Bretke 16th cent.)

There are adjectivizers, as in *tėv-išk-as* [father-ADJZ-NOM.SG] 'fatherly' (Endzelīns 1971:131). However, we have found no evidence that they allow the recursive embedding of NPs.

Old Lithuanian also has determinative compounds, e.g. *karal-krėslis* [king-chair] 'throne' (Endzelīns 1971:77-85), but they are, as far as we are able to establish, confined to two members. In the absence of other juxtapositions, we conclude that the type is not available as a strategy for recursive NP embedding in the language.

8 Celtic (Insular)

8.1 Breton, Middle

Middle Breton refers to the Breton spoken in Brittany between the 12th and 17th century CE. Up to the end of the 15th century, the langue is almost solely attested in glosses and proper names. After that, an ever-growing corpus of poems, plays, and prose texts of mostly religious character has been handed down to us.

In Middle Breton, there are genitives, adjectivizers and juxtapositions, but no adpositions or head marking. Adjectivizers do not allow NP recursion, but genitives and juxtapositions do.

Middle Breton has lost the Insular Celtic genitive along with the entire case morphology (Hemon 1975, Schrijver 2011) but has reanalyzed some former prepositions (e.g. *a* and *da*) into new phrasal case markers (Widmer 2017). They appear at the left edge of the NP and cannot be stranded. Thus, they are analyzed as genitives in our coding scheme. Genitives allow NP recursion:

(22) *an* *mister* *a-n* *incarnation* *a* *map* *doue*
ART mystery GEN-ART incarnation GEN son god
[N [NP-G [NP-G [NP-Ø]]]]
'the mystery of the incarnation of the son of god' (Ernault 1887: § 6)

Middle Breton has a restricted number of suffixes that derive adjectives from nouns, such as *-us* in *corff-us* [body-ADJZ] 'corporeal' (Hemon 1976–1998:1596). This kind of word formation is of limited productivity, and the derived adjectives cannot embed further constituents.

An NP can also be embedded into another NP by juxtaposition, i.e. without any further marking (Hemon 1975:45-46). As a rule, the head precedes the dependent and the dependent allows further recursive NP embedding:

(23) *a tut iesu a nazareth*
GEN people Jesus GEN Nazareth
[NP-G [NP-Ø [NP-G]]]
'of the followers of Jesus of Nazareth' (Le Berre 2011: l. 1663)

8.2 Breton, Modern

Modern Breton, like Middle Breton, uses genitives and juxtaposition to recursively embed NPs. There are also adjectivizers, but they do not allow NP recursion. Head marking is absent, and there are no adnominal adpositions.

Modern Breton has preserved the genitive that has developed from a preposition in Middle Breton (see Section 8.1) (Press 1986:212-213). Genitives allow NP recursion:

(24) *ur plac'hig yaouank a-n oad a bemzek vloaz*
a girl young GEN-ART age GEN 15 year
[N [NP-G [NP-G]]]
'a young girl of the age of fifteen years' (Hemon 1976–1998:34)

Adjectivizers exist, e.g. *tad-el* [father-ADJZ] 'paternal' (Vallé 1980:XXI). However, we have found no evidence in the literature that adjectivized NPs can embed further NP constituents. We thus conclude that this is not possible.

Modern Breton can form complex NP constructions by juxtaposing nouns without any further marking. The relevant constructions can recursively embed NP constituents:

(25) *toenn ti ar pesketaer*
roof house ART fisherman
[N [NP-Ø [NP-Ø]]]
'the roof of the fisherman's house' (Press 1986:211)

8.3 Irish, Old

Old Irish refers to the stage of the Irish language spoken from the 8th to the 9th century CE. A small portion of the texts, mostly glosses and commentaries on Latin texts, is contained in manuscripts of that period stored on the continent. The bulk of transmitted texts covers a wide range of genres (historical, legal, narrative, religious) and only survives in Irish vellum and paper manuscripts from later periods.

In Old Irish, there are two constructions that allow the recursive embedding of NPs: genitives and adpositions. Juxtaposition and adjectivizers exist, but do not allow NP recursion. There is no head-marking construction.

Old Irish has five cases, one of which is the genitive. It is expressed with a suffix accompanied by consonant / vowel alternations — e.g. *muir* sea.NOM.SG : *mor-o* sea.GEN.SG — in some declension classes. In others, it is marked by consonant / vowel alternations only, e.g. in the *o*-stem noun *fer* man.NOM.SG : *fir* man.GEN.SG (Thurneysen 1946:176–217).

The genitive allows NP recursion:

(26) *rúin* *ícc-e* *in-cheneli* *dóine*
mystery.ACC.SG salvation-GEN.SG ART.GEN.SG-race.GEN.SG man.GEN.PL
[N [NP-G [NP-G [NP-G]]]]
'the mystery of the salvation of the race of men' (Würzburg glosses 21d11)

Old Irish has a number of prepositions, which govern case (Thurneysen 1946:495–537). Such prepositional phrases allow the recursive embedding of additional NPs:

(27) *dechur* *eter* *corpu* *nem-d-i*
difference.NA.SG between body.ACC.PL heaven-ADJZ-ACC.PL
[N [NP-P [NP-A]]]
'the difference between heavenly bodies (and earthly bodies)' (Würzburg glosses 13c26)

Old Irish also has two productive adjectivizers, the suffixes *-de* and *-ach*: e.g. *nem-de* [heaven-ADJZ] 'heavenly' (as in Example 27) and *airther-ach* [eastern-ADJZ] 'eastern' (Thurneysen 1946:220–223). We have not found any evidence that adjectivized NPs can embed additional NPs and conclude that this is not possible.

The language makes ample use of compounds of all sorts, including determinative compounds. They are, however, generally limited to two members, as in *talam-chumscugud* [earth-quake] 'earth quake' (Thurneysen 1909:161). Accordingly, juxtaposition is not among the strategies that Old Irish make available for recursive NP embedding.

8.4 Irish, Modern

Modern Irish has genitives, adjectivizers, adpositions and juxtaposition, but there is no head marking. Adjectivizers do not allow NP recursion, but all other constructions do.

Like in Old Irish, nouns are inflected for case, one of which is the genitive. It is marked by consonant / vowel alternations (e.g. *fear* man.NOM.SG vs. *fir* man.GEN.SG) and / or a suffix (e.g. *muc* pig.NOM.SG vs. *muice* pig.GEN.SG) (Ó Siadhail 1989:149,151). Genitives allow recursive NP embedding:

(28) *teach* *athair* *an* *chara*
house father.GEN.SG ART friend.GEN.SG
[N [NP-G [NP-G]]]
'the house of the friend's father' (Mícheál Hoyne, p.c.)

Modern Irish has a number of prepositions, some of which govern case. The preposition *chun* 'to', for example, governs the genitive: *chun báis* [to death.GEN.SG] 'to death' (Ó Siadhail 1980:155). There is a case form (mostly referred to as dative or prepositional case) which is used solely after prepositions, e.g. *in Éirinn* [in Ireland.DAT] ‘in Ireland’. In earlier stages of the language, this was a productive dative case (Hickey 2011:279). Preposition-marked NPs can embed further NPs:

(29)

an	*chailleach*	*gan*	*fiacla*	*in*	*-a*	*béal*
the	witch	without	teeth	in	-3SG.F.GEN	mouth
[	N	[NP-P		[NP-P	[NP-G	]]]]

‘the witch without teeth in her mouth’ (Mícheál Hoyne, p.c.)

Adjectivizers exist, but they cannot be used for NP recursion (Mícheál Hoyne, p.c.).

Modern Irish can form complex NPs by juxtaposing two NPs in nominative case. Such strings of juxtaposed NPs can be expanded recursively, as the following example illustrates:

(30)

mac	*uncail*	*m-*	*athar*
son.NOM.SG	uncle.NOM.SG	1SG.GEN-	father.NOM.SG
[N	[NP-Ø	[[NP-G]	NP-Ø]]]

‘my father’s uncle’s son’ (Mícheál Hoyne, p.c.)

8.5 Welsh, Middle

Middle Welsh refers to the Welsh spoken from the mid 12^{th} century CE to the 15^{th} century CE. It is attested in a large number of texts, spanning narratives, translations, legal documents, religious texts and more (Willis 2009:118).

Middle Welsh has genitives and juxtaposition, both of which allow the recursive embedding of NPs. There are also adjectivizers, but they cannot embed other NPs. There are no adnominal adpositions and head marking is absent as well.

While the inherited case morphology was completely lost (Evans 1964), Middle Welsh has developed a new genitive construction from adpositions. The most common one is the phrasal prefix *o*, e.g. *heit o wenyn* [swarm GEN bee.PL] ‘a swarm of bees’ (Evans 1964:204), but other phrasal case prefixes exist as well (e.g. *y robert* ‘to/of Robert’). The new genitives allow recursively expanded NPs:

(31)

a	*thri*	*ugein-wyr*	*etholedigyon*	*o*	*degeingyl*	*o*	*gyuoeth*	*y*	*robert*
and	three	twenty-man.PL	select.PL	GEN	Degeingyl	GEN	kingdom	to	Robert
	[	N		[NP-G	]	[NP-G		[NP-G	]]]

‘and sixty select men of Degeingyl[4], of the kingdom of Robert’ (Peniarth 17.5)

Adjectivizers exist, for example *teyrn-aidd* [king-ADJZ] ‘royal, kingly’ (Zimmer 2000:478), or *neu-awl* [heaven-ADJZ] ‘heavenly’ (Morris Jones 1913:255), but they are not reported to be used for NP recursion.

NPs can modify other nouns without any marking of the embedding relationship, i.e. by juxtaposition (Willis 2009:133). Such constructions can embed further NPs:

[4] A place name referring to the region of modern-day Flintshire.

(32) *drws pebyll y brenhyn*
door tent ART king
[N [NP-Ø [NP-Ø]]]
'the king's tent's door' (Brut Dingestow t. 11.6)

8.6 Welsh, Modern

In Modern Welsh there are adpositions, adjectivizers and juxtaposition. Adpositions and juxtaposition both allow the recursive embedding of NPs, while this is not so for adjectivizers. There is no head-marking construction and no genitive construction.

The adposition *o* precedes the dependent and may host possessive markers and articles (Borsley et al. 2007:72). This construction allows NP recursion:

(33) *disgrifaid o-r rhes o dai*
description from-ART row from houses
[N [NP-P [NP-P]]]
'the description of a row of houses' (Borsley et al. 2007:72)

Recently, colloquial Modern Welsh has newly introduced preposition stranding of *o*, cf. Example 34, which was not licensed from Middle Welsh times up until the twentieth century CE. This innovation most certainly results from language contact with English. Thus, the language has re-assigned adpositional properties to *o*:

(34) *Lle 'dach chi 'n dod o?*
where be.PRS.2PL 2PL PROG come.INF from
'Where do you come from?' (Borsley et al. 2007:116)

Modern Welsh possesses a number of suffixes that derive adjectives from nouns, e.g. *-ol* (*wythnosol* 'weekly' < *wythnos* 'week') or *-ig* (*gwledig* 'rural' < *gwlad* 'country') (King 2003:86–89). However, we have found no instances of adjectivized NPs that embed additional NPs. We thus conclude that this is not possible.

Finally, Modern Welsh commonly uses juxtaposition to embed nouns into NPs. Such constructions can be expanded recursively, as the following example illustrates:

(35) *siop mab chwaer y meddyg*
shop son sister ART doctor
[N [NP-Ø [NP-Ø [NP-Ø]]]]
'the shop of the son of the sister of the doctor' (Borsley et al. 2007:184)

9 Germanic

9.1 Afrikaans

Afrikaans uses genitives, adpositions, juxtaposition and a head-marking construction to form recursive NP structures. There are also adjectivizers, but they cannot embed further constituents.

There is a genitive case, marked by *-s* or *-e*, but it has a limited distribution and is sometimes analyzed as linker in juxtaposition or compounds (while in our definition juxtapositions have no dedicated markers, i.e. no marking effects beyond regular morphophonology or stress patterns). The suffix *-e* conveys a notion of plurality (cf. *student-e-lewe* [student-GEN-life] ‘student life’), while it is completely unpredictable whether a certain noun takes *-s* or not. Sometimes combinations exist both with the genitive and as a plain juxtaposition (see below) without any change in meaning, e.g. *oorlog(-s)-museum* [war(-GEN)-museum] ‘war museum’ (Donaldson 1993:438). This genitive can only be used with nouns that are not accompanied by determiners, adjectives or the like. The construction allows recursion:

(36) *lugdiens- besprekning-s- kantoor*
airline- booking-GEN- office
[[[NP] NP-G] N]
‘airline booking office’ (Donaldson 1993:438)

There are a large number of prepositions and they can embed further constituents, such as a head-marking construction in Example 37. They govern case, but since there is no case marking on nouns, this can only be seen in combination with pronouns, which distinguish between a subject form and an object form (Donaldson 1993:123).

(37) *’n Kennis van my oom se pa*
ART friend of 1SG.GEN uncle POSS father
[N [NP-P [[NP-G] NP] N-H]]
‘a friend of my uncle’s father’ (Donaldson 1993:98)

Afrikaans possesses several adjectivizers, e.g. *-(e)rig* as in *korrel-rig* [grain-ADJZ] ‘grainy’ (Donaldson 1993:442), but we found no evidence that adjectivized NPs can embed additional NP constituents. Accordingly, it appears that such constructions are not grammatical.

The language also has juxtaposition, e.g. *waternood* ‘shortage of water’ (*water* ‘water’ + *nood* ‘need’) (Donaldson 1993:438). Such constructions allow NP recursion:

(38) *misdaad- toneel- bestuur*
crime- scene- managment
[[[NP-Ø] NP-Ø] N]
‘crime-scene managment’ (Alena Witzlack-Makarevich, p.c.)

Furthermore, there is a head-marking construction based on the possessive marker *se*. The particle *se* goes back to a third person singular possessive pronoun (Roberge 1996). This construction allows the recursive embedding of NPs:

(39) *ons bur-e se vriend-e se seun*
our neighbor-PL POSS friend-PL POSS son
[[[N] N-H] N-H]
‘our neighbors’ friends’ son’ (Donaldson 1993:98)

Evidence that *se* is a head marker comes from the fact that it shows allomorphy which depends on the structure of the head, not the dependent. In constructions in which the head is empty (zero), the allomorphs *s’n* or *s’ne* have to be used:

(40) *dis* *Amanda* *se* *ma* *s'n.*
this.be.3SG Amanda POSS mother POSS
[[[N] N-H] N-H]
'It's Amanda's mother's.' (Donaldson 1993:100)

9.2 English, Old

Old English refers to the English that was spoken from the 7th to the 11th century CE in modern-day England and southern Scotland. It is attested in inscriptions and manuscripts (Pilch 1970:28).

Old English has genitives, adjectivizers, adpositions, and juxtaposition as means for embedding NPs into NPs. There is no head-marking construction. Genitives, adpositions, and juxtaposition can be used for NP recursion.

Old English nouns are inflected for case, one of which is the genitive. The singular is marked by *-es/-e* in what is known as the strong declension class and by *-an* in the weak declension class. The *an*-allomorph is not a specialized genitive case, but a more general oblique: in the singular, it marks everything except the nominative singular (and accusative with feminines), in the plural it covers all cases except the genitive and dative. The genitive plural is marked by *-a* and *-en-a* (Pilch 1970:103-104). Genitives allow NP recursion:

(41) *xl* *monn-a* *his* *her-es*
forty man-GEN.PL 3SG.GEN army-GEN.SG
[N [[NP-G] NP-G]]
'forty men of his army' (Anglo-Saxon Chronicle 878)

Old English has several prepositions, some of which govern case (Cassidy & Ringler 1971:92). Like in many other IE languages, adposition-marked NPs allow recursive embedding of further NPs:

(42) *to þæm* *lond-um* *on* *þa* *healf-e* *munt-es*
to DEM.DAT.PL country-DAT.PL on DEM.ACC.SG.F side(F)-ACC.SG mountain-GEN.SG
[N [NP-P [NP-G]]]
'to the countries on the (other) side of the mountain' (Anglo-Saxon Chronicle 887)

Adjectivizers, by contrast, cannot recursively embed NPs. Old English has a number of adjectivizers, e.g. *-en* as in *æsċ-en* [ash-ADJZ] 'made of ash-wood' and *-iġ* as in *blōd-iġ* [blood-ADJZ] 'bloody' (Kastovsky 2006:241–242), but we found no evidence that they can be used to build recursive NP constructions.

Compounds are frequent and productive in Old English. Determinative compounds can have three members (Carr 1939), as in Example 43. Thus, juxtaposition allows the recursive embedding of NPs.

(43) *dēoful-* *gyld-* *hūs*
devil- guild- house
[[[NP-Ø-] NP-Ø-] N]
'heathen temple' (Carr 1939:199)

9.3 English, Middle

Middle English refers to the English used from about 1100 to about 1500 CE. Texts from the 11th and 12th century CE still show many parallels to the Old English standard language. Later documents often reflect the dialect of the scribe more closely (Brunner 1965:1).

Middle English uses genitives, adjectivizers, adpositions, head marking, and juxtaposition. Except for adjectivizers, all of these constructions allow the recursive embedding of NPs.

Nouns are inflected for case, although the system is greatly reduced compared to Old English. The genitive is marked by a suffix *-e/-es/-s*. During the Middle English period, the genitive case suffix was gradually reanalyzed as a phrasal clitic (Allen 2008:152-154). As a consequence, the genitive no longer had to be marked on its head noun, but could also appear on the right edge of the NP. The Middle English genitive can embed additional NPs, regardless of its syntactic placement:

(44) a. *my neighebore-s wyf*
1SG.GEN neighbor-GEN wife
[[[NP-G] NP-G] N]
'my neighbor's wife' (The Wife of Bath's Prologue, 242)

b. *God of Love-s servant-z*
God of Love-GEN servant-PL
[NP-G [NP-P]] N]
'the servants of the God of love' (Troilus and Criseyde, 1.15)

Middle English has a number of adpositions. These generally precede the noun they modify, but they can also follow it (Mossé 1952:125). Case government is only visible in pronouns, where adpositions govern the object form, e.g. *of mē* 'of me' vs. *i/ic* 'I' (Mossé 1952:54). Adpositions allow NP recursion:

(45) *alle the lord-is of my reme*
all ART.DEF lord-PL of 1SG.GEN realm
[N [NP-P [NP-G]]]
'all the lords of my realm' (Chronicle of the Reign of Henry IV, Anno A.D. 1401–2)

Middle English has a number of adjectivizers which form denominal adjectives. They are invariable, as only monosyllabic adjectives inflect (Mossé 1952:64), and cannot be used for NP recursion.

Nominal compounding appears to be less productive in Middle English than in Old English and Early Modern English (Van Gelderen 2014:139, Sauer 1992:7). Nevertheless, Middle English possesses numerous determinative compounds, e.g. *palm-twig* 'palm twig' (Sauer 1992:149). Juxtapositions of this kind are attested with recursive embeddings:

(46) *All- halow- masse- day*
all- saint- mass- day
[[[NP-Ø] NP-Ø] N]
'All Hallows' (Sauer 1992:324)

In addition to these inherited NP types, Middle English also has a head marking strategy to integrate NPs into NPs, the *his*-construction. This construction emerged during the Early Middle English period and gradually became obsolete again during the Early Modern English period, giving way to the *'s*-clitic.[5] The *his*-genitive can be used for NP recursion:

(47) *to fortefy hys brethren ys sayyngys*
to strengthen 3SG.GEN brethren ys comments
[[[NP-G] NP] N-H]
'to strengthen his brethren's comments' (Allen 2008:247)

9.4 English, Modern

Modern English by and large continues the NP types that were already present in Middle English, i.e. genitives, adpositions and juxtaposition. Adjectivizers exist, but they cannot embed further constituents. The Middle English *his*-genitive, a head marking strategy, fell into disuse during the Early Modern English period (Allen 2008:265).

The genitive clitic *-s* allows NP recursion, regardless of its position:

(48) a. *Peter's mother's house*
[[[NP-G] NP-G] N]

b. *the queen of England's crown*
[[NP-G [NP-P]] N]

Modern English has a number of prepositions, which can be used for NP recursion:

(49) *the board of directors of the company*
[N [NP-P [NP-P]]]

Prepositions can be stranded (e.g. *the bar I gave you the name of*), and they assign case (although case is visible only in pronouns, cf. *of me* vs. *I*).

Modern English adjectivizers such as *-al*, e.g. *nation-al* [nation-ADJZ] and *-ly* as in *father-ly* [father-ADJZ] are not able to embed further constituents and block fully-fledged recursion, i.e. the nominal base of adjectives is not accessible for unconstrained and recursive modification (e.g. **American Presidential plane*).[6]

Nominal compounding is a highly productive process of word formation in Modern English. Nominal compounds can be expanded recursively:

(50) *post office building*
[[[NP-Ø] NP-Ø] N]

[5] It is often assumed that the *'s*-clitic is the etymological continuation of the *his*-construction, but Allen (2008:250) shows that the two constructions evolved separately from each other, at roughly the same time.

[6] There is a rich literature on how to model what is possible and what not in this area, usually under the rubric of lexical integrity and discussions of the interface between morphology and syntax. We do not wish to enter this debate here. All that matters for our purposes is that English adjectivization does not tolerate unconstrained recursive embedding of the kind attested in such languages as Upper Sorbian, Albanian, Tocharian, or Hindi. In whatever model the facts are cast, the differences between adjectivization in English and these languages needs to be stated, and this is all we are interested in here.

9.5 German, Old High

Old High German (OHG) refers to a German dialect conglomerate spoken between about 600 and 1050 CE. OHG is attested in a small corpus including texts of various genres (Braune 2004:1-2).

In OHG, there are genitives, adpositions, adjectivizers and juxtaposition. With the exception of adjectivizers, all of these constructions allow NP recursion. Head marking is emergent, but not productive.

Nouns in OHG inflect for case. The genitive is marked by a suffix, which has various forms depending on the number and declension class of the noun (Braune 2004:182–184). Genitives can embed other NPs:

(51) a. *min-es truhtin-es muoter*
1SG.POSS-GEN.SG.M lord(M)-GEN.SG mother.NOM.SG
[[[NP-A] NP-G] N]
'my lord's mother' (Schrodt 2004:34)

b. *gerst-un korn-es hut*
barley-GEN.SG grain-GEN.SG peel.NOM.SG
[[[NP-G] NP-G] N]
'the peel of a barley grain' (Schrodt 2004:34)

There are a number of prepositions that assign case to the NP they govern (Schrodt 2004). Like genitives, they allow the recursive embedding of NPs:

(52) *si tharben bigan thes liob-es zi iro goman*
3SG.F.NOM live.in.want.INF begin.PST.3SG ART.DEF.GEN.SG.NT love(NT)-GEN.SG to 3SG.F.GEN man.DAT.SG
[N [NP-P [NP-G]]]
'since she began to be deprived of the love of her husband' (Otfried 1.16.5) (Schrodt 2004:35)

OHG possesses a range of derivational suffixes that derive adjectives from nominals, e.g. *-isc* as in *himil-isc* [heaven-ADJZ] 'heavenly' or *-līh* as in *got-līh* [god-ADJZ] 'divine' (Splett 2000:1219–1220). It appears, however, that adjectivized NPs cannot embed further NPs. In any case, we have found no evidence for such constructions.

Determinative compounds, such as *oli-faz* [oil-vessel] 'oil vessel', are common in OHG (Carr 1939:128–129). Juxtaposition of this kind allows recursive expansion:

(53) *hasal- nuz- cherno*
hazel- nut- kernel
[[[NP-Ø] NP-Ø] N]
'hazelkernel' (Carr 1939:197)

There are indications for an emergent head-marking construction involving pronouns as co-constituents of the head, but all examples are ambiguous between an adnominal and adverbal interpretation, e.g. Example 54.

(54) *demo* *balder-es* *uolon* *sin* *uuoz*
ART.DEF.DAT.SG Balder-GEN.SG foal.DAT.SG 3SG.M.GEN foot.NA.SG
[[[NP-G] NP] [NP-G] N]
'the foot of Balder's foal' (Fleischer & Schallert 2011:97)

Thus, even if head marking existed, it was a marginal strategy at best and developed only later, for example in Swiss German (on which see Section 9.12).

9.6 German, Middle High

Middle High German (MHG) refers to the German dialects spoken in southern Germany between 1050 and 1500 CE (Paul 1998:10).

MHG has all five of the surveyed structures, but only genitives and adpositions are available for NP recursion.

Nouns in MHG inflect for case. The genitive is marked by a suffix, which varies depending on the declension class and number of the noun. Masculines and neuters take *-(e)s* or *-en/-in* in the singular and *-e* or *-ôno* in the plural. Feminines have *-e* or *-ûn* in the singular and *-(e)n* or *-ôno* in the plural (Paul 1998). Genitives can be used for NP recursion:

(55) *wîb-es* *oug-en* *süeze* *und* *dâ* *bî* *wîb-es* *herz-en* *suht*
woman-GEN.SG eye-GD.PL sweetness.NOM.SG and there also woman-GEN.SG heart-GD.PL illness.NOM.SG
[[[NP-G] NP-G] N] [[[NP-G] NP-G] N]
'[he was] sweetness in women's eyes and illness in women's hearts' (Parzival 4.20f.) (Prell 2005:218)

There are a number of prepositions, which govern the genitive, dative or accusative case. They can be used adnominally, e.g. *der chünic uber Isrâhêlê* [the king over Israel] 'the king of Israel', and they allow recursive expansion (Prell 2005:216):

(56) *die* *starch-en* *in* *min-ero* *christenheit-i*
ART.NOM.PL strong-NOM.PL in 1SG.POSS-DAT.SG.F christianity(F)-DAT.SG
[N [NP-P [NP-A]]]
'the strong ones in my christianity' (Wiener Notker 51) (Prell 2005:216)

There are several suffixes which derive adjectives from nouns, e.g. *-el* as in *scham-el* [shame-ADJZ] 'bashful' and *-ic* as in *nîd-ic* [envy-ADJZ] 'envious' (Klein et al. 2009:278,285). Like other adjectives, they agree with their head noun in gender, number and case (Paul 1998:208). There is no evidence that they can embed further constituents.

Compounding exists and two-member compounds like *apfel-boum* [apple-tree] 'apple tree' occur frequently. But compounding does not seem to be used for NP recursion in MHG. Though three-member compounds exist (Ruthmann 2007:4), they are rare and none of the examples can be classified as a endocentric structure with certainty. An ambiguous case is *âbent-sunnen-schîn* [evening-sun-shine] 'evening sunshine', which can either be interpreted as non-recursive, with 'evening' modifying the compound *sunnen-schîn* or as recursive, with 'evening' modifying 'sun', which in turn modifies 'shine'. We conclude that juxtaposition was at best a highly marginal strategy for NP recursion.

There is some indication for an emergent head-marking construction involving a possessive pronoun as a co-constituent of the head noun. However, all relevant examples are ambiguous between an adnominal and an adverbal reading, as is the case for example in Example 57a. And even such ambiguous examples are rare. The first unambiguous specimens of this construction, such as the example in Example 57b, appear in a text from 1515, i.e. shortly after the MHG period (Fleischer & Schallert 2011:97-98):

(57) a.

Pharien-s	*wip*	*besach*	*yr-en*	*man*	*sin*	*wund-en*
Pharien-GEN.SG	woman.NA.SG	examine.PST.3SG	3SG.F.POSS-DAT.SG	man.DAT.SG	3SG.M.GEN	wound-ACC.PL
			[[[NP-A]	NP]	N-H	]

'Pharien's wife examined the wounds of her husband' (Prosalancelot 83.17) (Fleischer & Schallert 2011:97)

b.

Die	*paur-en*	*namen*	*dem*	*appt*	*von*	*Kempten*	*sein*	*Kloster*	*ein*
ART.NOM.PL	farmer-NOM.PL	take.PST.3PL	ART.DAT.SG.M	abbot(M)	of	Kempten	3SG.M.POSS.ACC.NT.SG	monastery(NT).NA.SG	in
			[[	NP	[NP-P	]]	N-H	]	

'the farmers took the the monastery of the abbot of Kempten' (*Chronica newer geschichten*, 1512-1527) (Fleischer & Schallert 2011:98)

This suggests that even if the construction already existed in MHG, it was marginal at best. It developed only later into a fully productive NP type, for example in modern Swiss German dialects (see Section 9.12).

9.7 Gothic

Gothic was spoken from the end of the 1st century to the 6th century CE. The language is almost exclusively attested in translations of the Greek New Testament from the 4th century CE. It is the only attested East Germanic language (Kotin 2012:13-21).

Gothic uses genitives, adjectivizers, adpositions, and juxtaposition to embed nominals into noun phrases. Genitives and adpositions allow NP recursion, while adjectivizers and juxtaposition cannot be used to build recursive NP structures. There is no head-marking construction.

Gothic nouns are inflected for case, one of which is the genitive (Braune & Ebbinghaus 1966:59). It is marked by the suffix *-(i)s* in the singular and by *-e/-o* in the plural. Genitives can embed other constituents:

(58) *qens* *Kusin-s* *fauragaggjin-s* *Herod-es*
wife.NOM.SG Chuza-GEN.SG steward-GEN.SG Herod-GEN.SG
[N [NP-G] [NP-G [NP-G]]]
'the wife of Chuza, of Herod's steward' (Luke 1.27)

There are a number of prepositions, which govern case (Braune & Ebbinghaus 1966:124). As mentioned above, they can be used for NP recursion:

(59) *Iosef* *us* *garda* *Daweid-is*
Joseph.NOM.SG from house.DAT.SG David-GEN.SG
[N [NP-P [NP-G]]]
'Joseph of the house of David' (Luke 1.27)

Gothic has a number of adjectivizers, e.g. *-isk-* as in *gud-isk-s* [god-ADJZ-NOM.SG] 'godly' and *-(e)in-* as in *aírþe-in-s* [earth-ADJZ-NOM.SG] 'earthen' (Wright 1954:178). There is no evidence that adjectivized NPs can embed additional NPs.

Gothic has determinative compounds, such as *auga-dauro* [eye-door] 'window', but none of them have more than two members (Carr 1939:197, Braune & Ebbinghaus 1966:62). Since there is no other juxtaposition construction for NPs, we conclude that this strategy was not available for NP recursion in Gothic.

9.8 Icelandic

Icelandic uses genitives, adjectivizers, adpositions, and juxtaposition for NP embedding, but there is no head marking. Adjectivizers do not allow NP recursion, but all the other constructions do.

Nouns are inflected for case, and there are several declension classes. The genitive singular is marked by a suffix which varies depending on the declension class of the noun. The two most common forms are *-s* (predominantly for masculines) and *-ar* (predominantly for feminines). The genitive plural is marked by *-a* throughout (Kress 1982:55ff.). Genitive-marked nouns can embed additional NPs:

(60) *hund-ur* *kærast-a* *systur* *minnar*
dog-NOM.SG boyfriend-GEN.SG sister(F).GEN.SG 1SG.POSS.GEN.SG.F
[N [NP-G [NP-G [NP-A]]]]
'my sister's boyfriend's dog' (Guðrún Þórhallsdóttir, p.c.)

There are a number of prepositions in Icelandic which govern case (Kress 1982:187–188). Like genitives, they allow NP recursion:

(61) *fugl-inn* *á* *grein-inni* *yfir* *tjörn-inni*
bird(M).NOM.SG-DEF.NOM.SG.M on branch(F).DAT.SG-DEF.DAT.SG.F above pond(F).DAT.SG-DEF.DAT.SG.F
[N [NP-P [NP-P]]]
'the bird on the branch over the pond' (Guðrún Þórhallsdóttir, p.c.)

There are three suffixes that derive adjectives from nouns: *-leg-*, *-sk-* and *-ug-*, e.g. *íslen-sk-ur* [Iceland-ADJZ-NOM.SG.M] 'Icelandic' (Thráinsson 1994:163). We found no evidence that the resulting denominal adjectives cannot embed additional NPs.

Compounds are very productive and noun-noun compounds are the most frequent type. Icelandic has both stem compounds, whose non-heads are bare stems (e.g. *snó-hús* [snow-house.NOM.SG] 'snow house'), and genitive compounds, whose non-heads are genitive forms (e.g. *barn-a-skóli* [child-GEN.PL-school.NOM.SG] 'children's school') (Thráinsson 1994:165). Both types of juxtaposition allow NP recursion:

(62) a.

járn-	*stál-*	*fótur*
iron	chair	leg
[[[NP-Ø]	NP-Ø]	N]

'leg of an iron chair' (Harðarson 2016:4)

b.

garð-s-	*vegg-jar-*	*staur*
garden-GEN.SG-	wall-GEN.SG-	stake.NOM.SG
[[[NP-G]	NP-G]	N]

'garden fence stake' (Guðrún Þórhallsdóttir, p.c.)

Note that these expressions can also have other, non-recursive readings (Harðarson 2016). For example, the expression in Example 62a can also be interpreted as 'iron [chair leg]', with 'iron' modifying the composite consituent 'chair leg'.

According to Thráinsson (1994:165), NP recursion with compounds is more common with genitives. However, the existence of examples such as those in Example 62 indicates that recursive compounding is also possible with plain juxtaposition, which is the only kind of juxtaposition recognized by our taxonomy. (We classify genitive-marked compounds as genitives.)

9.9 Old Norse

Old Norse refers to the language used in Norway, Iceland, the Faroes and in the Norse settlements in the British Isles and Greenland from the early 9th to the late 14th century CE. It is also referred to as Medieval West Nordic. The language is attested in manuscripts from Iceland and Norway. Sources from the other areas are very scarce: there are a few inscriptions and very few manuscripts (Faarlund 2004:1-2).

Old Norse uses genitives, adjectivizers, adpositions, and compounds to embed NPs into NPs. Of these, genitives and adpositions can be used for NP recursion. There is no head-marking construction.

Old Norse nouns are inflected for case. The genitive has different forms depending on the declension class and number of the noun. Genitive-marked nouns can embed additional NPs:

(63)

öl	*alda*	*sona*
beer.NOM.SG	mankind.GEN.PL	son.GEN.PL
[N	[[NP-G]	NP-G]]

'the beer of the sons of men' (Hávamál 12)

Old Norse has a number of prepositions, which govern case. Prepositions usually take the the accusative when expressing direction and the dative when indicating location. Other prepositions govern the genitive case (Faarlund 2004:116-117). Preposition-marked NPs can embed additional NPs:

(64) *brautir* *til* *Óðins* *landa*
road.ACC.PL to Odin.GEN.SG land.GEN.PL
[N [NP-P [NP-G]]]
'the road to Odin's realm' (Edda Hárbarðsljóð 56)

There are a number of adjectivizers that derive adjectives from nouns, e.g. *-sk-* as in *íslenzkr* 'Icelandic' derived from *Ísland* 'Iceland', and *-lig-* as in *konung-lig-r* [king-ADJZ-NOM.SG.M] 'royal, kingly' (Hægestads & Torp 1909:XXVIII-LXIII). There is no evidence, however, that the resulting adjectives can recursively embed additional NPs.

It is unclear whether compounds allow recursive expansion. According to Carr (1939:200-201), compounds consisting of more than two members are rare in Old Norse. Moreover, many of the relevant constructions are based on genitive-marked nouns, as in Example 65a. However, Carr (1939) also lists examples in which the individual compound members are not marked for genitive case, such as Example 65b, but these examples show no evidence of recursion.

(65) a. *guðs-* *reiðis-* *verk*
God.GEN.SG- anger.GEN.SG- act.NOM.SG
[[[NP-G] NP-G] N]
'act rousing the anger of God' (Carr 1939:200)

b. *hǫfuð-* *rað-* *gjafi*
head- advice- giver.NOM.SG
[[NP-Ø] [[NP-Ø] N]
'chief counsellor' (Carr 1939:201)

It thus appears that unmarked juxtaposition was at best only beginning to allow NP recursion in Old Norse. This development was completed in Modern Icelandic (Carr 1939:200).

9.10 Old Saxon

Old Saxon was spoken in today's northern Germany between the 6th and 11th century CE. The language is attested in manuscripts, most of them dating back to the 9th century CE. Most of the texts, including the well-known *Heliand*, are of a religious nature (Cathey 2000:7).

Old Saxon nouns are inflected for case, one of which is the genitive. It is marked by a suffix which has different forms depending on the number and declension class of the noun it attaches to (Cathey 2000). Genitives can be used for NP recursion:

(66) *folk* *god-es* *engil-o*
people.NA.SG god-GEN.SG angel-GEN.PL
[N [[NP-G] NP-G]]
'the folk of god's angels' (Heliand 1115) (Behaghel 1897:112)

There are several prepositions which govern case, usually the dative, accusative or instrumental. They can be used adnominally and can embed further constituents:

(67) *thea* *hoh-on* *burg-i* *umbi* *Sodom-o* *land*
ART.DEF.NA.PL high-NA.PL castle-NA.PL around Sodom-GEN.SG land.NA.SG
[N [NP-P [NP-G]]]
'the high castles of the land of Sodom' (Heliand 4367) (Behaghel 1897:119)

Old Saxon has several suffixes that derive adjectives from nouns, e.g. *-isk*, as in *kind-isk* [child-ADJZ] 'young', and *-în*, as in *bôm-in* [tree-ADJZ] 'wooden' (Cathey 2000). There is no evidence that adjectivized NPs can embed additional NPs.

Nominal compounding is frequent in Old Saxon. Examples of determinative compounds are *meri-strôm* [ocean-stream] 'ocean stream' and *gast-seli* [guest-hall] 'guest hall' (Cathey 2000:55-56). Examples like the following suggest that this type of juxtaposition allows recursive NP embedding:

(68) *erth-* *lîƀ-* *giskapu*
earth- life- fate
[[[NP-Ø] NP-Ø] N]
'fate of life on earth' (Carr 1939:197)

9.11 Swedish

Swedish uses genitives, adjectivizers, adpositions, and juxtaposition to embed NPs into NPs, but has no head-marking construction. NP recursion is possible with genitives, adpositions and juxtaposition.

Swedish has almost entirely lost case inflection on nouns (Holmes & Hinchliffe 2013:35ff.). The only case form that survives is the genitive, which is marked by the clitic =*s*. The genitive marker was originally a stem affix, but later developed into a phrasal affix that commonly occurs at the right edge of the NP, but can also be placed on the head if the head is not the last constituent in the NP (cf. Norde 2013). Regardless of its syntactic position, the genitive-marked constituents can embed additional NPs:

(69) a. *drottning-en s av* *England* *vapen*
queen-DEF GEN of England coat.of.arms
[[NP-G [NP-P]] N]
'the queen of England's coat of arms' (Norde 2013:303)

b. *drottning-en av* *England=s* *krona*
queen-DEF of England=GEN crown
[[NP-G [NP-P]] N]
'the queen of England's crown' (Norde 2013:302)

Swedish also uses adpositions as a means for NP building. There are a number of prepositions, which are usually unstressed and a small number of postpositions, which are usually stressed (Holmes & Hinchliffe 2013:154-155). Adposition-marked NPs allow recursive expansion:

(70)

Herr	*Ek*	*från*	*vårt*	*svensk-a*	*kontor*
Mister	Ek	of	1PL.POSS.SG.NT	swedish-DEF	office(NT)
[N		[NP-P	[NP-A	]	]]

'Mister Ek of our Swedish office' (Holmes & Hinchliffe 2013:474)

There are several adjectivizers that derive adjectives from nouns, e.g. *-sk* as in *Stockholm-sk* [Stockholm-ADJZ] 'pertaining to Stockholm' or *-lig* as in *mänsk-lig* [human-ADJZ] 'human (adj.)' (Holmes & Hinchliffe 2013:624), but the resulting adjectives are not able to embed additional NPs.

Determinative compounds are productive and examples like the following suggest that this type of juxtaposition allows recursive NP embedding:

(71)

språk-	*lärar-*	*utbildning*
language-	teacher-	education
[[[NP-Ø-]	NP-Ø-]	N]

'language teacher education' (Per Baumann, p.c.)

9.12 Swiss German

Our description is based on the Bernese dialect, because the IELEX data of Chang et al. (2015) also comes from this dialect.

Bernese German has all five constructions that we investigated: genitives, adjectivizers, adpositions, juxtaposition and head marking. Adjectivizers are the only NP construction that does not allow the recursive embedding of NPs.

The genitive case has a very limited distribution in contemporary Bernese German, as it can only occur on nouns that denote definite human referents. (In other dialects of Swiss German, e.g. in St. Gallen, the genitive is completely lost). Nonetheless, the genitive allows NP recursion:

(72)

Peter-s	*Brueder-s*	*Huus*
Peter-GEN	brother-GEN	house
[[[NP-G]	NP-G]	N]

'Peter's brother's house'

However, such structures are not acceptable to all speakers of Bernese German; some speakers of the city dialect consider it ungrammatical and only allow one level of embedding.

Prepositions are frequently used for NP embedding, with no constraints on recursion (cf. Example 73). Prepositions assign case, which can be seen on the article.

(73)

ds-	*Huus*	*vo-*	*m*	*Brueder*	*vo-*	*m*	*Peter*
ART.NOM.SG.NT-	house(NT)	of-	ART.DAT.SG.M	brother(M)	of-	ART.DAT.SG.M	Peter(M)
[	N	[NP-P			[NP-P		]]]

'the house of the brother of Peter'

Bernese German also has a number of adjectivizers, e.g. *-isch* as in *italiänisch* 'Italian', derived from the noun *Italie* 'Italy', or *-lech* as in *länd-lech* [country-ADJZ] 'rural'. However, the resulting adjectives block NP recursion entirely.

Juxtaposition is a productive process for NP embedding, and it allows recursion:

(74) *Fänschter- rahme- farb*
window- frame- paint
[[[NP-Ø] NP-Ø] N]
'paint for window frames'

There is also a head-marking construction involving a possessive pronoun, frequently used as an alternative to the genitive. This type also allows NP recursion:

(75) a. *dr- Mueter ir-es Huus*
ART.DAT.SG.F- mother(F) 3SG.F.POSS-NOM.SG.NT house(NT)
'mother's house'

b. *dr- Mueter vo- r Anna ire-s Huus*
ART.DAT.SG.F- mother(F) of- ART.DAT.SG.F Anna(F) 3SG.F.POSS-NOM.SG.NT house(NT)
[[NP [NP-P]] N-H]
'Anna's mother's house'

c. *er Anna ire-m Brueder si-s Huus*
ART.DAT.SG.F Anna(F) 3SG.F.POSS-DAT.SG.M brother(M) 3SG.M.POSS-NOM.SG.NT house(NT)
[[[NP] N-H] N-H]
'Anna's brother's house'

10 Greek

10.1 Greek, Ancient

Ancient Greek refers to the language used in Greece and Greek colonies from the 9th to the 5th century BCE. There were many dialects, most of which are only attested in inscriptions. The dominant literary dialect was Attic, although there are also literary works in Aeolic, Doric and Ionic. The following description is based on the Attic dialect.

Greek exhibits four of the surveyed NP structures: genitives, adjectivizers, adpositions and juxtaposition. All of these types except juxtaposition allow NP recursion. There is no head-marking construction.

Nouns are inflected for case. The genitive is marked by a suffix, which varies in form depending on the number and declension class of the noun it attaches to. In the singular, the genitive is marked either by *-s* or *-ou*, in the plural it is marked by *-ō̃n* throughout. Genitives allow the recursive embedding of NPs:

(76) *metaskʰṑn toũ oíkou tẽs mistʰṓseōs tõn*
partake.AOR.PTCP.NOM.SG.M ART.GEN.SG.M estate(M).GEN.SG ART.GEN.SG.F rent(F).GEN.SG ART.GEN.PL
[NP-G(rent) [NP-G(estate) [NP-G(children)
paídōn toũ Nikíou
child(M).GEN.PL ART.GEN.SG.M Nikias(M).GEN.SG
[NP-G(Nicias)]]]][7]
'having become part-lessee of the estate of the children of Nicias' (Isaeus 2.9)

There are a number of prepositions which govern case, many of them the accusative. Some govern different cases depending on their function (Schwyzer 1990a:432-433). Like the genitive, they can embed a further NP:

(77) *tḗn es toùs polémous hypèr tḗs patrídos andragathían*
ART.ACC.SG.F in ART.ACC.PL.M war(M).ACC.PL over ART.GEN.SG.F motherland(F).GEN.SG steadfastness(F).ACC.SG
[[NP-P [NP-P]] N]
'the steadfastness in the battles in defense of (his) country' (Thucydides, *The Peloponnesian War* 2.42.3)

Adjectivizers also exist. The most common suffix in this function is *-io/-ia*. Examples include *pátr-ios* [father-ADJZ.NOM.SG] 'paternal', *khthón-ios* [earth-ADJZ.NOM.SG] 'of the earth' and the like. While these suffixal adjectivizers do not support NP recursion, there is another adjectivizing construction that does: this involves the definite article functioning as a 'linker'. In this function, the article signals the embedding of a dependent phrase into an NP, and since the article confers adjectival properties to the embedded NP in the form of case, number and gender agreement, it qualifies as an adjectivizer in our taxonomy. Adjectivizers of this kind are frequently used for recursive embedding of NPs, for example:

(78) *eis tàs kṓmas tàs en toĩs ánkesi* (…)
to ART.ACC.PL.F village(F).ACC.PL ADJZ.ACC.PL.F in ART.DAT.PL.NT hollow(NT).DAT.PL (…)
[N [NP-A [NP-P (…)
tõn oréōn
ART.GEN.PL.M mountain(M).GEN.PL
[NP-G]]]]
'the villages in the hollows of the mountains.' (Xenophon, Anabasis 4.1.7.3)

Ancient Greek makes frequent use of compounds, but determinative compounds are the rarest type and presumably of younger age than other types (Schwyzer 1990b:453). An example is *sōmato-phýlax* [body-guard.NOM.SG] 'body-guard' (Smyth 1920:248). There are compounds with three members, but they are all of the type of *batrakho-mȳo-makhía* [frog-mouse-battle] 'battle of the frogs and mice', with two elements together modifying the third element, which means there is only one level of embedding (Smyth 1920:247). We conclude that juxtaposition was not among the strategies that Ancient Greek allowed for recursive NP embedding.

10.2 Greek, Modern

Modern Greek has genitives, adpositions, adjectivizers and juxtaposition, but no head marking. Genitives and adpositions allow NP recursion, while adjectivizers and juxtaposition do not.

Modern Greek nouns are described has having four cases – nominative, accusative, genitive and vocative – although there is only one declension class in which all of these have separate endings. In all other classes, some of the forms are syncretic. The genitive is marked by a suffix,

[7] The constituent 'the estate of the children of Nicias' is discontinuous, therefore the bracketing does not correspond exactly to the linear order of the sentence.

which has different forms depending on the declension class of its head noun. In most cases, the genitive is the only formally distinct case. In the plural, the genitive is marked by the suffix *-on* throughout. In the singular, it is marked by *-ø* with some masculine nouns, by *-s* with feminine nouns and by *-ou* with neuter nouns and some masculines (Mackridge 1985:135-139). Note that in most declension classes, there are only two different forms for the singular and the plural. In the plural, the genitive always has a separate form. In the singular, it is either the genitive or the nominative that has a separate form. As in Ancient Greek, genitives are compatible with recursive NP structures, cf. Example 79.

(79)

hē	*hypostḗria*	*tē̃s*	*oikogéneia-s*	*tou*
art.na.sg.f	support(f).na.sg	art.gen.sg.f	family(f)-gen.sg	3sg.m.poss.gen.sg
[	n	[np-g	[np-g	[np-a]]]]

‘the support of his family’ (Mackridge 1985:65)

There are a number of prepositions, all of which generally govern the accusative. There are some syntactic constructions, in which a nominative or genitive can be used after a preposition (Mackridge 1985:203-204). Like genitives, they allow the recursive embedding of NPs:

(80)

to	*spíti*	*dípla*	*sto*	*magazí*	*tēs*	*Anna-s*
art.na.nt	house(nt).na	next.to	at	store.acc.sg	art.gen.f	Anna(f)-gen
[	n	[np-p			[np-g	]]]

‘the house next to Anna’s store’ (Elisa Papathanassiou, p.c.)

There is a limited number of adjectivizing suffixes, namely *-eidēs* ‘like X’, *-énios* ‘made of X’, *-ístikos* and *-i(a)kós*, the latter being by far the most productive. There is no indication that they can embed a further constituent. Unlike in Ancient Greek, the article cannot be used as an adjectivizer anymore. Instead, genitives or adnominal adpositions are used, though the former is the more common choice (Elisa Papathanassiou, p.c.).

Modern Greek has determinative compounds, such as *petrelaio-kēlída* [oil-stain] ‘oil-slick’, but they are not very productive. Recursive compounding in Greek is quite rare in general. There are a few examples of three-member determinative compounds, e.g. *agroto-trapezo-dáneio* [farmer-bank-loan] ‘bank loan for farmers’, but this and similar examples discussed by Ralli (2013:94-95) are not recursive in our sense of embedding nouns in nouns: here *agroto-* ‘farmer’ modifies the compound *trapezo-dáneio* ‘bank loan’ as a whole and does not modify the embedded element *trapezo-* ‘bank’. Examples with recursive embedding, such *lemonó-dendro-k^{h}órap^{h}o* [lemon-tree-field] ‘field of lemon trees’ based on *lemonó-dendro* [lemon-tree] ‘lemon tree’ have “disputed acceptability” only (Ralli 2013:95). We conclude that juxtaposition does not allow NP recursion in Modern Greek.

11 Indo-Aryan

11.1 Hindi

Hindi has only one structure that allows the recursive embedding of NPs, namely adjectivizers. Adpositions exist, but they are not used adnominally. Juxtaposition also exist as a strategy, but it does not support recursive embedding. There is no head marking and no genitive.

The inherited Indo-Aryan case system survives in the form of a ternary distinction between direct, oblique, and vocative. The case forms are marked by suffixes on nouns. However, not all noun classes distinguish all three case forms. Feminine nouns, for example, no longer make a distinction in the singular (Sandahl 2000). The old genitive case has been lost. The oblique may either occur in isolation or in combination with adpositions, which occur as clitics on their head nouns. These adpositions, which have developed from relator nouns (see Bubenik 1998), specify the grammatical role of the relevant argument (Montaut 2004:56-57). Both case suffixes and adposition clitics are hardly ever used adnominally in Hindi, see the ungrammatical examples in the following:

(81) a. **tīn baj-e sabhā*
three o'clock-OBL.SG meeting.NOM.SG
intended: 'The meeting at 3 o'clock.' (Verma 1971:146)

b. **mez-par kitāb*
table-on book.NOM.SG
intended: 'The book on the table.' (Verma 1971:146)

Rather, case- and adposition-marked nouns have to be integrated into an NP by means of the adjectivizers *==k- k-* or *-vāl-*:

(82) a. *tīn baj-e=k-ī sabhā*
three o'clock-OBL.SG=ADJZ-NOM.SG.F meeting(F).NOM.SG
'The meeting at 3 o'clock.' (Verma 1971:146)

b. *mez-par=vāl-ī kitāb*
table-on=ADJZ-NOM.SG.F book(F).NOM.SG
'The book on the table.' (Verma 1971:146)

These adjectivization strategies are the only NP type that allows full recursive embedding of NPs. Example 83 illustrates the recursive use of the adjectivizer *=k-*, which agrees in case, number and gender with its immediate head:

(83) *Khannā=k-ī bahin=k-e kutt-e=k-ā nām*
Khanna=ADJZ-OBL.SG.F sister(F).OBL.SG=ADJZ-OBL.SG.M dog(M)-OBL.SG=ADJZ-NOM.SG.M name(M).NOM.SG
[[[[NP-A] NP-A] NP-A] N]
'Khanna's sister's dog's name' (Snell & Weightman 2003:66)

Apart from this, there are several other adjectivizing suffixes, most of them taken over from Sanskrit (Montaut 2004:148-149). Unlike *=k-* and *-vāl-*, these suffixes cannot embed additional NPs.

Hindi possesses determinative compounds, e.g. *gaṅgā-jal* [Ganga-water] 'Ganga water '(Montaut 2004:156). However, we have not found evidence that this kind of juxtaposition allows NP recursion.

11.2 Kashmiri

Kashmiri has adjectivizers, adpositions, and juxtaposition. There is no genitive and no head-marking construction. Only adjectivizers allow the recursive embedding of NPs.

There are adjectivizing clitics, which attach to the embedded NP and index gender and number of the head. There are different sets depending on whether the embedded NP is headed by a proper noun (Example 84a), a singular inanimate noun (Example 84b) or any other noun (Example 84c). The embedded NP appears in the ablative case in combination with the first two and in the dative case with the latter. Such adjectivizers allow NP recursion, as shown in Example 84d.

(84) a. *mohn=un* *bo:y* */ mohn-ɨn'* *beni*
Mohan=ADJZ.SG.M brother(M) / Mohan=ADJZ.SG.F sister(F)
'Mohan's brother / Mohan's sister' (Wali & Koul 1996:164)

b. *duka:n=uk* *mə:likh*
shop.ABL.SG=ADJZ.SG.M owner(M)
'the owner of the shop' (Wali & Koul 1996:165)

c. *ləḍk-an=hɨnd'* *ma:sṭar*
boy-DAT.PL=ADJZ.PL.M teacher(M).PL
'the teachers of boys' (Wali & Koul 1996:165)

d. *me:nis* *bəḍ-is* *bə:y=sɨnzi* *ko:ri=hɨnz* *za:m*
1SG.DAT elder-DAT.SG.M brother(M).DAT.SG=ADJZ.SG.F daughter(F.).DAT.SG=ADJZ.SG.F sister.of.husband(F)
[[[[NP-G] NP-A] NP-A] N
yiyi *paga:h.*
come.FUT tomorrow
]
'My older brother's daughters' sister-in-law will come tomorrow.' (Wali & Koul 1996:103)

Apart from these clitics, there are other adjectivizers, most of them of Persian origin. An example is *sarkə:r-i:* [government-ADJZ] 'governmental' (Wali & Koul 1996:273-275). We have found no evidence that adjectivized NPs can embed further NPs.

Kashmiri has postpositions, which assign the dative or ablative case to the noun they govern (Wali & Koul 1996:152). We have found no evidence for the adnominal use of postpositions, which suggests that such constructions do not exist in Kashmiri. Rather, it appears that postpositions are integrated into NPs with the adjectivizer:

(85) *me:z-ɨ* *p'aṭh=uk* *me:zposh* *chunɨ* *sa:ph.*
table-ABL.SG on=ADJZ.SG.M tablecloth(M) be.NEG clean
'The tablecloth on the table is not clean.' (Wali & Koul 1996:101)

There are compounds in Kashmiri, but the description by Wali & Koul (1996:284–287) suggests that determinative compounds are extremely rare. In addition, Wali & Koul (1996) do not provide any evidence for recursive compounding and we are not aware of any other kind of NP embedding by means of juxtaposition. We thus conclude that juxtaposition is not available for NP recursion.

11.3 Maithili

The trees in Chang et al. (2015) include Magahi as a representative of what is traditionally called the Bihari group. Here we sample Maithili instead, which is another representative of the group and which can be assumed to have diverged lexically from the other languages in Indo-Aryan languages at about the same time depth. We chose Maithili because it is better described and better accessible to us.

Maithili allows recursive embedding only with genitives. Endocentric juxtaposition is not productive, adjectivizers are limited to a few lexical formations without recursion, adpositions cannot be used adnominally, and there is no head-marking construction.

While the inherited Indo-Aryan case system is lost (except for a few traces that show up as case-induced stem allomorphies), Maithili has developed a new genitive *-k* (Grierson 1909, Yadav 1996). This marker derives from the same etymon as the Hindi adjectivizer *=k-*, i.e. a participle of the root *kr̥-* 'do', but has lost all properties of agreement and case government. An example with recursion is the following:

(86) *Rām-ak kāki-k ghar*
Ram-GEN aunt-GEN house
[[[NP-G] NP-G] N]
'Ram's aunt's house' (Yogendra Yādava, p.c.)

There are no other adnominal cases. The locative *-me*, for example, cannot be used adnominally, and locative adpositions are relational nouns. An expression like 'take the mango in the basket under the tree (not this one)' cannot be translated by an adnominal locative (Example 87a). Instead, a participle construction is required (Example 87b).

(87) a. * *gāch tar chiti-me ām lia*
tree under basket-LOC mango take.IMP

b. *gāch tar-ak chiti-me rākh-al ām lia*
tree under-GEN basket-LOC keep-PTCP mango take.IMP
'take the mango in the basket under the tree!' (Yogendra Yādava, p.c.)

NP embedding by means of juxtaposition is avoided, apart from compounds borrowed from Sanskrit. Expressions like 'chicken meat curry' require a genitive:

(88) *murgā maus-ak jhor*
chicken meat-GEN curry
'chicken meat curry' (Yogendra Yādava, p.c.)

11.4 Marathi

In Marathi, only adjectivizers allow NP recursion. Juxtaposition is possible, but is limited to two members. There are no adnominal adpositions, no genitive, and no head-marking construction.

Marathi continues the ancient Indo-Aryan case system in form of a binary contrast of direct vs. oblique. The direct form serves as a nominative, the oblique form occurs in all other contexts.

The oblique combines with a number of different postpositions that are cliticized to the oblique-marked noun and specify its grammatical role (Pandharipande 1997:273-275). The bare oblique form serves a vocative.

Marathi cases and adpositions are not attested in adnominal position. In any case, such constructions are not described by Pandharipande (1997:149-158). Rather, it appears that case- and adposition-marked nouns can only be embedded into an NP by means of the adjectivizer:

(89)	*sudhā-tS-yā*	*gharā-tS-yā*	*samor-ts-a*	[...]	*dzhāḍ*
	Sudha-ADJZ-OBL.SG	house-ADJZ-OBL.SG	in.front-ADJZ-NOM.SG.NT	[...]	tree(NT).NOM.SG
	[[[[NP-A]	NP-A]	NP-A]		N]

‘The [...] tree in front of Sudha’s house’ (Pandharipande 1997:156)

The adjectivizer construction here builds on the clitic *-tS-* ~*-ts-*, which probably goes back to a participial form of the verb stem *kr̥-* ‘do’ (Masica 1991:243). This adjectivizer attaches to the oblique form of an NP and assigns adjective properties to it, including agreement with the head in terms of number, gender, and case. The construction allows NP recursion, as shown by the example in Example 89.

Apart from this adjectivizer, there are several others, such as *-i* (e.g. *sarkār-i* [government-ADJZ] ‘pertaining to the government’), *-dār* (e.g. *tsaw-dār* [taste-ADJZ] ‘tasty’), or *-īya* (e.g. *bhārat-īya* [India-ADJZ]. These adjectivizers do not allow NP recursion, as far as we can tell.

Marathi also uses juxtaposition as a means of NP building (Pandharipande 1997:518-524). The language is rich in determinative compounds, e.g. *warṣā-rutu* [rain-season] ‘rainy season’ or *rāj-wāḍā* [king-palace] ‘royal palace’. However, we have not encountered any instances of recursive determinative compounds. Pandharipande (1997:520) describes the three-member compound *tan-man-dhan* [body-mind-money] ‘devotion’. However, the relationship between the single compound members is not hierarchical. Moreover, the author notes that such complex compounds are ‘not very common’. We thus conclude that juxtaposition is not a strategy that Marathi uses for NP recursion.

11.5 Nepali

In Nepali, recursive NP embedding is only possible by means of genitive case marking. There are no other case markers or adpositions that could be used adnominally, no adjectivizers that allow recursion and no endocentric juxtaposition, except for compounds borrowed from Sanskrit. There is also no head-marking construction.

Nepali has lost the old Indo-Aryan case system and replaced it by phrasal suffixes derived from various sources. Dative *-lāi*, for example, derives from a conjunct participle (converb) *lā-ī* ‘having taken’ (Schikowski 2013:213). A new genitive *-ko* has developed from the same participial base of the root *kr̥-* ‘do’ as the Hindi adjectivizer *=k-* (Masica 1991:243). For conservative speakers and in the written language, *-ko* still agrees in gender with the head noun and so it could be analyzed as an adjectivizer. In colloquial Nepali, however, agreement is completely lost and *-ko* is now a simple genitive:

(90) Nepali
us-ko *sāthi-ko* *didi-ko* *ghar* *(usko sathiko didiko ghʌr)*
3SG-GEN friend-GEN elder.sister-GEN house
[[[[NP-G] NP-G] NP-G] N]
'his friend's elder sister's house' (fieldnotes, B. Bickel)

Conservative speakers would have the feminine form *sāthi-kī* here.

Pronouns retain traces of an old direct vs. oblique system, but oblique stems can only occur in combination with the innovative case suffixes and do never occur on their own (e.g. *u* 'she' vs. *us-ko* '3SG-GEN'; or *yo* 'this' vs. *yas-mā* 'DEM-LOC'). We analyze these traces as stem allomorphy, conditioned by the case suffixes.

Determinative noun compounds are restricted to Sanskritized jargon in newspapers, scientific discourse etc. (mostly *tadbhavas*). In these jargons, just like in Sanskrit, they can be recursively expanded:

(91) Nepali
a. *viśva-* *vidhyālaya-* *sevā-* *āyog* *(bisso-biddyalae-sewa-ayog)*
world school service commission
[[[[NP-Ø] NP-Ø] NP-Ø] N]
'University Service Commission' (`http://www.tuservicecommission.edu.np`)

b. *sthā-* *nāma-* *koś* *(istha-nama-kos)*
place name dictionary
[[[NP-Ø] NP-Ø] N]
'Dictionary of place names' (title of a book published by the Nepal Academy)

The equivalents of determinative compounds in modern Nepali invariably require a genitive, e.g. 'chicken curry' translates as *kukhurā-ko māsu-ko tarkāri*, not **kukhurā-māsu-tarkāri* (by contrast, for example, to Sinhala *kukul-mas-kariə*, cf. Section 11.8). There is no other type of juxtaposition and so we conclude that the strategy is not available for NP recursion in Nepali.

Adpositions cannot be used adnominally. To translate an expression like '(take) [the book [on the table]]' one would either use a genitive without specifying the spatial relationship *ṭebul-ko kitāb* or use a relator noun such as *māthi* 'surface, top side' (*ṭebul māthi-ko kitāb*) (Narayan Gautam, p.c.).

11.6 Oriya

Oriya has genitives, adjectivizers, juxtaposition and adpositions, but there is no head marking. Genitives, adpositions and juxtaposition are available for recursive embedding, while adjectivizers are not.

The Oriya case system retains a direct vs. oblique distinction in personal pronouns (e.g. *mu* vs. *mo*). In all other nominal word classes (including demonstratives serving as substitutes for third person pronouns), the distinction is only marked in honorific contexts by a newly innovated oblique case -*nkɔ*. The oblique case serves as a genitive and is also governed by adpositions.[8] Genitives allow the recursive embedding of NPs:

(92) *mɔ mamu-nkɔ sikhyɔkɔ*
1SG.OBL uncle-OBL teacher
[[[NP-G] NP-G] N]
'my uncle's teacher' (Sadananda Das, p.c.)

Postpositions come from two sources: from relator nouns (Bubenik 1998) and, in the case of the postposition -*rɔ*, from an adjectivizer going back to a participle of the root *kr̥*- 'do' (Masica 1991:243). They all govern the oblique case (Neukom & Patnaik 2003:47,313). The postposition -*rɔ* can be used adnominally (as well as for marking experiencers and a few other functions) and allows NP recursion:

(93) *mo-rɔ sangɔ-rɔ bapa-nkɔ-rɔ ghɔrɔ*
1SG.OBL-of friend-of father-OBL-of house
[[[[NP-P] NP-P] NP-P] N]
'the house of the father of my friend' (Manideepa Patnaik, p.c.)

The postposition -*rɔ* has given up all adjectival properties in modern Oriya.

The other postpositions cannot be used adnominally (Neukom & Patnaik 2003:126):

(94) a. * *gãã-ru jhiɔ-ṭi*
village-from girl-ART
'the girl from the village'

b. * *ṭebul upɔr-e bɔhi*
table top-on book
'the book on the table'

What is used instead is the semantically neutral postposition -*rɔ* or relator noun constructions (Neukom & Patnaik 2003:126):

(95) a. *gãã-rɔ jhiɔ-ṭi*
village-of girl-ART
'the girl from the village'

b. *ṭebul upɔrɔ-rɔ bɔhi*
table top-of book
'the book on the table'

There are also adjectivizing suffixes that integrate nominals into NPs. One such adjectivizers is -*ia*, e.g. *sɔrɔkar-ia* [government-ADJZ] 'governmental' (Neukom & Patnaik 2003:68). We have

[8] Neukom & Patnaik (2003) treat examples like 92 as the result of case dropping because they analyze the oblique not as a case but as a case base. We are not aware of any arguments for this analysis, but note that it ma-kes the morphology more complex and misses the generalization that the oblique is both triggered by what they call 'case' and by what they call 'adpositions'.

not found any evidence, however, that adjectivized NPs can embed additional NPs and thus conclude that this is not possible in Oriya.

Oriya also displays determinative compounds, e.g. *phulɔ bɔgica* [flower garden] 'flower garden' (Neukom & Patnaik 2003:19). Juxtaposition of this kind allows the recursive embedding of further NPs, as the following example demonstrates:

(96) *kukuḍa mansɔ tɔrkari*
chicken meat curry
[[[NP-Ø] NP-Ø] N]
'chicken curry' (Manideepa Patnaik, p.c.)

11.7 Pāli

The term Pāli refers to the language which was used in the Buddhist traditions of Sri Lanka and Southeast Asia. The dating of the formation of the Pāli canon is controversial, but according to one tradition some canonical texts were brought to Sri Lanka in the 3rd century BCE (Hinüber 2001:62).

Pāli uses genitives, adjectivizers and juxtaposition for NP embedding. Genitives and juxtaposition allow recursive embedding. Adpositions exist, but they are not used adnominally. Head marking is absent altogether.

Pāli nouns are inflected for case. The genitive is marked by a suffix, which varies depending on the declension class and number of the noun. In some declension classes, there is only one form covering the instrumental, ablative and genitive (Oberlies 2001:140–149). Genitives allow NP recursion:

(97) *migāra- mātu pāsāde*
Migara- mother.GEN.SG palace.LOC
[[[NP-Ø-] NP-G] N]
'in the palace of Migara's mother' (Aṅguttaranikāya 2.35)

There are postpositions, which can govern any case except the nominative and vocative. They are rare in general and mostly appear as first members of compounds. Fahs (1989:102) observes that what is expressed by other languages by adpositions, is covered in Pāli by relator nouns, absolutives, adjectives and participles etc. Furthermore, Fahs (1989:104–105) explicitly mentions that relator nouns are being used instead of case forms, but does not mention adpositions in this context. In addition, we are not aware of any example of an adposition that is used adnominally. Taking all these observations together, we conclude that adnominal adposition constructions were unavailable in Pāli.

There are several suffixes which derive adjectives from nouns (and in most cases also nouns from nouns). Examples are *-ya* in *gamma*[9] [village.ADJZ] 'vulgar' and *-ima* in *majjh-ima* [middle-ADJZ] 'medium, middling' (Warder 2005 [1991]:254). Adjectives agree with their head noun in gender, number and case (Duroiselle 1997:160). There is no evidence whatsoever that derived adjectives can embed other NPs.

[9] The suffix *-ya* undergoes assimilation to the previous consonant.

Juxtapositions (compounds) are very frequent and become more complex over time: in the older language, they mostly consist of up to three stems, but in the later language more extensive ones can be found (Duroiselle 1997:129). Endocentric juxtapositions are frequent as well. Examples include *nadi-tīraṁ* [river-bank] 'river bank' and *rāja-putto* [king-son] 'prince' (Duroiselle 1997:132). As with other types of juxtapositions, they frequently have three or more members, which modify each other (Warder 2005 [1991]:78):

(98) *kūṭā- gāra- sālā*
gable- house- hall
[[[NP-Ø-]NP-Ø-]N]
'hall of the house with a gable' (Warder 2005 [1991]:78)

Like in Vedic Sanskrit and Avestan (cf. 11.9 and 12.1), the non-head can also be modified by an external genitive:

(99) *brāhmaṇa-ssa pada- saddena*
brahmin-GEN.SG foot- sound-INS.SG
[[[NP-G] NP-Ø-] N]
'by the sound of the footsteps of the Brahmin' (Wijesekera 1936:229)

11.8 Sinhala

Sinhala has two constructions that allow NP recursion: genitives and juxtaposition. Adpositions exist, but they cannot be used adnominally. Adjectivizers that allow NP recursion appear *in statu nascendi*, but are not yet fully grammaticalized. Head marking is absent altogether.

Sinhala nouns are inflected for case, which is marked by suffixes. There is one case which covers genitive and locative functions. It has different forms for animate and inanimate nouns (Chandralal 2010:79,81). On animates, it is indicated by the suffix *-ge*. On inanimates, however, the genitive is marked by the suffix *-ee* in the singular Example 100a and *-wələ/-walwələ* in the plural. These genitives can be used for NP recursion, as in Example 100b.

(100) a. *winaasy-ee mulə*
destruction-GEN.SG beginning.NOM.SG
'the beginning of destruction' (Chandralal 2010:118)

b. *apee ayya-ge noona-ge malli-ge duwə*
1PL.GEN elder.brother-GEN.SG wife-GEN.SG younger.brother-GEN.SG daughter.NOM.SG
[[[[[NP-G] NP-G] NP-G] NP-G] N]
'our elder brother's wife's younger brother's daughter' (Chandralal 2010:10)

There are no adnominal adpositions (in our sense) in Sinhala. Instead of adpositions, Sinhala uses relator nouns, which are inflected for case:

(101) *wattə-ṭə pahalə / wattə-ṭə pahal-in*
estate-DAT.SG below.NOM.SG / estate-DAT.SG below-INS.SG
'below the estate / from below the estate' (Gair 1970:33)

When such relator nouns modify other nouns, they are embedded by a construction involving a participle. With inanimates, *æti* is used, which is most probably a past participle derived from the existential copula *æt-*. With animates, the present participle of the verb 'to be, to stand', *sitinə*, is used. These are incipient adjectivizer constructions that seem to recapitulate a development similar to that of of adjectivizers from *kr̥-* 'do' in other Indo-Aryan languages several centuries earlier:

(102) a. *kaamər-ee æti meese uḍə æti mall-ee æti potə*
room-GEN.SG be.PTCP table on be.PTCP bag-GEN.SG be.PTCP book.NOM.SG
'the book in the bag on the table in the room[10]' (Tilakaratne 1992:164)

b. *siilimə-hi sitinə makuluwə*
ceiling.NOM.SG-on stand.PTCP spider.NOM.SG
'the spider on the ceiling' (Tilakaratne 1992:163)

The construction does not appear to have been fully grammaticalized, although further fieldwork would be needed to firmly establish this. There are also adjectivizing suffixes, e.g. -*muwaa* (*dæwə-muwaa* [wood-ADJZ] 'wooden') or (*-ikə aarth-ikə* [economy-ADJZ] 'economical') (Chandralal 2010:84–85), but we are not aware that they allow the recursive embedding of NPs.

Juxtaposition is a very productive process in Sinhala, which also includes endocentric constructions such as *baicikal roode* [bicycle wheel] 'bicycle wheel'. Although instances seem to have two members, it is also possible to have recursive stacking of more members:

(103) *kukul- mas- kariə*
chicken- meat- curry
[[[NP-Ø-] NP-Ø-] N]
'chicken curry' (Chandralal 2010:89)

11.9 Vedic Sanskrit

Vedic Sanskrit refers to the language of the Vedas, ritual texts that were compiled around the mid-second century to the mid-first century BCE. It is the oldest well-attested language of the Indo-Aryan subbranch.

Vedic Sanskrit has genitives, adjectivizers, adpositions and juxtapositions. Adjectivizers do not allow the recursive embedding of NPs, but the other two constructions do. Head marking is absent altogether.

The genitive is expressed by a suffix, which varies according to the declension class and number of the noun it attaches to. Genitives are unconstrained with regard to recursion. In Example 104a, the genitive *amítrāṇām* 'of the enemies', which modifies the head noun, is in turn modified by the genitive of the 3rd sg. pronoun *asya* 'his'. In Example 104b, the genitive *viśā̃m* 'of the clans' is modified by the denominal adjective *daívīnām* 'divine'. Note that the adjective agrees with its head noun in case, gender and number, like when it is the only modifier (cf. Example 105 below).

[10] Note that we adjusted the orthography and glosses used by Tilakaratne (1992) to the orthography and glosses used by Chandralal (2010).

(104) a. *amítr-āṇām a-sya sénā-ṃ*
enemy-GEN.PL 3SG-GEN.SG troop-ACC.SG
[[NP-G [NP-G]] N]
'the troop of his enemies' (Taittirīya Saṃihtā 3.4.8.4)

b. *viś-ā́ṃ daív-ī-nām utá pūrva-yā́-vā*
clan(F)-GEN.PL god-ADJZ.F-GEN.PL and before-go-NOM.SG
[[NP-G [NP-A]] N]
'and the leader of the clans of the gods' (Rig Veda 3.34.2d)
(cf. *dev-ā́nāṃ víś-as* [god-GEN.PL clan-GEN.PL] 'clans of the gods')

There are a number of prepositions in Vedic Sanskrit, which govern case. They cannot be used adnominally; instead, they always modify predicates (Macdonell 1916:208-210,285).

There are several suffixes used as adjectivizers, the most prominent one being *-(i)ya-*. With some of these suffixes, the stem undergoes vowel gradation, traditionally referred to as *Vṛddhi*. Such adjectives agree with their head noun in case, number and gender:

(105) *nákir-mā dáiv-y-aṃ sáh-o vara-te*
nobody.NOM.SG-1SG.ACC god-ADJZ-NA.SG.NT power(NT)-NA.SG hinder-3SG.PRS.MD
'no divine power holds me back' (Rig Veda 4.42.6ab)

We have not come across any examples in which a denominal adjective is modified by an other NP. We thus conclude that they cannot embed NPs.

Juxtapositions in the form of compounds are abundant in Vedic Sanskrit, but those of interest, the determinatives, are the rarest type. An example of a determinative compound is *rāja-putrá-* [king-son-] 'son of a king' (Wackernagel 1905:243). Determinative compounds can be recursively expanded, but only with genitive-marked NPs:

(106) a. *árvato māṃsa- bhikṣā́m*
horse.GEN.SG meat- request.ACC.SG
[[[NP-G] NP-Ø] N]
'the request for the meat of the steed (s.c. the aforementioned horse which is being cooked during the horse sacrifice)' (Rig Veda 1.162.12c) (Wackernagel 1905:31)

b. *vīra-hā́ vā́ eṣá devā́nāṃ*
man-slayer(M).NOM.SG PTCL DEM.NOM.SG.M god.GEN.PL
[*hā́* (N) [*vīra-* (NP-Ø) [*devā́nāṃ* (NP-G)]]]
'now he is the slayer of the men of the gods' (Taittirīya Saṃihtā 1.5.2.5)

In the period that followed Vedic Sanskrit, this constraint is resolved and determinative compounds can be expanded directly inside the compound. The determinative compound in Example 107 has four members. The first two are coordinated (*veda* and *vedāṅga*) and together, they are embedded in the phrase headed by *-jñaḥ* 'scholar'.

(107) Epic Sanskrit

veda- *veda-* *aṅga-* *tattva-* *jñaḥ*
veda- veda- limb- truth- scholar
[[[[NP-Ø] [[NP-Ø] NP-Ø]] NP-Ø] N]
'a scholar in the essence of the Vedas and the limbs of the Vedas, too'[11] (Rāmāyaṇa 1.1.13c)

11.10 Vlax Romani

Vlax Romani refers to the dialects of Romani spoken in Southeastern Europe. There is no genitive and no head-marking construction. Juxtaposition exists, but seems to be very marginal. Adjectivizers and adpositions are productive and allow the recursive embedding of NPs.

Nouns are inflected for case, though the system is heavily reduced. There is a direct/nominative case and an oblique/accusative case. These are mainly used for subjects and direct objects respectively. All the oblique relations are expressed by adding a suffix to the oblique/accusative (Igla 1996:23).

One of the main constructions for NP recursion builds on an adjectivizer. The construction involves a clitic *-k-/-g-* which governs oblique NPs and agrees with the head noun in number and (in the singular) also in gender, just like any other adjective (Hancock 1995:72):

(108) a. *e* *čha-əs-k-o* *dad*
ART.DEF boy-OBL-ADJZ-SG.M father(M)
'the boy's father' (Igla 1996:24)

b. *e* *čha-əs-k-i* *dej*
ART.DEF boy-OBL-ADJZ-SG.F mother(F)
'the boy's mother (Igla 1996:24)

c. *e* *čha-əs-k-e* *phralá*
ART.DEF boy-OBL-ADJZ-PL brother.PL
'the boy's brothers' (Igla 1996:24)

Adjectivizer constructions allow NP recursion. The adjectivizer always agrees with the immediate head:

(109) *murr-e* *dad-əs-k-o* *amal-əs-k-o* *kher*
1SG.POSS-SG.OBL.M father(M)-OBL-ADJZ-SG.M friend(M)-OBL-ADJZ-SG.M house(M)
[[[[NP-A] NP-A] NP-A] N]
'the house of my father's friend' (Anthony Grant, p.c.)

There are a number of prepositions, which may govern the locative case[12] unless there is a definite article, in which case the direct/nominative case is used, e.g. *ande o veš* [in ART.DEF forest] 'in the forest' vs. *pàša man-de* [near 1SG.OBL-LOC] 'near me' (Hancock 1995:73). Prepositions can be used adnominally (Example 110a), and in this function can also embed further NPs (Example 110b):

[11] The *vedāṅga* (lit. 'limb of the veda') is a term for a certain class of predominantly scholastic and exegetic works regarded as auxiliary to the Veda.

[12] It is often not used with ordinary nouns, but always with pronouns.

(110) a. *khoj katar e bakrja*
tallow of ART.DEF sheep.PL
'tallow of the sheep' (Boretzky 1994:108)

b. *e klishka pe e sinija ande e soba*
ART.DEF book on ART.DEF table in ART.DEF room
[N [NP-P [NP-P]]]
'the book on the table in the room' (Ian Hancock, p.c.)

Compounds play a marginal role in Romani and are not productive to the most part. Noun-noun compounds are especially rare and there are only a few lexicalized cases (Matras 2002). There is no other juxtaposition construction, and we conclude that this type is not available for NP recursion in Vlax Romani.

12 Iranian

12.1 Avestan

Avestan is the language of the *Avesta*, the sacred book of the Zoroastrians. This collection of mostly ritual texts is generally assumed to have been composed in the late second and early first millennia BCE. Old Avestan represents an earlier form of the language than Young Avestan, but Old Avestan is unlikely to be the direct ancestor of Young Avestan. Our description includes both varieties (as indicated below).

In Avestan, we find all surveyed structures: genitives, adjectivizers, compounds, head marking, and adpositions. The first three allow NP recursion. Adjectivizers cannot be used for recursive NP embedding, and adpositions are not attested in adnominal function.

Avestan nouns are inflected for case. There are three genders and three numbers, and all eight cases that are reconstructed for PIE are preserved in Avestan, including the genitive (Hoffmann & Forssman 1996:114-115). Avestan allows recursive NP embedding with genitives:

(111) *Zaraθuštr-ō nmān-ahe Pourusasp-ahe*
Zaraθuštra-NOM.SG house-GEN.SG Pourusasp-GEN.SG
[N [NP-G [NP-G]]]
'Zarathustra of the house of Pourusaspa' (Young Avestan, Yasna 9.13)

There are various suffixes for deriving adjectives from nouns, for example *-a-* (*maniiauu-a-* 'of the spiritual world'), *-i-* (*āhuir-i-* 'of Ahura (Mazda)'), and *-ya-* (*tūir-iia-* 'from Tura') (Skjærvø 2009:167–168). Denominal adjectives derived in this way are very common in Avestan, but the grammars and descriptions do not mention any ability to embed further constituents, and we have not come across any examples in the corpus. This suggests that they are highly marginal, if they exist at all. But there is another adjectivizing construction which does allow NP recursion. This construction involves the use of the relative pronoun as a linker of adjectives or embedded NPs:

(112) a. *aoi yąm ast-uuait-ī-m gaēϑąm*
against REL.ACC.SG.F bone-ADJZ-F-ACC.SG world(F).ACC.SG
[NP-P [NP-A]]

‘against the material world’ (Young Avestan, Yasna 9.8)

b. *aspa-cit̰* *yōi* *miϑrō-* *drująm*
horse(M).NOM.PL.-even REL.NOM.PL.M contract- infringer.GEN.PL
[N [NP-A [[NP-Ø] NP-G]]]
‘even the horses of the infringers of contracts’ (Young Avestan, Yašt 10.20)

Avestan also allows the recursive extension of juxtapositions, not by adding a third member to the construction, but by adding an external genitive NP. This is illustrated in (Example 113), in which the embedded genitive *daēuuanąm* ‘of the old gods’ modifies the embedded noun *kamərəδa-* ‘head’. Note that the linear order does not correspond to the phrase structure, so we omit an interlinear display of the hierarchical structure:

(113) *kamərəδō-jan-ō* *daēuua-nąm*
head-smash-NOM.SG old.god-GEN.PL
[*janō* (N) [*kamərəδō-* (NP-Ø) [*daēuuanąm* (NP-G)]]]
‘smasher of the heads of the old gods’ (Young Avestan, Yasna 57.33)

In Young Avestan, the relative pronoun often takes the invariable from *yat̰*, which formally corresponds to the nominative/accusative neuter singular. The lack of agreement morphology suggests that the particle *yat̰* does not assign any adjectival properties anymore and so effectively functions as a head marking *ezāfe* that links adjectives (Example 114a) and NPs (Example 114b) to the head. However, we are not aware of any instances of recursive embedding with *yat̰*. This is attested only in younger members of the Iranian family.

(114) a. *ahmi* *aŋhuuō* *yat̰* *ast-uuaiṇt-i*
this.LOC.SG.M life(M).LOC.SG (EZ|REL.NA.SG.NT) bone-ADJZ-LOC.SG.M
‘in this material life’ (Young Avestan, Vidēvdād 5.39)

b. *puθr-əm* *yat̰* *pourušasp-ahe*
son(M)-ACC.SG (EZ|REL.NA.SG.NT) Pourušaspa(M)-GEN.SG
[N-H [NP-G]]
‘the son of Pourušaspa’ (Young Avestan, Yašt 5.18)

12.2 Persian, Middle

Middle Persian is the direct successor of Old Persian and was the language used by local rulers (from the late 3rd century BCE) in modern-day south-western Iran (*Fārs*) after the fall of the Old Persian empire. From 224 CE onwards it was the official language of the Sasanid empire until the Arab conquest in ca. 650 CE and continued to be used by Zoroastrians for centuries after the spoken language had further evolved into Early Modern Persian. The bulk of Middle Persian documents is known from inscriptions and mostly religious (Manichean, Zoroastrian, and Christian), historical, and legal literature written on parchment and papyri found in the Near East and along the Silk Road (Skjærvø 2009:196–197, Durkin-Meisterernst 2014:14–25).

In Middle Persian there is no genitive case, but the language has adjectivizers, adpositions, juxtapositions, and a head-marking construction, the *ezāfe*. The latter three allow the recursive embedding of NPs.

As for case, Middle Persian makes a two-way distinction between *rectus* and *obliquus.* This case distinction is clearly in decay, but there is reliable evidence that the system is likely to have survived until late Middle Persian (Skjærvø 2009:205). The two cases are still fairly well distinguished in the 1SG pronoun, as shown in Example 115. The *rectus* form *an* is employed for S/A arguments, e.g. in the non-past tense as illustrated in Example 115a. The oblique form *man* is used for P arguments in the non-past tense (Example 115b), and for both A and P arguments-in the past tense. When governed by an adposition, it is the oblique form *man* which is used (Example 115c). Therefore Middle Persian *pad*, and *andar*, are still to be classified as prepositions, which is also in accordance with their etymology.

(115) a. *an hēm ādur*
1SG.RCT am fire
'I am the fire' (Middle Persian from Turfan, Brunner 1977:55)

b. *man wēnēd*
1SG.OBL look.IMP.2PL
'Look at me!' (Middle Persian from Turfan, Brunner 1977:55)

c. *kū-tān hān … pad man wināst*
QUOT-2PL this against 1SG.OBL sin.PST
'you … sinned against me' (Middle Persian from Turfan, MacKenzie 1979:508)

The following example shows the embedding of an NP inside a preposition-marked NP (Gignoux & Tafazzoli 1993:36):

(116) *tār-kirb-ān pad čihr ud dēs ī Azdahāg*
darkness-body-PL in shape and form EZ Azdahāg
[N [NP-P N-H [NP]]]
'creatures of darkness in the shape and appearance of (the dragon) Azdahāg' (Gignoux & Tafazzoli 1993:36)

This suggests that the prepositional type of NPs is compatible with recursive NP structures.

The genitive case was lost in Late Old Persian and is absent from Middle Persian. Instead, juxtaposition is used. This constructions allows the recursive embedding of NPs (Skjærvø 2009:265):

(117) *šāh-ān šāh ērān (= šāh šāh-ān ērān)*
king-PL king Erān [N [NP-Ø [NP-Ø]]]
'king of kings of Erān (i.e. of Iranian peoples)' (Skjærvø 2009:265)

Adjectivizers, as in *āb-īg* [water-ADJZ] 'watery, aquatic', are quite productive (Skjærvø 2009:262–263, Durkin-Meisterernst 2014). However, the literature does not mention that adjectivizers allow the recursive embedding of NPs. This suggest that adjectivizers cannot be used to build recursive NP constructions.

Head marking occurs in a construction usually referred to as *ezāfe*. It consists of a linker *ī(g)* that marks the embedding relationship (Skjærvø 2009:263):[13]

[13] The linear oder does not entirely correspond to the hierarchical structure since the *ezāfe* modifies the element *kār* 'deed' which is embedded in the topmost head noun *nāmag* 'book' right after it.

(118) Middle Persian
kār- nāmag ī Ardašīr ī Pābagān
deed- book EZ Ardašīr EZ clan.of.Pābag
[NP-Ø N] H [N-H [N-H NP]]
'book of (the) deeds of Ardašīr (son) of Pābag' (Skjærvø 2009:263)

The dependent noun is in the oblique case although this is of course only visible where the forms are available. Historically, the Middle Persian *ezāfe* particle *ī(g)* has pronominal origins. It is historically related to the Old Persian relative pronoun *haya-* (Skjærvø 2009:100):

(119) Old Persian
kāra haya manā
army(F).NOM.SG REL.NOM.SG.F 1SG.GEN
'my army' (Darius Behistun 2.25, Skjærvø 2009:100)

12.3 Persian, Modern

Modern Persian uses genitives, adjectivizers, juxtaposition, and head marking for NP embedding. Genitives and head marking allow NP recursion.

Modern Persian has lost all inherited case marking and has reanalyzed adpositions as new phrasal case markers (according to our taxonomy, where adpositions are required to show case government or stranding behavior). The preposition *az* 'from' comes close to a semantically neutral genitive, although it is not as neutral as the parallel developments we find in, for example, German *von* 'from' or French *de* 'from', and still conveys an ablative notion. At any rate, NPs embedded by means of *az* allow recursive expansion (Example 120):

(120) *yek kolāh az pashm-e shotor*
one hat from wool-EZ camel
[N [NP-G N-H [NP]]]
'a hat made of camel wool' (Sascha Völlmin, p.c.)

There are several adjectivizers that derive adjectives from nouns, e.g. the suffix *-i/-ī* as in *Irān-i* [Iran-ADJZ] 'Iranian' (Windfuhr & Perry 2009:527). However, there is no evidence that adjectivized NPs embed additional NPs.

Modern Persian also occasionally uses juxtaposition as a strategy to build complex NP structures, e.g. *hafte-nāme* [week-document] 'weekly periodical' (Windfuhr & Perry 2009:528), but instances of recursive applications of this are extremely rare (Sascha Völlmin, p.c.). We have only come across one juxtaposition that consists of more than two nominal elements, viz. *shotor-gāv-palang* [camel-cow-leopard] 'giraffe'. However, the express in question is not recursive. Accordingly, there is no evidence that Modern Persian exploits juxtaposition as a productive means for building recursive NP structures.

The construction that is most commonly used to build complex NP structures is a head-marking construction. In this construction, the head is marked by what is known as the *ezāfe* particle. The embedded constituent can be headed by any word class, e.g. a noun, an adjective, etc. The *ezāfe* construction can embed additional NP constituents, as the following example illustrates:

(121) *ketāb-e pedar-e Hasan*
book-EZ father-EZ Hasan
[N-H [N-H [NP]]]
'the book of Hasan's father' (Lazard 1992:67)

12.4 Northern Pashto

In Pashto there is only one construction that allows the recursive embedding of NPs, namely adpositions. Adjectivizers and compounds exist as well, but they cannot embed further constituents. There is no genitive and no head-marking construction.

In Pashto, there are pre- and postpositions and they can combine into circumpositions. Such circumpositions are used more frequently than either pre- or postpositions (Babrakzai 1999:41). Prepositions usually govern the oblique case (Tegey & Robson 1996:158). While this is not always evident with nouns, it can be seen with pronouns: the preposition *də*, which regularly reduces to a sibilant before pronouns, e.g. *z-ma* 'of me', requires the first person singular oblique pronoun *ma* (rather than the direct form *zə*) (Babrakzai 1999:30-31). Adpositions can embed further NP constituents:

(122) *də kitab χane də mudir də wror kor.*
of book house of manager of brother house
[[[[NP-P [[NP-Ø] N]] NP-P] NP-P] N]
'The house of the brother of the manager of the library.' (Babrakzai 1999:31)

There are a number of adjectivizing suffixes that derive adjectives from nouns, e.g. *-áy* (*jāpān-áy* [Japan-ADJZ] 'Japanese') (David 2014:150-152). It seems that denominal adjectives cannot embed other NPs. In any case, we have not encountered any examples in the available material.

Pashto also has compounds, e.g. *lmar-xātə* [sun-rising] 'east'. However, many of them have been borrowed from other languages, especially Persian (David 2014:101). We have not come across any nominal compounds that consists of more than two members. This suggests that juxtaposition is not among the strategies that allow recursive embedding of NPs in Pashto.

12.5 Ossetic

In Ossetic, we find genitives, adpositions, juxtaposition and head marking. Except for juxtaposition, all these constructions are available for NP recursion.

Ossetic nouns are inflected for case, one of which is the genitive. It is marked by the suffix *-i* (Digor dialect) or *-ɨ* (Iron dialect), as in Example 123, and derives from an ancient genitive ending *-ah* (Thordarson 2009:132). For most nominals it also marks accusative and inessive, i.e. it is syncretic and therefore commonly glossed as oblique (David Erschler, p.c.). The genitive can embed additional NP constituents:

(123) *soslan-i bɐχ-i k'ɐʤelɐ*
Soslan-OBL horse-OBL tail
[[[NP-G] NP-G] N]
'Soslan's horse's tail' (Digor dialect; David Erschler, p.c.)

There is one preposition and several postpositions in Ossetic, which predominantly express location and govern various cases. Adposition-marked NPs can embed further NPs, no matter whether the NP is embedded by a preposition (Example 124a) or a postposition (Example 124b):

(124) a. *fɐndɐg mɐ- fid-i χɐdzar-i wɐngɐ*
road 1SG.OBL- father-OBL house-OBL until
[N [[[NP-G] NP-G] NP-P]]
'the/a road up to my father's house'

b. *seχʷar ɐnɐ mɐ- mad-i fidgun(-ɐj)*
breakfast without 1SG.OBL- mother-OBL meat.pie(-ABL)
[N [NP-P [[NP-G] NP-G]]]
'the/a breakfast without my mother's meat pie' (Digor dialect; David Erschler, p.c.)

Adjectives and nouns cannot be clearly distinguished based on morphology (Erschler 2016:3158). With regard to their function however, some nominal stems occur almost exclusively either as head or modifier (Thordarson 2009:86). Also, there are various derivational suffixes which form either nouns as in Iron *xorz-ʒinad* 'good-ness', or adjectives *pedagog-on* 'pedagogical' (cf. Thordarson 2009:86-87). We have found no evidence that adjectivized noun can embed additional NP constituents and thus conclude that this is not possible in Ossetic.

Ossetic makes frequent use of compounds, such as *aχsɐn-nez* [stomach-illness] 'stomach illness' (Erschler 2016:3162-3163). Such collocations are quite common, but not with three or more members, suggesting that juxtaposition does not allow recursion.

Ossetic has developed a head-marking construction involving a possessive pronominal clitic on the head. The embedded NP is assigned dative case (similar to the Swiss German construction described in Section 9.12). This construction allows NP recursion:

(125) *mɐ- fid-i limɐn-ɐn ɐ- kizgɐ*
1SG.GEN- father-OBL friend-DAT 3SG.GEN- daughter
[[[[NP-G] NP-G] NP] N-H]
'my father's friend's daughter' (Digor dialect; David Erschler, p.c.)

In addition to this, there is also the trace of an other, presumably older, head-marking construction in the form of a suffix *-i* (Digor dialect) or *-ɨ* (Iron dialect). The form happens to be formally identical to the genitive case resulting in the apparent synchronic puzzle of a genitive marking a head:

(126) *mæ fid-ɨ zærond*
1SG.GEN father-LNK old
'my old father' (Iron dialect; Thordarson 2009:109)

This construction displays an atypical syntactic structure for Ossetic NPs, as the modifier occurs in postnominal rather than prenominal position. In addition, the construction is semantically specialized for specific kinds of physical and mental properties and states such as 'old', 'good', 'stupid', etc. (Thordarson 2009:109). The *i*-construction is thus limited to a few fixed expressions only (Belyayev 2010:298) and, unlike the head marking type based on possessive forms (Example 125), does not support recursive NP embedding (David Erschler, p.c.).

12.6 Southern Baluchi

There are three main varieties of Baluchi: Southern Baluchi (spoken in South Pakistan and Southeast Iran), Eastern Baluchi (spoken in Pakistan) and Western Baluchi (spoken in Pakistan, Iran, Afghanistan, and Turkmenistan). The varieties are quite different in terms of grammatical structure (Jahani & Korn 2009), but their divergence from the most common recent ancestor (i.e. theoretical branch lengths) can be assumed to be sufficiently similar for the purpose of phylogenetic modeling. We base our analysis on Southern Baluchi because we have much better access to data from this variety than from the others.

Southern Baluchi has genitives, adpositions, juxtaposition and adjectivizers. Only genitives and adpositions can embed other constituents. There is a head-marking construction, but it is not productive.

In Southern Baluchi, nouns are inflected for case, number, and indefiniteness. The language has a genitive case, which is marked by the suffix *-ay*, *-ē* or *-ī* in the singular and by the suffix *-ānī* in the plural (Jahani & Korn 2009). Genitive-marked nouns can embed additional NPs:

(127) *manī* *dost-ē* *pit-ē* *lōg*
1SG.GEN friend-GEN.SG father-GEN.SG house.NOM
[[[[NP-G] NP-G] NP-G] N]
'my friend's father's house' (Riaz Ahmed, p.c.)

Southern Baluchi has prepositions, postpositions, and circumpositions. Prepositions usually govern the oblique, while postpositions and circumpositions govern the genitive case (Jahani & Korn 2009). Most postpositions have developed from relator nouns in the oblique case and in many cases they are still best analyzed as such rather than as adpositions. Prepositions, however, constitute a lexical class in their own right. Preposition-marked nouns allow recursive NP structures:

(128) *yak wahag-ē pa(r) taī dōst-ē pit-ā*
one gift-INDF for 2SG.GEN friend-GEN.SG father-OBL.SG
[N [NP-P [[NP-G] NP-G]]]
'a gift for your friend's father' (Riaz Ahmed, p.c.)

There is an adjectivizer *-ī*, which derives adjectives from nouns, e.g. *iran-ī* [Iran-ADJZ] 'Iranian'. The resulting adjectives cannot embed other NPs, however (Riaz Ahmed, p.c.).

The language also has nominal compounds, e.g. *log-banok* [house-lady] 'wife' and *insān-dōst* [human.begin-friend] 'philanthropist'. Recursive compounding, however, is not productive in Southern Baluchi (Riaz Ahmed, p.c.), and we thus conclude that juxtaposition is not a strategy for NP recursion.

Southern Baluchi also has a head marking *ezāfe* construction, but this strategy has been borrowed only recently from Persian and is confined to fixed phrases (Jahani & Korn 2009:656), such as *nam-ē huda* [name-EZ God] 'the name of God' (cf. the native genitive construction *huda-ē nam-ā* [God-GEN name-OBL.SG] 'the name of God'). Thus, the *ezāfe* cannot be considered a productive means for the building of recursive NP structures in Southern Baluchi.

13 Italic

13.1 French

Of the five constructions surveyed, French has three: genitives, adjectivizers and juxtaposition, of which only the first can embed additional NPs. There are no adpositions and no head marking.

French possesses a number of NP-building constituents that are commonly referred to as prepositions in reference grammars. In our coding scheme, these constituents count as genitives rather than prepositions because they do not govern case in contemporary French. To be sure, French 'prepositions' require a special form of pronouns, the so-called disjunctive pronouns (Batchelor & Chebli-Saadi 2011:446). However, disjunctive pronouns cannot be analyzed as case-marked pronominal forms because they can refer to participants in both subject and object function (cf. Example 129a and Example 129b). French genitives can embed further constituents as Example 129c illustrates.

(129) a. *Lui, il ne peut pas ven-ir.*
3SG.M.DISJ 3SG NEG can.PRS.3SG NEG come-INF
'He cannot come.'

b. *Je ne le vois pas, lui.*
1SG NEG 3SG.M.OBL see.PRS.1SG NEG 3SG.M.DISJ
'I do not see him.'

c. *le chien de la voisin-e de Martha*
ART.SG.M dog(M) GEN ART.SG.F neighbor-F.SG GEN Martha(F)
[N [NP-G [NP-G]]]
'the dog of Martha's neighbor'

Adjectivizers exist, e.g. *familial* [family.ADJZ] 'domestic' or *côtier* [coast.ADJZ] 'coastal', but cannot recursively embed additional NPs.

Compounding is a productive process in French, though noun-noun compounds are not the most frequent type. Often, the first element is the head and the second the modifier, as in *timbre-poste* [stamp-mail] 'postage stamp'. These compounds are limited to two members. Thus, juxtaposition is not available for recursive NP embedding in French.

13.2 Italian

Italian uses adpositions, adjectivizers, and juxtaposition for NP embedding. Of these three strategies, only adpositions and juxtaposition can be used for recursive embedding. There is no genitive and no head marking.

Italian has a wide range of prepositions, which can occur in adnominal position. Prepositions govern an oblique case, which is however only visible in pronouns (Maiden & Robustelli 2007:171), as in Example 130a. Adposition-marked nominals can embed an additional NP, as Example 130b demonstrates.

(130) a. *un regalo per te*
ART.SG.M gift(M).SG for 2SG.OBL
'a gift for you' *[ANONYMIZED]*

b. *il* *libro* *sul* *tavolo* *nel* *soggiorno*
ART.SG.M book(M).SG on.M.SG table(M).SG in.M.SG living.room(M).SG
[N [NP-P [NP-P]]]
'the book on the table in the living room' (Francesco Gardani, p.c.)

There are a number of adjectivizing suffixes, which derive adjectives from nouns, e.g. *-ale* (*nazionale* [nation.ADJZ] 'national') or e.g. *-ano* (*italiano* [Italia.ADJZ] 'Italian') (Maiden & Robustelli 2007:443-444). Such denominal adjectives cannot embed additional NPs, however.

Italian also uses plain juxtaposition as a means of NP building (Bisetto 2010). Such NPs are typically left-headed, e.g. *capo-stazione* [master-station] 'station master' or *transmissione radio* [transmission-radio] 'radio transmission'. These NPs can be recursively expanded, although we note that these constructions are limited stylistically and to some extent also semantically (Bisetto 2010):

(131) a. *programma* *riciclo* *materiali*
program recycling stuff
[N [NP-Ø [NP-Ø]]]
'stuff recycling program' (Bisetto 2010:28)

b. *ufficio* *responsabile* *reparto* *giocattoli* *e* *attrezzature* *sport-iv-e*
office manager section toys and facilities sport-ADJZ-PL
[N [NP-Ø [NP-Ø [[NP-Ø] [NP-Ø [NP-A]]]]]]
'toys and sports facilities section manager office' (Bisetto 2010:27)

13.3 Latin

Latin is attested as early as the sixth century BCE. The examples presented below are all from later times, mostly from the Classical Latin period (1st century BCE to 3rd/4th century CE) and, in the case of the Bible translation (Weiss 2009:23-24), the Late Latin period (3rd and 4th century CE).

Latin uses genitives, adpositions and, to a minor extent, juxtaposition and adjectivizers for NP embedding. However, only genitives and adpositions are productive strategies to build recursive NP constructions. Head marking is absent altogether.

The genitive is the primary means of expressing relations between nouns and is marked by a suffix whose phonological form depends on the declension class of a given noun. Genitives allow the recursive embedding of NPs:

(132) *morb-us* *et* *aegrotatio,* *ex* *totius* *valetudin-is* *corpor-is* *conquassation-e*
disease-NOM.SG and sickness.NOM.SG out.of all.GEN health-GEN.SG body-GEN.SG disturbance-ABL.SG
[[NP-G [NP-G]] N
et *perturbation-e* *gignuntur*
and disorder-ABL.SG produce.3PL.PRS.PASS
N]
'disease and sickness are entirely produced by disturbance and disorder of the health of the body' (Cicero, Tusculanae Disputationes 4.29)

Latin has a range of prepositions, which govern case. Preposition-marked NPs are compatible with recursive NP structures:

(133) *librum* *de* *vita* *eius*
book.ACC.SG of life.ABL.SG DEM.GEN.SG
[N [NP-P [NP-G]]]
'a book about his life' (Pliniues, Epistolae 4.7.2)

Numerous adjectivizers are present in every period of Latin, but they play a marginal role in the language (Baldi & Nuti 2010:356,364). Unlike genitives, derived adjectives almost never control additional NPs. We have come across only one example from an early comic play by Terentius Afer, first performed 161 BCE, in which a derived adjective embeds a further constituent:

(134) *pater-n-um* *amic-um* *me* *assimul-abo* *uirgin-is*
father-ADJZ-ACC.SG.M friend(M)-ACC.SG 1SG.ACC feign-1SG.FUT girl-GEN.SG
NP-A N NP-G
[*amicum* (N) [*paternum* (NP-A) [*virginis* (NP-G)]]]
'I'll make believe I'm a friend of the girl's father' (Terentius, Phormio 128) (Baldi & Nuti 2010:364)

In this example, the genitive *uirginis* 'of the girl' is a dependent of the derived adjective *paternum* 'of the father' and not the head noun *amicum* 'friend'. However, this example is generally considered very odd, and the lack of additional instances suggests that there is a strong general tendency, grammatical or stylistical, against the modification of derived adjectives and therefore against NP recursion with adjectivizers. This conclusion is supported by the observation that derived adjectives were never very common to begin with. Also, while adjectivizers are explicitly mentioned and discussed by Baldi & Nuti (2010:363–364) as modifiers of possessive pronouns and genitives, adjectivizers modified by another NP are never mentioned. It is conceivable that they may have been more common in the older stages of the language and then gradually fell out of use. However, adjectivized NPs were clearly not a productive means for building recursive NP structures in Classical Latin or later stages of the language.

Compounds exist but are quite rare (Fruyt 2002:259) and there is no evidence whatsoever for recursive compounds. The attested compounds are mostly lexicalized (e.g. *capri-fīcus* [male.goat-fig] 'fig tree' or *angi-portus* 'narrow passage' Fruyt 2002:266-267). The only compounds that are more common are those whose first element is a preverb or preposition as in *con-discupulus* [with-pupil] 'a fellow pupil', not a noun. Thus, juxtaposition does not seem to have been available for recursive NP embedding in Latin.

13.4 Romanian

Romanian has two NP structures that allow recursive NP embedding: genitives and adpositions. The language also makes use of endocentric juxtaposition, but such constructions are not available for NP recursion. Adjectivizers only play a minor role in Romanian and head marking is absent altogether.

Romanian has two cases, an unmarked nominative-accusative that is primarily used for subjects and direct objects, and a genitive-dative case that is used for indirect objects and possession, among other functions. With masculine and neuter nouns, the genitive-dative case is only marked on the definite and indefinite article that accompanies the noun. The noun itself remains unchanged (Example 135a and Example 135b). With feminine nouns, an additional suffix *-i/-e/-ele* (Example 135c and Example 135d) is added (Cojocaru 2004:32):

(135) a. *un* *pom* / *unui* *pom*
ART.INDF.NA.SG.M tree(M) / ART.INDF.GD.SG.M tree(M)
'a tree / of a tree'

b. *pom-ul* / *pom-ului*
tree(M)-ART.DEF.NA.SG.M / tree(M)-ART.DEF.GD.SG.M
'the tree / of the tree'

c. *o* *cas-ă* / *unei* *cas-e*
ART.INDF.NA.SG.F house(F)-NA.SG / ART.INDF.GD.SG.F house(F)-GD.SG
'a house / of a house'

d. *cas-a* / *cas-e-i*
house(F)-ART.DEF.NA.SG.F / house(F)-GD.SG-ART.DEF.GD.SG.F
'the house / of the house' (Cojocaru 2004:33)

The genitive-dative can embed other constituents, as is shown in Example 136, where the genitive *sângelui* 'of the blood' embeds a possessive pronoun.

(136) *izvor-ul* *sânge-lui* *ei*
source/fountain(M)-ART.DEF.NA.SG.M blood(M)-ART.DEF.GD.SG.M 3SG.F.GEN
[N [NP-G [NP-G]]]
'the source/fountain of her blood' (Mark 6.18)

Romanian makes frequent use of prepositions, which govern different cases (Cojocaru 2004:171). Most of them do not occur alone, but as a composite form with the preposition *de* 'of'. Like the genitive case, they can embed further NPs (cf. Example 137).

(137) *cas-a* *de* *lângă* *magazin-ul* *An-e-i*
house(F)-ART.DEF.NA.SG.F of by store(M)-ART.DEF.NA.SG.M Ana(F)-GD.SG-ART.DEF.GD.SG.F
[N [NP-P [NP-G]]]
'the house next to Ana's store' (Selim Özgür, p.c.)

There are several adjectivizing suffixes, e.g. *-iu* as in *argint-iu* [silver-ADJZ] 'silvery' or *-al* as in *anu-al* [year-ADJZ] 'yearly'(Gönczöl-Davies 2008), but we have found no evidence that such derived adjectives can embed other constituents.

Compounds do not play a major role in the Romanian language and determinative compounds such as *floarea-soarelui* [flower-sunshine] 'sunflower' have a low productivity or are even obsolete (Grossmann 2012:148,153). Moreover, we have not come across any examples of recursive compounding. We thus conclude that juxtaposition is not available for recursive NP embedding.

13.5 Sardinian

Our description is based on Cagliari Sardinian. The language has only one construction that allows the recursive embedding of NPs: the genitive. Adjectivization and juxtaposition exist, but these strategies do not allow recursive embedding of NP constituents. There are no adpositions and head marking is absent as well.

The genitive is marked by the proclitic *de*, which is traditionally referred to as a preposition. We analyze it as a genitive though, because it has lost its adpositional behavior such as case government. Unlike in Italian, case is not detectable in pronouns anymore because the oblique pronoun forms only ever occur after a preposition, i.e. there is no situation where the oblique form can be analyzed as case assigned independently of the preposition. The reason for this is that an oblique pronoun in object function requires a differential object marker. The old direct vs. oblique alternation is thus best analyzed as a stem alternation triggered by the case marker that precedes the pronoun, very similar to what happened in several modern Indo-Aryan languages (e.g. in Maithili or Nepali, as described in Sections 11.3 and 11.5 respectively.) The new *de*-genitive allows the recursive embedding of other constituents, such as another genitive or possessive pronoun:

(138) *su cane 'essu piseddu de sorre mia*
ART.SG.M dog(M) GEN.ART.SG.M[14] boyfriend(M) GEN sister(F) 1SG.POSS.SG.F
[N [NP-G [NP-G [NP-A]]]]
'my sister's boyfriend's dog' (Antonello Porcu, p.c.)

There are a number of adjectivizers, such as *-inu* in *berbek-inu* [sheep-ADJ.SG.M] 'ovine' (Simone Pisano, p.c.), but we found no evidence that they can embed further constituents.

Compounding is not a productive process in Sardinian and the recursive expansion of determinative compounds is not possible. Rather, the genitive *de* is used to render what in an other language (e.g. Italian) might be a compound:

(139) *programma de ritzicru de materiales*
program GEN recycling GEN materials
[N [NP-G [NP-G]]]
'stuff recycling program' (Simone Pisano, p.c.)

This suggests that juxtaposition is not available for recursive NP embedding.

13.6 Spanish

In Spanish, only adpositions allow the recursive embedding of NPs. The language also features adjectivization and juxtaposition, but these strategies do not allow recursive NP embedding. Genitives and head marking are absent from the language.

Spanish displays a number of adpositions (Batchelor 2006:205–206). Their status as adpositions is based on the fact that they govern case. Most adpositions govern the prepositional case in combination with first and second person singular pronouns (cf. Example 140c and Example 140d) as well as third person reflexive pronouns, but the nominative case in all other contexts. In addition, there are some adpositions that exclusively govern the nominative case (Butt & Benjamin 1994:106):

[14] The clitic *de* regularly combines with the article and the initial *d* is elided in certain contexts.

(140) a. *Yo* *compr-o* *una* *casa.*
1SG.NOM buy-PRS.1SG ART.INDF.F house(F)
'I buy a house.' (Batchelor 2006:154)

b. *Tú* *compr-as* *una* *casa.*
2SG.NOM buy-PRS.2SG ART.INDF.F house(F)
'You buy a house.' (Batchelor 2006:154)

c. *Habl-an* *de* *mí.*
speak-PRS.3PL of 1SG.PREP
'They speak of me.' (Batchelor 2006:157)

d. *Lo* *hag-o* *por* *ti.*
3SG.M.ACC do-PRS.1SG for 2SG.PREP
'I do it for you.' (Batchelor 2006:157)

Adpositions allow the recursive embedding of NPs:

(141) *el* *perro* *de-l* *padre* *de* *mi* *amigo*
ART.SG.M dog(M) of-ART.SG.M father(M) of 1SG.GEN friend(M)
[NP [NP-P [NP-P [NP-G]]]]
'the dog of my friend's father' (Martha Mariani, p.c.)

Spanish possesses a set of derivational suffixes that derive adjectives from nominals, e.g. *-al* (e.g. *central* 'central' < *centro* 'center') or *-oso* (e.g. *aceitoso* 'oily' < *aceite* 'oil') (Española 2010:133). In the literature, adjectivizers are not described as allowing the embedding of additional NPs. We thus conclude that recursive embedding of NPs is not possible with adjectivizers.

Spanish makes use of juxtaposition to form complex NPs. Such juxtapositions are typically left-headed, e.g. *telaraña* 'spiderweb' < *tela* 'cloth' + *araña* 'spider' or *bocacalle* 'intersection' < *boca* 'mouth' + *calle* 'street entrance, side street' (Española 2010:196-197). We checked with three native speakers whether recursive nominal compounding is possible. We presented each speaker with hypothetical three-member compounds from different semantic fields (e.g. **puerta valla jardín* [gate-fence-garden] 'garden fence gate', **programa reciclaje material* [program-recycling-stuff] 'stuff recycling program', **puchero carne cabra* [stew-meat-goat] 'goat meat stew'). All of them rejected the relevant constructions as ungrammatical. We conclude that juxtapositions is not available for the recursive embedding of NPs in Spanish.

14 Slavic

14.1 Bulgarian

Bulgarian deviates from other Slavic languages in that it only has one major strategy for recursive NP structures, namely adpositions. There is no genitive and no head-marking constructions, and while juxtapositions and adjectivization exist, these strategies do not allow the recursive embedding of NPs.

In Bulgarian there are prepositions which govern case. Because there is no case marking on nouns in the spoken language, this can only be seen on pronouns: e.g. *az* [1SG.NOM] 'I' vs. *na mene* [at 1SG.OBL] 'at me' (Max Wahlström, p.c.). Prepositions can embed other constituents and allow NP recursion:

(142) *tova e domăt na brat mi*
this is house.DEF of brother 1SG.DAT
[N [NP-P [NP-G]]]
'This is the house of my brother.' (Scatton 1993:237)

The language makes use of denominal adjectives to embed nominals into NPs. This strategy is commonly used to express possessive relationships, as in *brat-ov* [brother-ADJZ] 'brother's' (Scatton 1993:221). However, denominal adjectives can never embed an additional NP (Corbett 1987:310).

Juxtaposition is productive in Bulgarian, but recursive structures are marginal at best. It only exists in specialized technical vocabularies. Determinative compounds such as *disko-moda* [disco-fashion] 'disco fashion' exist, but they are confined to two members (Max Wahlström, p.c.).

14.2 Old Church Slavic

Old Church Slavic (henceforth OCS) is attested in a text corpus that is commonly dated to the late 10th century CE. The individual texts are translations of Greek ecclesiastical texts and were first translated into OCS about one hundred years earlier. The texts primarily display dialectal features representative of Balkan Slavic varieties, but also show some influence of Slavic varieties spoken in Moravia (Huntley 1993:125).

OCS uses genitives, adjectivizers, adpositions, and juxtaposition to embed nominals into NPs. Of these, genitives and adpositions are commonly used to build embed additional NP constituents, while adjectivizers are only marginally attested in this function. Juxtaposition does not allow recursive NP embedding, and head marking is absent altogether.

OCS nouns are inflected for case. Recursive NP constructions are commonly formed with the genitive:

(143) *iz očese bratra tvoego*
out.of eye.GEN.SG brother(M).GEN.SG 2SG.POSS.SG.M
[NP-P [NP-G [NP-A]]]
'from your brother's eye' (Codex Marianus, Matthew 7.5)

OCS also uses adpositions, which govern case, for the formation of recursive NP constructions:

(144) *dъva otъ učenikъ svoixъ*
two.ACC.DU of disciple(M).GEN.PL 3SG.POSS.PL.M
[N [NP-P [NP-A]]]
'two of his disciples' (Codex Marianus, Mark 11.1)

There are several adjectivizers in OCS. The resulting denominal adjectives agree in gender, case and number with their head noun. There is marginal evidence that OCS denominal adjectives can embed additional NPs:

(145) *otъ uzdy koń-ьnyj-ę cěsarę*
from bridle(F).GEN.SG horse(M)-ADJZ-GEN.SG.F emperor(M).ADJZ.GEN.SG.F
[N [NP-A [NP-A]]]
'from the bridle of the horse of the emperor' (Codex Suprasliensis 193.9)

However, as Huntley (1993) points out, there are only three instances of a denominal adjective controlling an additional constituent in the entire OCS corpus. In a complex NP in which a possessor NP controls an additional possessor NP or stands in adposition to an additional possessor NP, both dependents are usually in the genitive (Huntley 1993: 218–219; Trubetzkoy 1954:188, 192; Vaillant 1964:133-4). Accordingly, the embedding properties of denominal adjectives are extremely limited, and the genitive is clearly preferred over denominal adjectives for the formation of recursive NP constructions.

A survey of the PROIEL materials (Haug & Jøhndal 2008) confirms this analysis, with one potentially important counterexample: the 76-generation genealogy of Jesus in Luke 3.23 is a string of recursively embedded patronymics in the form of possessive adjectives:

(146) *synъ syi jěko mьnimъ bě Iosif-ov-ъ Ili-ev-ъ Mattat-ov-ъ*
son(M).NOM.SG being as thought was Joseph(M)-ADJZ-NOM.SG.M Eli(M)-ADJZ-NOM.SG.M Matthat(M)-ADJZ-NOM.SG.M
Levi-in-ъ Melxi-ev-ъ Ianna-ev-ъ … (68 more generations)
Levi(M)-ADJZ-NOM.SG.M Melchi(M)-ADJZ-NOM.SG.M Jannai(M)-ADJZ-NOM.SG.M …
Adam-ov-ъ božei
Adam(M)-ADJZ-NOM.SG.M God(M).ADJZ.NOM.SG.M
'being (as was supposed) the son of Joseph, which was the son of Heli, which was the son of Matthat, which was the son of Levi, which was the son of Melchi, which was the son of Jannai … which was the son of Adam, which was the son of God' (Codex Marianus, Luke 3.23)

The Greek original has a recursive string of genitives:

(147) *huiós […] Iōsḕf toũ Ēlì toũ Mat^h^t^h^àt toũ Leuì*
son.NOM.SG […] Joseph(M) ART.GEN.SG.M Eli(M) ART.GEN.SG.M Matthat(M) ART.GEN.SG.M Levi(M)
toũ Melk^h^ì toũ Iannaì […] toũ Adàm toũ T^h^eoũ
ART.GEN.SG.M Melchi(M) ART.GEN.SG.M Jannai(M) […] ART.GEN.SG.M Adam(M) ART.GEN.SG.M God(M).GEN.SG

'the son … of Joseph, which was the son of Heli, which was the son of Matthat, which was the son of Levi, which was the son of Melchi, which was the son of Janna … which was the son of Adam, which was the son of God' (Luke 3.23)

The Slavic rendition faithfully copies the recursion, but uses possessive adjectives rather than calquing the Greek morphology. It is impossible to know whether this OCS example is vestigial or embryonic, but it shows that recursion was a marginal possibility for medieval Slavic, at least in the specific context of patronymics. Given the limits of the evidence and the strong preference for genitives, we conclude that adjectivization is not available as a general strategy for NP recursion in OCS.

OCS makes use of juxtaposition for complex NPs, e.g. *bogo-mati* [god-mother] 'mother of god' (Vaillant 1974:738). However, we have not come across any such constructions with more than two members. This suggests that recursive embedding was not common with juxtaposition in OCS, if it existed at all. The other Slavic language with a sizable medieval attestation, Old Russian (including its distinctive Old Novgorod variant), appears to have been identical to OCS in all these respects (see also footnote 14 in Section 14.3 just below.)

14.3 Russian

Russian uses genitives, adjectivizers, adpositions, and juxtaposition for NP embedding, but only genitives and adpositions are able to embed an additional NP. There is no head-marking construction.

Russian nouns are inflected for case, one of which is the genitive. Genitives allow NP recursion:

(148) *teksty pesen raznyx ispolnitelej i grupp*
text.NOM.PL song.GEN.PL various.GEN.PL performer.GEN.PL and group.GEN.PL
[N [NP-G [NP-G and NP-G]]]
'texts of songs of various performers and groups' (Timberlake 2004:205)

Russian also has prepositions. Some of them govern only one case; others can combine with multiple cases (Andrews 2001:66). They too can embed additional NPs:

(149) *mal'čik iz derevni na reke*
boy.NOM.SG from village.GEN.SG on river.PREP.SG
[N [NP-P [NP-P]]]
'the boy from the village by the river' (Jekaterina Mažara, p.c.)

Russian has adjectivizers that are frequently used to derive denominal adjectives, in particular to express possessive relationships (Timberlake 2004:127). However, such adjectives cannot embed additional NPs (Corbett 1987:308-309):[15]

(150) (**Ivan-ov-a/Ivan-a*) *mam-in-a kniga*
Ivan-ADJZ-SG.F/Ivan-GEN.SG mother(F)-ADJZ-SG.F book(F)
Intended: 'Ivan's mother's book.'

Juxtaposition is not a common strategy to build recursive NP structures. Schönle (1975:151) lists a number of compounds with three members, e.g. *ugle-rudo-voz* [coal-ore-load] 'load of

[15] There are a very few examples suggesting that this may have been possible in Old Russian (Corbett 1987:309):

toě Marf-in-ymŭ mužemŭ
that.GEN.SG.F Marfa(F)-ADJZ-INS.SG.M husband(M).INS.SG
[[[DEM] NP-A] N]
'that Marfa's husband' (Corbett 1987:309)

coal ore'. However, as Schönle (1975:153) points out, such instances of recursive compounding are quite rare. They might be productive only for very few final elements, e.g. *samo-leto-stroenie* [[self-fly-]building] 'airplane building', but note that the first compound is already lexicalized. Moreover, they are confined to specific technical jargons.

14.4 Upper Sorbian

Upper Sorbian has genitives, adjectivizers, adpositions, and juxtaposition. Both genitives and denominal adjectives can be used to build recursive NP structures, while juxtaposition does not allow recursive embedding. Head marking is absent from Upper Sorbian.

Upper Sorbian has seven cases, one of which is the genitive (Schuster-Šewc 1996:65). Genitives allow NP recursion:

(151) *praw-o* *naš-ich* *muž-ow*
right-NOM.SG 1PL.POSS-GEN.PL husband-GEN.PL
[N [[NP-A] NP-G]]
'the right of our husbands' (Corbett 1987:302)

Adjectivizers are very productive in Upper Sorbian (Schuster-Šewc 1996:104), and they can recursively embed further constituents. In Example 152, the denominal adjective is modified by another denominal adjective:

(152) *přez* *Mar-in-eje* *maćer-n-u* *smjerć*
through Marja-ADJZ-GEN.SG.F mother(F)-ADJZ-ACC.SG.F death(F).ACC.SG
[[[NP-A] NP-A] N]
'through Marja's mother's death' (Lötzsch 1965:378)

Denominal adjectives agree in case, gender and number with their immediate head. Thus, in Example 152, the first level adjective *maćernu* 'mother's' agrees with the head noun *smjerć* 'death', and the second level adjective *Marineje* 'Maria's' with the nominal head contained in the adjectivized form *maćernu*.

There are several prepositions in Upper Sorbian which govern case (Schuster-Šewc 1996:218). They allow NP recursion:

(153) *hólc* *z* *wjeski* *při* *rěce*
boy.NOM.SG from village.GEN.SG at river.LOC.SG
[N [NP-P [NP-P]]]
'the boy from the village by the river' (Sonja Wölke, p.c.)

Stone (1993:651) describes a number of determinative compounds with two members, but does not mention any instances of recursive compounding. Schuster-Šewc (1991) lists many compounds, none of which have more than two members. According to Sonja Wölke (p.c.), nominal compounding is indeed strongly restricted in Upper Sorbian to the extent that a compound cannot have more than two members. Thus, the language does not allow juxtaposition for recursive NP embedding.

15 Tocharian

15.1 Tocharian B

Tocharian refers to a branch of Indo-European which is attested in two closely related languages or dialects, Tocharian A (eastern variant) and B (western variant). They are attested from the 5th to 9th century CE in manuscripts found around the Turpan Oasis in the modern-day Chinese province Xinjiang. Most of the documents are translations of Sanskrit Buddhist texts (Krause & Thomas 1960:37-38). Our description is based on Tocharian B, as this is the language also sampled in the IELEX data used by Chang et al. (2015).

Tocharian B exhibits four of the NP structures that we surveyed: it has case, adjectivizers, adpositions, and juxtaposition. Both the genitive and the adjectivizers can embed further NPs. Adpositions cannot be used adnominally and head marking is absent altogether.

Genitives allow NP recursion as shown in (Example 154):[16]

(154) *oṅkolmä-ṃts* *lānte* *sayi*
elephant-GEN.PL king.GEN.SG son.GEN.SG
[[NP-G] NP-G] N]
'of the son of the king of elephants' (THT 74a4) (Adams 2009:304)

In the literature (Carling 2000, Krause & Thomas 1960:171), we found no evidence that pre- and postpositions in Tocharian are used adnominally in Tocharian B and conclude that such constructions do not exist in the language.

The suffixes *-(i)ye* , *-ññe* , or *-ṣṣe* are used as adjectivizers, with *-ṣṣe* being the most widely used suffix (Adams 2009:306-307). As with genitives, NP recursion is possible, i.e. an adjectivized noun can be modified by an other constituent. If the other constituent includes an adjectivizer as well, the second-level embedded adjective agrees in gender with the underlying noun of the first level adjective and not with the topmost head noun (Adams 2009:304-305), i.e. agreement is normally with the immediate head:

(155) *laksa-ññ-ai* *klautsai-ṣṣ-e* *ṣpe[l]*
fish-ADJZ-OBL.SG.F ear(F)-ADJZ-NO.SG.M poultice(M).NO.SG
[[[NP-A] NP-A] N]
'a poultice of fish ears' (PK AS 3Bb2) (Adams 2009:304)

The ability to form compounds is quite limited in Tocharian, and many of the attested compounds are loan translations from Sanskrit. Determinative compounds do exist, as for example *syalle-were* [sweat-smell] 'smell of sweat' or *tana-mot* [grain-alcoholic.drink] 'grain brandy' (Bernhard 1958:166,167). However, we have not come across any example with three members in the extensive survey by Bernhard (1958), and it is thus safe to conclude that juxtaposition was not available for recursive NP embedding in Tocharian.

[16] Writing conventions: [ ]=damaged sign, ( )=restored parts of missing signs.

References

Adams, Douglas Q. 2009. Genitive and Adjective in Tocharian. *Journal of Indo-European Studies* 37(3&4). 299–320.

Allen, Cynthia. 2008. *Genitives in early English: typology and evidence.* Oxford: Oxford University Press.

Ambrazas, Vyautas. 1997. *Lithuanian grammar.* Lithuania: Baltos Lankos.

Andrews, Edna. 2001. *Russian.* München: Lincom Europa.

Babrakzai, Farooq. 1999. *Topics in Pashto Syntax*: University of Hawaii at Manoa PhD dissertation.

Baldi, Philip & Andrea Nuti. 2010. Possession. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New Perspectives on Historical Latin Syntax. Constituent Syntax: Quantification, Numerals, Possession, Anaphora*, vol. 3, Berlin, Boston: Mouton De Gruyter.

Batchelor, Ronald Ernest. 2006. *A student grammar of Spanish.* Cambridge University Press.

Batchelor, Ronald Ernest & Malliga Chebli-Saadi. 2011. *A reference grammar of French.* Cambridge University Press.

Bauer, Anna. 2014. *Morphosyntax of the Noun Phrase in Hieroglyphic Luwian.* Leiden: Brill.

Behaghel, Otto. 1897. *Die Syntax des Heliand.* Prag, Wien, Leipzig: F. Tempski.

Belyayev, Oleg. 2010. Evolution of case in Ossetic. *Iran and the Caucasus* 14. 287–322.

Bernhard, Franz. 1958. *Die Nominalkomposition im Tocharischen*: Universität Göttingen PhD dissertation.

Bisetto, Antonietta. 2010. Recursiveness and Italian compounds. *SKASE Journal of theoretical linguistics* 7. 14–35.

Boretzky, Norbert. 1994. *Romani.* Wiesbaden: Harrassowitz.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The Syntax of Welsh.* Cambridge: Cambridge University Press.

Braune, Wilhelm. 2004. *Althochdeutsche Grammatik I. Bearbeitet von Ingo Reiffenstein.* Tübingen: Niemeyer 15th edn.

Braune, Wilhelm & Ernst A. Ebbinghaus. 1966. *Gotische Grammatik mit Lesestücken und Wörterverzeichnis.* Tübingen: Niemeyer.

Brunner, Christopher. 1977. *A syntax of western Middle Iranian* (Persian studies series 3). Delmar, NY: Caravan Books.

Brunner, Karl. 1965. *An outline of Middle English Grammar (translated by Grahame Johnston).* Cambridge, MA: Harvard University Press.

Bubenik, Vit. 1998. *A historical syntax of late Middle Indo-Aryan (Apabhramsa).* Amsterdam, Philadelphia: John Benjamins.

Bubenik, Vit. 2009. The rise and development of the possessive construction in Middle Iranian with parallels in Albanian. In Vit Bubenik, John Hewson & Sarah Rose (eds.), *Grammatical change in Indo-European languages*, 89–101. Amsterdam: Benjamins.

Buchholz, Oda & Wilfried Fiedler. 1987. *Albanische Grammatik.* Leipzig: Verlag Enzyklopädie.

Butt, John & Carmen Benjamin. 1994. *A new reference grammar of Modern Spanish.* Edward Arnold 2nd edn.

Carling, Gerd. 2000. *Die Funktionen der lokalen Kasus im Tocharischen.* Berlin: Mouton de Gruyter.

Carr, Charles Telford. 1939. *Nominal compounds in Germanic.* London: St. Andrews University.

Cassidy, Frederic G. & Richard N. Ringler. 1971. *Bright's Old English grammar & reader.* New York: Holt, Rinehart and Winston 3rd edn.

Cathey, James E. 2000. *Old Saxon.* München: Lincom Europa.

Chandralal, Dileep. 2010. *Sinhala.* Amsterdam, Philadelphia: John Benjamins.

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

Cojocaru, Dana. 2004. *Romanian grammar.* Durham, NC: Slavic and East European Language Research Center (SEELRC), Duke University.

Corbett, Greville G. 1987. The morphology/syntax interface: evidence from possessive adjectives in Slavonic. *Language* 299–345.

David, Anne. 2014. *Descriptive Grammar of Pashto and its Dialects.* Boston: Walter de Gruyter.

Donaldson, Bruce C. 1993. *A grammar of Afrikaans.* Berlin, New York: Mouton De Gruyter.

Dum-Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian*, vol. 14 London Oriental and African Language Library. Amsterdam, Philadelphia: John Benjamins.

Durkin-Meisterernst, Desmond. 2014. *Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch).* Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Duroiselle, Charles. 1997. *A Practical Grammar of the Pāli Language.* Lismore, Australia: Buddha Dharma Education Association 3rd edn.

Endzelīns, Jānis. 1971. *Comparative phonology and morphology of the Baltic languages.* The Hague & Paris: Mouton.

Ernault, Emile. 1887. La vie de sainte Catherine. *Revue Celtique* 8. 76–95.

Erschler, David. 2016. Ossetic. In Peter O. Mueller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation. An International Handbook of the languages of Europe*, vol. 5, chap. 173, 3157–3179. Berlin, Boston: Mouton De Gruyter.

Española, Real Academia. 2010. *Nueva gramática de la lengua española.* Espasa.

Evans, D. Simon. 1964. *A Grammar of Middle Welsh.* Dublin: Dublin Institute for Advanced Studies.

Faarlund, Jan Terje. 2004. *The Syntax of Old Norse. With a survey of the inflectional morphology and a complete bibliogaphy.* Oxford: Oxford University Press.

Fahs, Achim. 1989. *Grammatik des Pali.* Leipzig: Verlag Enzyklopädie 2nd edn.

Fleischer, Jürg & Oliver Schallert. 2011. *Historische Syntax des Deutschen. Eine Einführung.* Tübingen: Narr.

Fruyt, Michèle. 2002. Constraints and productivity in Latin nominal compounding. *Transactions of the Philological Society* 100(3). 259–287.

Gair, James W. 1970. *Colloquial Sinhalese clause structures.* The Hague: Mouton De Gruyter.

Gignoux, Philippe & A. Tafazzoli (eds.). 1993. *Anthologie de Zādspram. Édition critique du texte pehlevi.* Paris: Association pour l'Avancement des Etudes Iraniennes.

Godel, Robert. 1975. *An introduction to the Study of Classical Armenian.* Wiesbaden: Reichert.

Gönczöl-Davies, Ramona. 2008. *Romanian: An Essential Grammar.* New York: Routledge.

Grierson, George A. 1909. *An introduction to the Maithilī dialect of the Bihārī language as spoken in North Bihār.* Calcutta: Asiatic Society of Bengal [2nd edition].

Grossmann, Maria. 2012. Romanian compounds. *Probus* 24(1).

Hancock, Ian. 1995. *A handbook of Vlax Romani.* Columbus, OH: Slavica Publishers.

Harðarson, Gísli Rúnar. 2016. Peeling away the layers of the onion: on layers, inflection and domains in Icelandic compounds. *Journal of Comparative German Linguistics* 19. 1–47.

Haug, Dag T. T. & Marius L. Jøhndal. 2008. Creating a Parallel Treebank of the Old Indo-European Bible Translations. In Caroline Sporleder & Kiril Ribarov (eds.), *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH)*, 27–34.

Hawkins, J. David. 2000. *Corpus of Hieroglyphic Luwian Inscriptions. Vol. 1 Inscriptions of the Iron Age.* Berlin/New York: de Gruyter.

Hemon, Roparz. 1975. *A historical morphology and syntax of Breton.* Dublin: Dublin Institute for Advanced Studies.

Hemon, Roparz. 1976–1998. *Geriadur istorel ar brezhoneg.* Plomelin: Preder 2nd edn.

Hickey, Raymond. 2011. *The dialects of Irish: Study of a changing landscape.* Berlin, Boston: De Gruyter.

Hinüber, Oskar von. 2001. *Das ältere Mittelindisch im Überblick*. Vienna: Österreichische Akademie der Wissenschaften 2nd edn.

Hoffmann, Karl & Bernhard Forssman. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Hoffner, Harold & H. Craig Melchert. 2008. *A Grammar of the Hittite language*. Winona Lake, ID: Eisenbrauns.

Holmes, Philip & Ian Hinchliffe. 2013. *Swedish: a comprehensive grammar*. London: Routledge.

Huntley, David. 1993. Old Church Slavonic. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 125–187. London: Routledge.

Hægestads, Marius & Alf Torp. 1909. *Gamalnorsk ordbok*. Kristiania: Norske Samlaget.

Igla, Birgit. 1996. *Das Romani von Ajia Varvara: deskriptive und historisch-vergleichende Darstellung eines Zigeunerdialekts*. Wiesbaden: Harrassowitz.

Jahani, Carina & Agnes Korn. 2009. Balochi. In Gernot Windfuhr (ed.), *The Iranian Languages*, 634–692. New York: Routledge.

Kastovsky, Dieter. 2006. Vocabulary. In Richard Hogg & David Denison (eds.), *A history of the English language*, chap. 4, 199–270. Cambridge: Cambridge University Press.

King, Gareth. 2003. *Modern Welsh: A Comprehensive Grammar*. London, New York: Routledge.

Klein, Thomas, Hans-Joachim Solms & Klaus-Peter Wegera. 2009. *Mittelhochdeutsche Grammatik: Wortbildung*, vol. 3. Tübingen: Niemeyer.

Kotin, Michail L. 2012. *Gotisch im (diachronen und typologischen) Vergleich*. Heidelberg: Winter.

Krause, Wolfgang & Werner Thomas. 1960. *Tocharisches Elementarbuch – Grammatik*, vol. 1. Heidelberg: Winter.

Kress, Bruno. 1982. *Isländische Grammatik*. München: Hueber.

Lazard, Gilbert. 1992. *A grammar of contemporary Persian*. Costa Mesa, CA: Mazda Publishers.

Le Berre, Yves. 2011. *La passion et la résurrection bretonnes en 1530*. Brest: C.R.B.C.

Lötzsch, Ronald. 1965. Das sog. Possessivadjektiv im Slawischen, speziell im Sorbischen, und seine Stellung im System der Redeteile. *Forschungen und Fortschritte* 39. 377–379.

Macdonell, Arthur Anthony. 1916. *A Vedic Grammar for Students*. Oxford: Clarendon Press.

MacKenzie, David N. 1979. Mani's Šāhburagān I. *Bulletin of the School of Oriental and African Studies* 41. 499–511.

Mackridge, Peter. 1985. *The Modern Greek Language*. Oxford: Oxford University Press.

Maiden, Martin & Cecilia Robustelli. 2007. *A reference grammar of modern Italian*. London: Routledge 2nd edn.

Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Matras, Yaron. 2002. *Romani – A Linguistic Introduction*. Cambridge: Cambridge University Press.

Melchert, H. Craig. 2008. Luvian. In Roger D. Woodard (ed.), *The Ancient Languages of Asia Minor*, Cambridge: Cambridge University Press.

Montaut, Annie. 2004. *A grammar of Hindi*. München: Lincom Europa.

Morris Jones, John. 1913. *A Welsh Grammar*. Oxford: Clarendon.

Mossé, Fernand. 1952. *A Handbook of Middle English*. Baltimore: Johns Hopkins Press.

Neukom, Lukas & Manideepa Patnaik. 2003. *A grammar of Oriya* (ASAS 17). Universität Zürich.

Newmark, Leonard, Philip Hubbard & Peter Prifti. 1982. *Standard Albanian: a reference grammar for students*. Stanford, CA: Stanford University Press.

Norde, Muriel. 2013. Tracing the origins of the Swedish group genitive. In Anne Carlier & Jean-Christophe Verstraete (eds.), *The Genitive*, 299–332. Amsterdam, Philadelphia: John Benjamins.

Ó Siadhail, Mícheál. 1980. *Learning Irish: an introductory self-tutor*. Dublin: Dublin Institute for Advanced

Studies.

Ó Siadhail, Mícheál. 1989. *Modern Irish. Grammatical Structure and Dialectal Variation.* Cambridge: Cambridge University Press.

Oberlies, Thomas. 2001. *Pāli: A Grammar of the Language of the Theravāda Tipiṭaka.* Berlin, New York: Walter de Gruyter.

Pandharipande, Rajeshwari V. 1997. *Marathi.* London: Routledge.

Paul, Hermann. 1998. *Mittelhochdeutsche Grammatik. überarbeitet von Peter Wihl und Siegfried Grosse.* Tübingen: Niemeyer 24th edn.

Payne, Annick. 2004. *Hieroglyphic Luwian.* Wiesbaden: Harrassowitz.

Pilch, Herbert. 1970. *Altenglische Grammatik.* München: Hueber.

Prauliņš, Dace. 2012. *Latvian: An essential grammar* Routledge Essential Grammars. Routledge.

Prell, Heinz-Peter. 2005. Zur Neubearbeitung der mittelhochdeutschen Syntax von Ingeborg Schröbler. Mit einem Vergleich von Vers-und Prosatexten am Beispiel der Nominalphrase. In Claudia Wich-Reif & Yvon Desportes (eds.), *Syntax: Althochdeutsch-Mittelhochdeutsch: eine Gegenüberstellung von Metrik und Prosa: Akten zum internationalen Kongress an der Freien Universität Berlin, 26. bis 29. Mai 2004*, 209–222. Berlin: Weidler.

Press, Ian. 1986. *A grammar of modern Breton.* Berlin, New York: Mouton de Gruyter.

Ralli, Angela. 2013. *Compounding in modern Greek.* Berlin: Springer.

Roberge, Paul T. 1996. Multilevel syncretism and the evolution of Afrikaans periphrastic possessives with se. *Stellenbosch Papers in Linguistics* 30. 123–153.

Ruthmann, Davina. 2007. *Entwicklung des Wortschatzes vom Mittelhochdeutsch bis zum Gegenwartsdeutsch.* Bergische Universität Wuppertal MA thesis.

Sandahl, Stella. 2000. *A Hindi reference grammar.* Leuven: Peeters.

Sauer, Hans. 1992. *Nominalkomposita im Frühmittelenglischen.* Tübingen: Mouton De Gruyter.

Scatton, Ernest A. 1993. Bulgarian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, 188–248. London, New York: Routledge.

Schikowski, Robert. 2013. *Object-conditioned differential marking in Chintang and Nepali*: University of Zürich Doctoral dissertation.

Schmalstieg, William R. 1987. *A Lithuanian historical syntax.* Columbus, OH: Slavica.

Schönle, Paul Walter. 1975. *Zur Wortbildung im modernen Russisch.* München: Sagner.

Schrijver, Peter. 2011. *Middle and Early Modern Breton* 359–429. Bremen: Hempen.

Schrodt, Richard. 2004. *Althochdeutsche Grammatik II.* Tübingen: Niemeyer.

Schuster-Šewc, Heinz. 1991. Zur lexikalischen Nomination im Sorbischen. *Zeitschrift für Slawistik* 36(2). 261–270.

Schuster-Šewc, Heinz. 1996. *Grammar of the Upper Sorbian Language.* München, Newcastle: Lincom Europa.

Schwyzer, Eduard. 1990a. *Griechische Grammatik*, vol. 2: Syntax und Syntaktische Stilistik. München: C.H.Beck 6th edn.

Schwyzer, Eduard. 1990b. *Griechische Grammatik*, vol. 1: Lautlehre, Wortbildung, Flexion. München: C.H.Beck 6th edn.

Skjærvø, Prods Oktor. 2009. Old Iranian. In Gernot Windfuhr (ed.), *The Iranian Languages*, 43–195. New York: Routledge.

Skjærvø, Prods Oktor. 2009. *Middle West Iranian* 197–278. New York: Routledge.

Smyth, Herbert Weir. 1920. *Greek Grammar for Colleges.* New York: American Book Company.

Snell, Rupert & Simon Weightman. 2003. *Teach yourself Hindi.* London: Hodder and Stoughton Educational.

Splett, Jochen. 2000. Wortbildung des Althochdeutschen. In Werner Besch, Anne Betten, Oskar Reichmann & Stefan Sonderegger (eds.), *Geschichte der deutschen Sprache und ihrer Erforschung*, vol. 2.2 Handbooks of Linguistics and Communication Science, chap. 76, 1213–1222. Berlin/New York: Walter de Gruyter 2nd edn.

Stang, Christian S. 1966. *Vergleichende Grammatik der baltischen Sprachen.* Oslo/Bergen/Tromsö: Universitetsforlaget.

Stone, Gerald. 1993. Sorbian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, London, New York: Routledge.

Tegey, Habibullah & Barbara Robson. 1996. *A Reference Grammar of Pashto.* Washington, D.C.: Center for Applied Linguistics.

Thordarson, Fridrik. 2009. *Ossetic grammatical studies.* Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Thráinsson, Höskuldur. 1994. Icelandic. In Ekkehard König & Johan Van der Auwera (eds.), *The Germanic Languages*, chap. 6, 142–189. London, New York: Routledge.

Thurneysen, Rudolf. 1909. *Handbuch des Alt-Irischen.* Heidelberg: Winter.

Thurneysen, Rudolf. 1946. *A Grammar of Old Irish.* Dublin: Dublin Institute for Advanced Studies revised and enlarged edition with supplement edn.

Tilakaratne, Sunanda. 1992. Spatial Expressions in Sinhala: Appearance of Verb Forms. *Kansas Working Papers in Linguistics* 17(1). 159–189.

Timberlake, Alan. 2004. *A reference grammar of Russian.* Cambridge: Cambridge University Press.

Vaillant, André. 1974. *Grammaire comparée des langues slaves: La formation des noms*, vol. IV. Paris: Ed. Klincksieck.

Vallé, François. 1980. *Grand dictionnaire français-breton.* Glornel: Association bretonne de culture 2nd edn.

Van Gelderen, Elly. 2014. *A history of the English language.* John Benjamins revised edition edn.

Verma, Manindra K. 1971. *The structure of the noun phrase in English and Hindi.* Delhi: Motilal Banarsidass.

Wackernagel, Jacob. 1905. *Altindische Grammatik: Einleitung zur Wortlehre. Nominalkomposition*, vol. 2 1. Göttingen: Vandenhoeck & Ruprecht.

Wali, Kashi & Omkar N. Koul. 1996. *Kashmiri.* London, New York: Routledge.

Warder, A. K. 2005 [1991]. *Introduction to Pāli.* Oxford: The Pali Text Society 3rd edn.

Weiss, Michael L. 2009. *Outline of the Historical and Comparative Grammar of Latin.* Ann Arbor, MI: Beech Stave PressInc.

Widmer, Paul. 2017. Cases, paradigms, affixes and indexes. Selecting grammatical relations in Middle Breton. In Erich Poppe, Karin Stüber & Paul Widmer (eds.), *Referential properties and their impact on the syntax of Insular Celtic languages*, Münster: Nodus.

Wijesekera, O.H. de A. 1936. *Syntax of the Cases in Pali Nikayas.* Sri Lanka: The Postgraduate Institute of Pali and Buddhist Studies, Unversity of Kelaniya PhD dissertation.

Willis, David. 2009. Old and Middle Welsh. In Martin J. Ball & Nicole Müller (eds.), *The Celtic Languages*, 117–160. London, New York: Routledge.

Windfuhr, Gernot & John R Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian Languages*, 416–544. London, New York: Routledge.

Wright, Joseph. 1954. *Grammar of the Gothic Language.* Oxford: Clarendon Press.

Yadav, Ramwatar. 1996. *A reference grammar of Maithili.* Berlin: Mouton de Gruyter.

Yoshida, Daisuke. 1987. *Die Syntax des althethitischen substantivischen Genitivs.* Heidelberg: Winter.

Zimmer, Stefan. 2000. *Studies in Welsh Word-Formation.* Dublin: Dublin Institute for Advanced Studies.

NP Recursion Over Time – Supporting Material 2: Phylogenetic Analyses

Balthasar Bickel

November 30, 2016

Contents

List of Tables

List of Figures

1 Data to Tree Mapping

The data are available in binary form as Supporting Material 3:

```
ie.data <-  read.csv('som3_database.csv', stringsAsFactors=F, strip.white=T)
```

Our phylogenetic models are chiefly based on the latest published estimates. This is the phylogeny developed by Chang et al. (2015). The posterior probability sample of trees are taken from the Supporting Material of this paper, available at https://muse.jhu.edu/journals/language/v091/91.1.chang01.html. We chose the ancestry-constrained tree that shows the closest fit in taxa with our data. This is tree `c1` in their dataset. The tree covers most of our languages, and at the same time has an excess of only 54 languages. (`c2` would include just as many of our data, but it has a higher excess of data that we do not need. Its broader data basis is not necessarily an advantage because this also entails more empty cells. See Chang et al. (2015) for discussion). There is only one `c1` tree, which we used.

The Supporting Materials contain a `BEAST XML` specification file and a Maximum Clade Credibility (MCC) tree summarizing the posterior probability sample. We re-ran the `BEAST` analysis according to Chang et al.'s recipe and generated a posterior probability sample of 20,001 trees.

In order to assess whether our results are robust against different assumptions in phylogeny estimates, we also model our data on the earlier phylogeny developed by Bouckaert et al. (2012). The posterior probability sample of this phylogeny, as well as an MCC summary tree, was made available to use by Remco Bouckaert. All tree files were read in as follows. To avoid ambiguity in the script, we call the trees from Chang et al. (2015) simply `ie.tree(s)`, those from Bouckaert et al. (2012) `ie.b.tree(s)`. In the text we furthermore distinguish them as "C Trees" vs. "B Trees".

```
ie.tree.sum <- read.nexus('a1-c1-d0-g1-l2-s1-t1-z3/mcc.trees')
ie.b.tree.sum <- read.nexus('ie-trees-bouckaertetal2012/IE2011_RelaxedCovarion_AllSingletonsGeo_Combined
ie.trees <- read.nexus('changetal_c1_rerun/ieo.trees')
ie.b.trees <- read.nexus('ie-trees-bouckaertetal2012/IE2011_RelaxedCovarion_AllSingletonsGeo_Combined.ti
```

In some cases we have sampled slightly different language varieties or dialects than either of earlier studies, but the branch lengths and topologies are virtually identical. (For discussion of each difference in the sample, see the relevant language descriptions in Supporting Material 1.) To accomodate these cases, and also cases where we chose a more distinctive name (e.g. Modern Persian vs. Persian), we edit the language names in the trees. At the same time we remove the tips that are not in our sample:

```
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Tosk'] <- 'Albanian'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Eastern_Armenian'] <- 'Modern_Armenian'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Lithuanian'] <- 'Modern_Lithuanian'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Old_West_Norse'] <- 'Old_Norse'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Digor_Ossetic'] <- 'Ossetic'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Cagliari'] <- 'Sardinian'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Breton'] <- 'Modern_Breton'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Irish'] <- 'Modern_Irish'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Welsh'] <- 'Modern_Welsh'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'English'] <- 'Modern_English'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Persian'] <- 'Modern_Persian'
ie.tree.sum$tip.label[ie.tree.sum$tip.label %in% 'Bihari'] <- 'Maithili'
ie.tree.sum <- drop.tip(ie.tree.sum, tip=setdiff(ie.tree.sum$tip.label, ie.data$language))

ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Armenian_Mod"] <- "Modern_Armenian"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Greek_Mod"] <- "Modern_Greek"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Albanian_Top"] <- "Albanian"
```

```
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Afghan"] <- "Pashto"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Persian_List"] <- "Modern_Persian"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% 'Digor_Ossetic'] <- 'Ossetic'
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Nepali_List"] <- "Nepali"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Bihari"] <- "Maithili"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Romanian_List"] <- "Romanian"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Sardinian_C"] <- "Sardinian"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Breton_ST"] <- "Modern_Breton"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Irish_A"] <- "Modern_Irish"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Welsh_N"] <- "Modern_Welsh"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "English_ST"] <- "Modern_English"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "German_ST"] <- "Swiss_German"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Dutch_List"] <- "Afrikaans"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Icelandic_ST"] <- "Icelandic"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Swedish_List"] <- "Swedish"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Old_Church_Slavonic"] <- "Old_Church_Slavic"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Lusatian_U"] <- "Upper_Sorbian"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Lithuanian_ST"] <- "Modern_Lithuanian"
ie.b.tree.sum$tip.label[ie.b.tree.sum$tip.label %in% "Luvian"] <- "Luwian"
ie.b.tree.sum <- drop.tip(ie.b.tree.sum, tip=setdiff(ie.b.tree.sum$tip.label, ie.data$language))

ie.trees <- lapply(ie.trees, function(tree) {
    tree$tip.label[tree$tip.label %in% 'Tosk'] <- 'Albanian'
    tree$tip.label[tree$tip.label %in% 'Eastern_Armenian'] <- 'Modern_Armenian'
    tree$tip.label[tree$tip.label %in% 'Lithuanian'] <- 'Modern_Lithuanian'
    tree$tip.label[tree$tip.label %in% 'Old_West_Norse'] <- 'Old_Norse'
    tree$tip.label[tree$tip.label %in% 'Digor_Ossetic'] <- 'Ossetic'
    tree$tip.label[tree$tip.label %in% 'Cagliari'] <- 'Sardinian'
    tree$tip.label[tree$tip.label %in% 'Breton'] <- 'Modern_Breton'
    tree$tip.label[tree$tip.label %in% 'Irish'] <- 'Modern_Irish'
    tree$tip.label[tree$tip.label %in% 'Welsh'] <- 'Modern_Welsh'
    tree$tip.label[tree$tip.label %in% 'English'] <- 'Modern_English'
    tree$tip.label[tree$tip.label %in% 'Persian'] <- 'Modern_Persian'
    tree$tip.label[tree$tip.label %in% 'Bihari'] <- 'Maithili'
    tree <- drop.tip(tree, tip=setdiff(tree$tip.label, ie.data$language))
    })

ie.b.trees <- lapply(ie.b.trees, function(tree) {
    tree$tip.label[tree$tip.label %in% "Armenian_Mod"] <- "Modern_Armenian"
    tree$tip.label[tree$tip.label %in% "Greek_Mod"] <- "Modern_Greek"
    tree$tip.label[tree$tip.label %in% "Albanian_Top"] <- "Albanian"
    tree$tip.label[tree$tip.label %in% "Afghan"] <- "Pashto"
    tree$tip.label[tree$tip.label %in% "Persian_List"] <- "Modern_Persian"
    tree$tip.label[tree$tip.label %in% 'Digor_Ossetic'] <- 'Ossetic'
    tree$tip.label[tree$tip.label %in% "Nepali_List"] <- "Nepali"
    tree$tip.label[tree$tip.label %in% "Bihari"] <- "Maithili"
    tree$tip.label[tree$tip.label %in% "Romanian_List"] <- "Romanian"
    tree$tip.label[tree$tip.label %in% "Sardinian_C"] <- "Sardinian"
    tree$tip.label[tree$tip.label %in% "Breton_ST"] <- "Modern_Breton"
    tree$tip.label[tree$tip.label %in% "Irish_A"] <- "Modern_Irish"
    tree$tip.label[tree$tip.label %in% "Welsh_N"] <- "Modern_Welsh"
    tree$tip.label[tree$tip.label %in% "English_ST"] <- "Modern_English"
    tree$tip.label[tree$tip.label %in% "German_ST"] <- "Swiss_German"
```

```r
    tree$tip.label[tree$tip.label %in% "Dutch_List"] <- "Afrikaans"
    tree$tip.label[tree$tip.label %in% "Icelandic_ST"] <- "Icelandic"
    tree$tip.label[tree$tip.label %in% "Swedish_List"] <- "Swedish"
    tree$tip.label[tree$tip.label %in% "Old_Church_Slavonic"] <- "Old_Church_Slavic"
    tree$tip.label[tree$tip.label %in% "Lusatian_U"] <- "Upper_Sorbian"
    tree$tip.label[tree$tip.label %in% "Lithuanian_ST"] <- "Modern_Lithuanian"
    tree$tip.label[tree$tip.label %in% "Luvian"] <- "Luwian"
    tree <- drop.tip(tree, tip=setdiff(tree$tip.label, ie.data$language))
    })
```

We then graft each of our sample languages that are not already included onto the tree(s) as a sister or ancestor of an existing reference tip, by keeping the total branch lengths of the existing tips and nodes. Thus, given a lineage N0 – N1 – N2 – N3 – `reference.tip`, we insert a new node N′ between two nodes (e.g. between N1 and N2) from which the new language evolved with a certain length, keeping the original distances between N0…N3 all intact (e.g. the length between N1 and N2 is the same as the sum of the lengths between N1 and N′ and between N′ and N2). Grafting is implemented in a function `graft`[1] which takes as input (i) a tree, (ii) the name of a language to be grafted, (iii) a reference tip from which the location of the new language is defined, (iv) the time between the reference tip and the point in which the new language split off, and (v) the time after the split in which the new language evolved on its own. We estimate the relevant times from the earliest and youngest attestations of the added languages. When grafting onto the MCC summary trees we take the mean of these attestation boundaries. When grafting onto the trees in the posterior samples, we randomly draw from a uniform distribution bounded by these limits to allow for uncertainty:

```r
ie.tree.sum <- graft(ie.tree.sum, 'Middle_English', 'Modern_English', mean(c(516, 866)), 0) %>%
               graft(., 'Middle_Welsh', 'Modern_Welsh', mean(c(566, 866)), 0) %>%
               graft(., 'Middle_Persian', 'Modern_Persian', mean(c(1316, 2216)), 0) %>%
               graft(., 'Luwian', 'Hittite', mean(c(1600, 1000)), mean(c(1800,2400))) %>%
               graft(., 'Middle_High_German', 'Swiss_German', mean(c(516, 966)), 0) %>%
               graft(., 'Middle_Breton', 'Modern_Breton', mean(c(366, 916)), 0) %>%
               graft(., 'Old_Saxon', 'Modern_English', mean(c(1216, 1716)), mean(c(260,560))) %>%
               graft(., 'Old_Lithuanian', 'Modern_Lithuanian', mean(c(316, 516)), 0) %>%
               graft(., 'Pali', 'Nepali', mean(c(2800,3100)), mean(c(400, 1100)))

ie.b.tree.sum <- graft(ie.b.tree.sum, 'Middle_English', 'Modern_English', mean(c(516, 866)), 0) %>%
               graft(., 'Middle_Welsh', 'Modern_Welsh', mean(c(566, 866)), 0) %>%
               graft(., 'Middle_Persian', 'Modern_Persian', mean(c(1316, 2216)), 0) %>%
               graft(., 'Middle_High_German', 'Swiss_German', mean(c(516, 966)), 0) %>%
               graft(., 'Middle_Breton', 'Modern_Breton', mean(c(366, 916)), 0) %>%
               graft(., 'Old_Saxon', 'Modern_English', mean(c(1216, 1716)), mean(c(260,560))) %>%
               graft(., 'Old_Lithuanian', 'Modern_Lithuanian', mean(c(316, 516)), 0) %>%
               graft(., 'Pali', 'Nepali', mean(c(2800,3100)), mean(c(400, 1100)))

ie.trees <- lapply(ie.trees, function(t) {
              graft(t, 'Middle_English', 'Modern_English', runif(1, 516, 866), 0) %>%
              graft(., 'Middle_Welsh', 'Modern_Welsh', runif(1, 566, 866), 0) %>%
              graft(., 'Middle_Persian', 'Modern_Persian', runif(1, 1316, 2216), 0) %>%
              graft(., 'Luwian', 'Hittite', runif(1, 1000, 1600), runif(1, 1800, 2400)) %>%
              graft(., 'Middle_High_German', 'Swiss_German', runif(1, 516, 966), 0) %>%
              graft(., 'Middle_Breton', 'Modern_Breton', runif(1, 366, 916), 0) %>%
              graft(., 'Old_Saxon', 'Modern_English', runif(1, 1216, 1716), runif(1, 260, 560)) %>%
              graft(., 'Old_Lithuanian', 'Modern_Lithuanian', runif(1, 316, 516), 0) %>%
```

[1] available at https://github.com/IVS-UZH/phylo-convert

```
              graft(., 'Pali', 'Nepali', runif(1, 2600,2900), runif(1, 200, 900))
              })

ie.b.trees <- lapply(ie.b.trees, function(t) {
               graft(t, 'Middle_English', 'Modern_English', mean(c(516, 866)), 0) %>%
               graft(., 'Middle_Welsh', 'Modern_Welsh', mean(c(566, 866)), 0) %>%
               graft(., 'Middle_Persian', 'Modern_Persian', mean(c(1316, 2216)), 0) %>%
               graft(., 'Middle_High_German', 'Swiss_German', mean(c(516, 966)), 0) %>%
               graft(., 'Middle_Breton', 'Modern_Breton', mean(c(366, 916)), 0) %>%
               graft(., 'Old_Saxon', 'Modern_English', mean(c(1216, 1716)), mean(c(260,560))) %>%
               graft(., 'Old_Lithuanian', 'Modern_Lithuanian', mean(c(316, 516)), 0) %>%
               graft(., 'Pali', 'Nepali', mean(c(2800,3100)), mean(c(400, 1100)))
              })


# bring the list of trees back into the correct phylo class:
class(ie.trees) <- 'multiPhylo'
attributes(ie.trees)$TipLabel <- ie.trees[[1]]$tip.label
class(ie.b.trees) <- 'multiPhylo'
attributes(ie.b.trees)$TipLabel <- ie.b.trees[[1]]$tip.label
```

Figures 1 and 2 show the resulting summary trees, pruned and grafted so as to exactly match our dataset.[2] Figures 3 and 4 show `DensiTree` representations of the posterior samples (Bouckaert & Heled 2014), giving an impression of the amount and locations of phylogenetic uncertainty. Figures 5 and 6 show the MCC summary tree together with the data. (Figures 3 and 5 are also printed in the main paper.)

[2] Unless noted otherwise, we use the `ggplot2` (Wickham 2009) framework for visualizations, with the `ggtree` (Yu et al. 2016) extension for phylogenies.

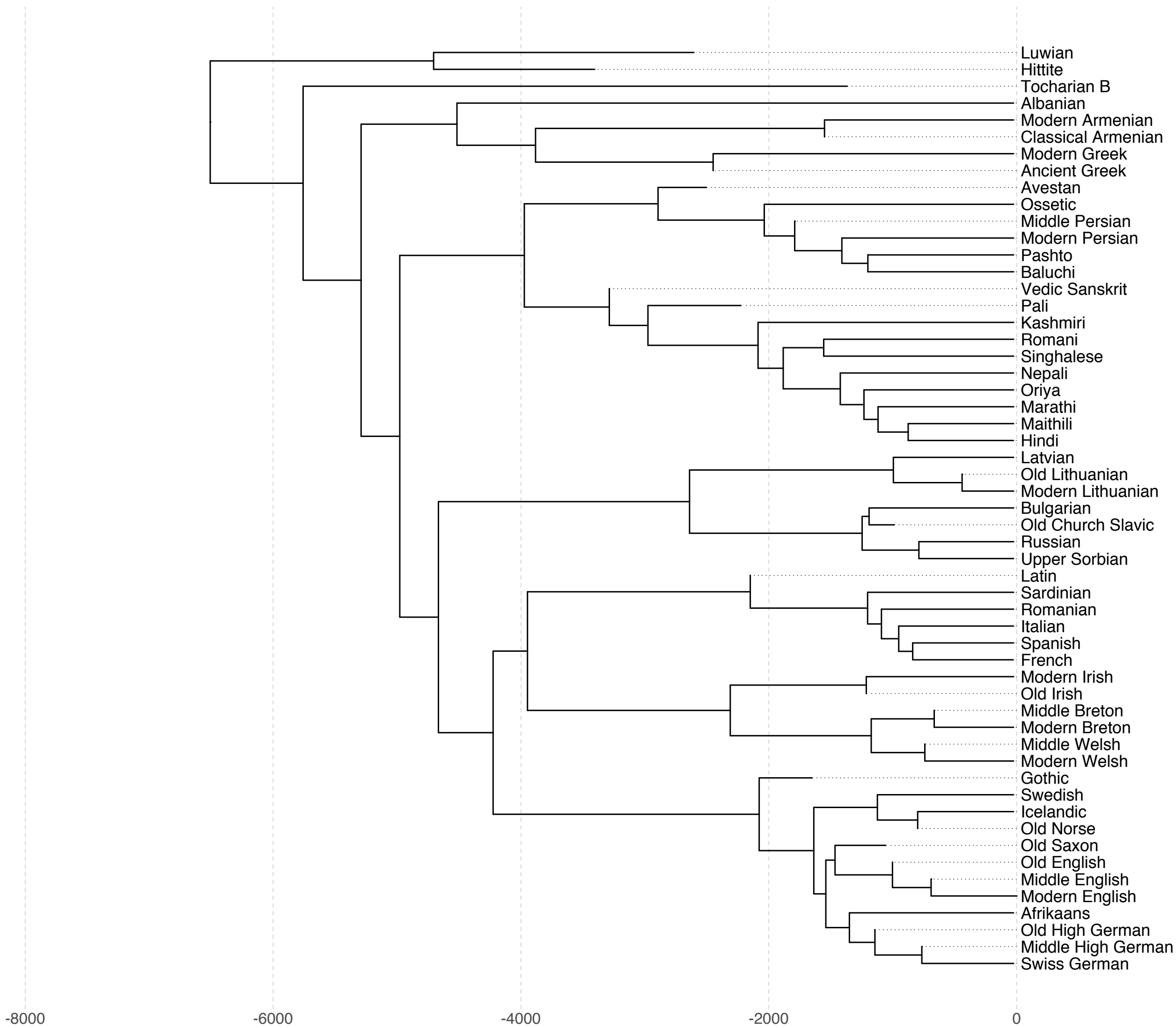

Figure 1: Indo-European Maximum Clade Credibility tree `c1` in Chang et al. (2015), matched to the present dataset

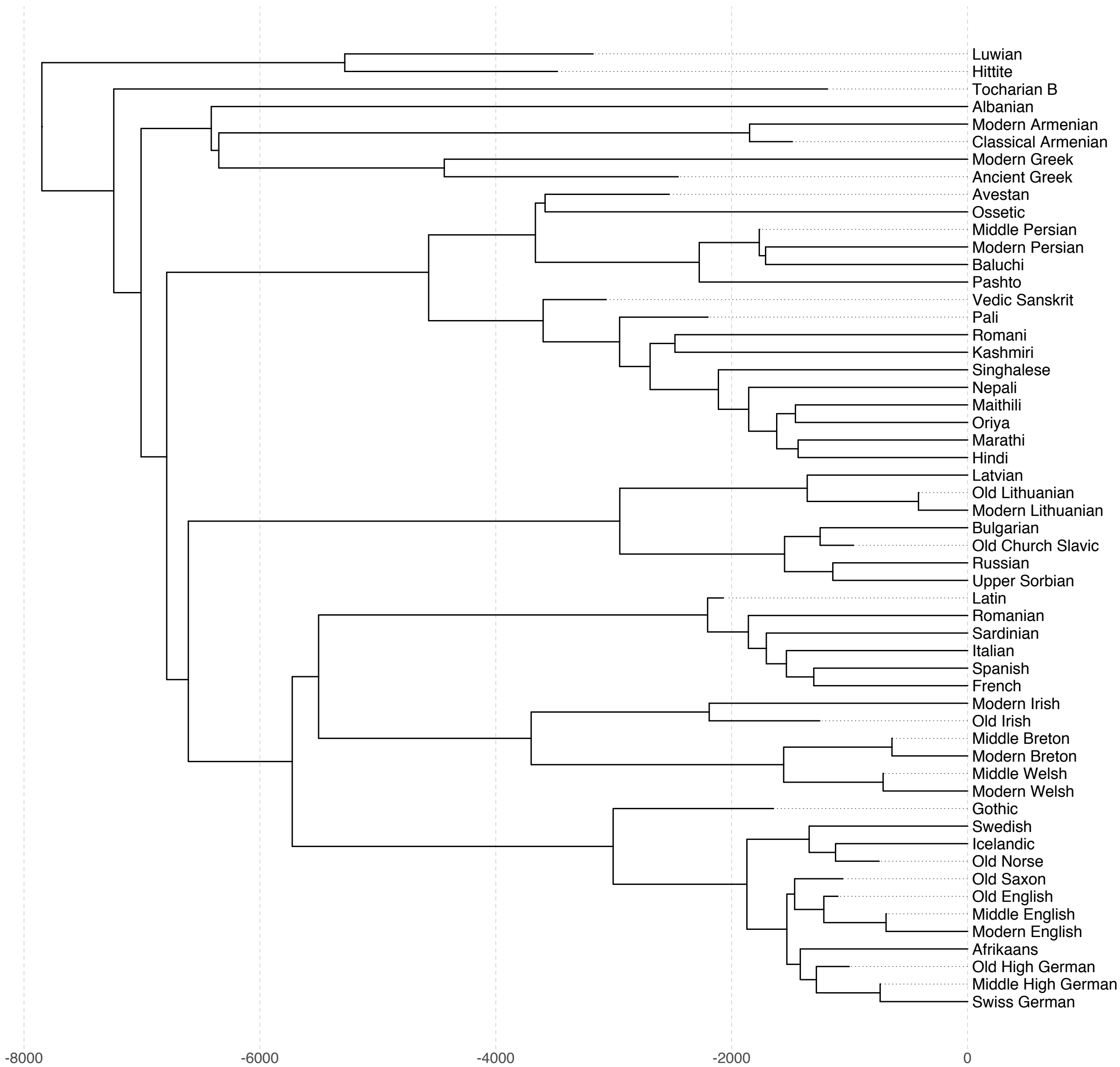

Figure 2: Indo-European Maximum Clade Credibility tree in Bouckaert et al. (2012), matched to the present dataset

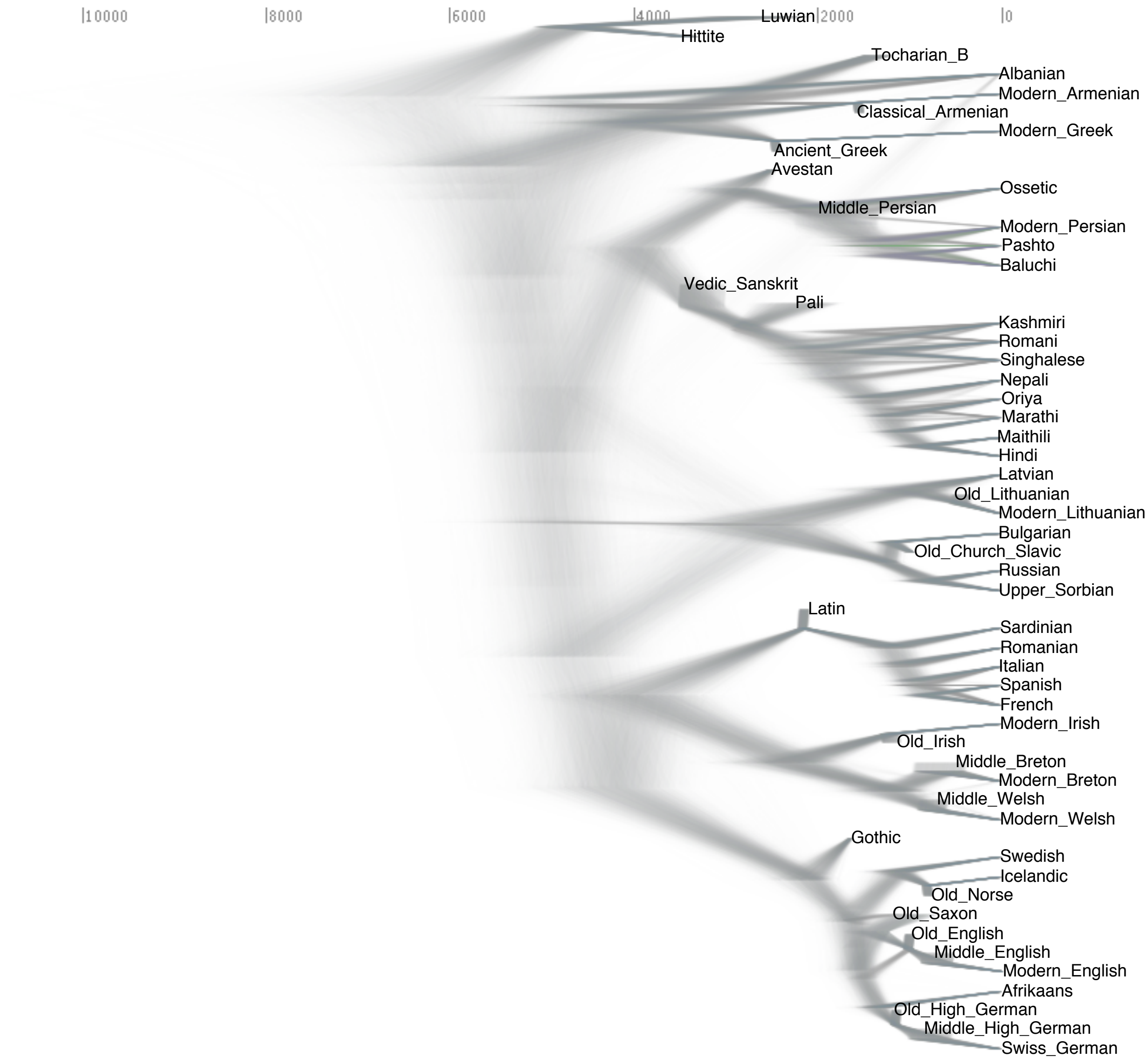

Figure 3: Densitree representation of the posterior tree sample c1 in Chang et al. (2015), matched to the present dataset

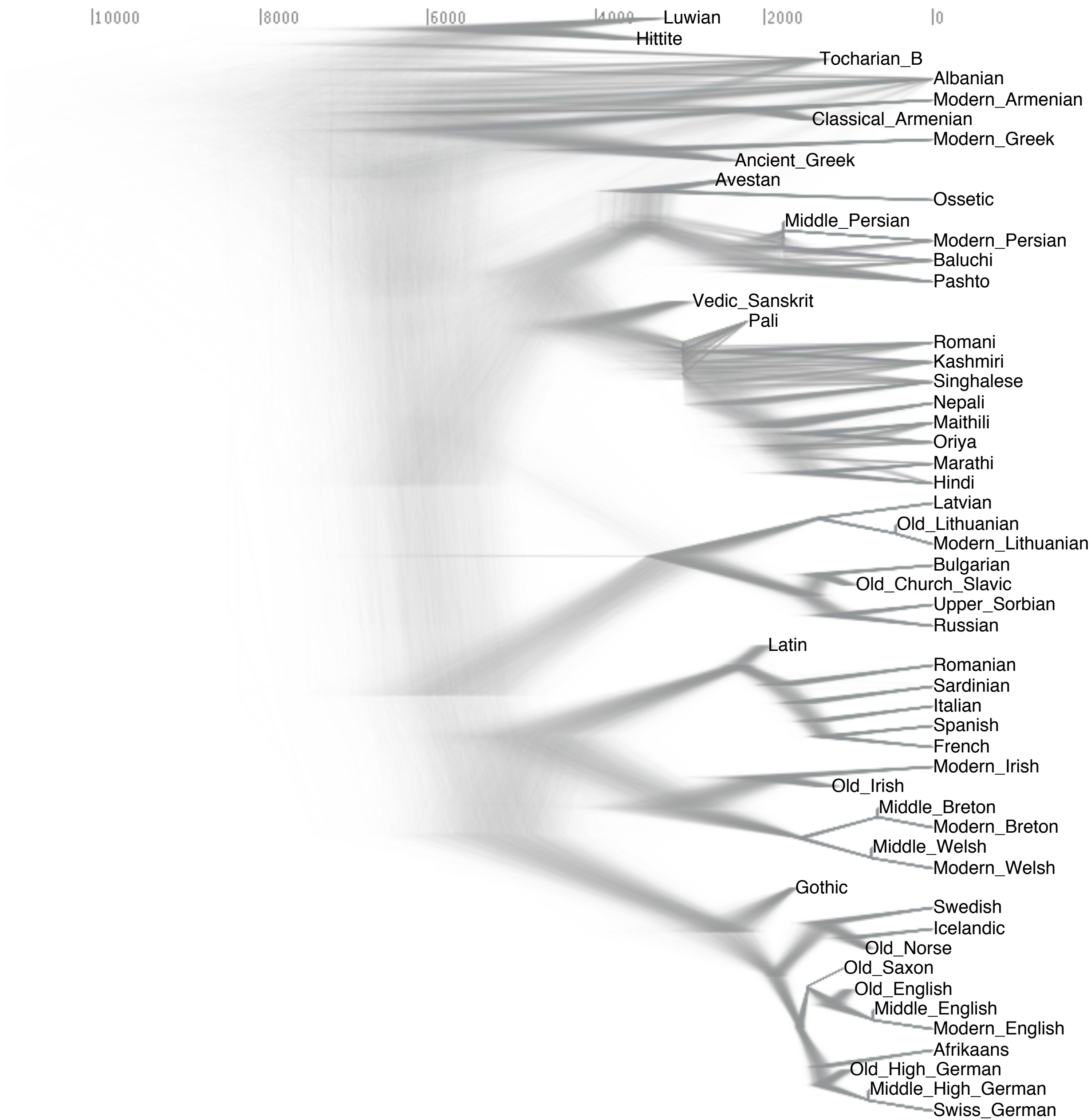

Figure 4: Densitree representation of the posterior tree sample in Bouckaert et al. (2012), matched to the present dataset

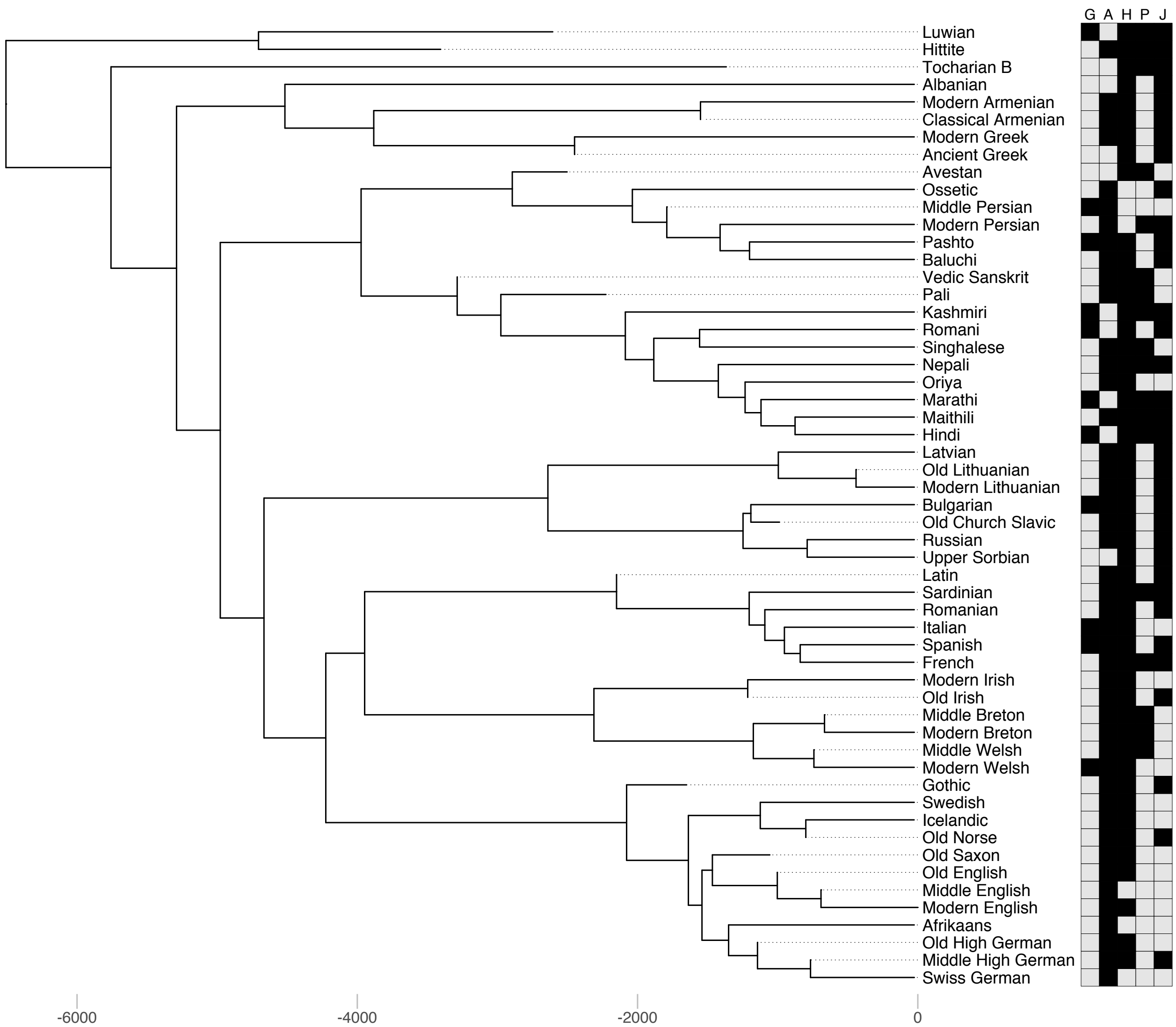

Figure 5: Summary of the analysis in Supporting Material 1, with the phylogeny from Chang et al. (2015). Types of signaling the embedding relation a complex NP: G = genitives, A = adjectivizers, P = adpositions, H = head marking, J = juxtaposition. Grey: type is available for recursion; Black: type is not available

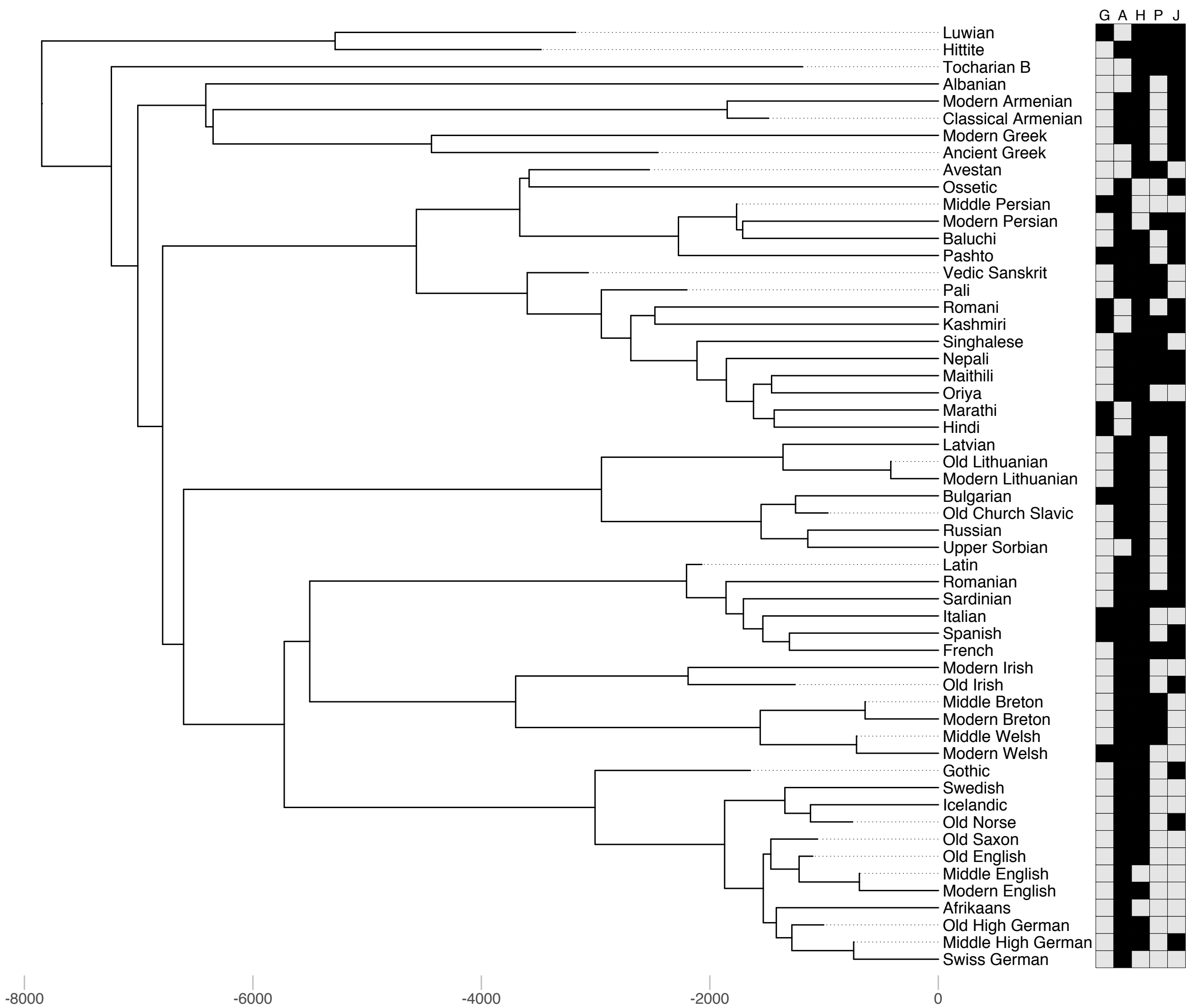

Figure 6: Summary of the analysis in Supporting Material 1, with the phylogeny from Bouckaert et al. (2015). Types of signaling the embedding relation a complex NP: G = genitives, A = adjectivizers, P = adpositions, H = head marking, J = juxtaposition. Grey: type is available for recursion; Black: type is not available

2 Modeling Transition Rates

We basically follow here the same procedure as Dunn et al. (2011) or Bickel et al. (2015). The basic idea is to model language change as Continuous Time Markov Chains with two states:

- Y: yes, the construction is available for recursive embedding
- N: no, the construction is not available for recursive embedding

We then estimate the rates of change q between two states in a Bayesian framework (Pagel 1999; Huelsenbeck et al. 2003; Pagel et al. 2004; Ronquist 2004):

- `qYN`: instantaneous rate of loosing a construction for recursive embedding (Yes>No)
- `qNY`: instantaneous rate of gaining a construction for recursive embedding (No>Yes)

In order to assess whether there is a diachronic bias in either direction, e.g. whether gains are favored over losses, we compare two models for each construction:

- ARD: all rates different, i.e. rates of change towards vs away from recursion are different: qYN≠qNY. This model therefore has two parameters.
- ER: equal rates, i.e. rates of change towards vs away from recursion are the same: qYN=qNY. This model therefore has one parameter.

While for the main analysis we assume constant rates within each tree, we also explore simple covarion models as implemented in `BayesTraits`. These models allow different rates across branches and are arguably more realistic, although it is not clear what variation they react to (perhaps contact effects, perhaps chance). However, in line with what Levinson et al. (2011) report in their study, covarion models gave results that are very similar to constant-rate models. This does not justify the heavily increased computing time required by covarion models.

In each case, the models are compared with Bayes Factors (BF), which are defined as double the difference between the log marginal likelihood of the more complex model, i.e. the ARD model, and the log marginal likelihood of the simpler model, i.e. the ER model. BF > 2 are considered to give "positive evidence" for the better-fitting model; between 5 and 10 "strong evidence" and > 10 "very strong evidence".

Following the recommendations of Pagel & Meade (2013), we replicate all MCMC analyses several times and report the median and the median absolute deviation across these replications. The parameters of each run of the analysis were determined on the basis of visual inspections of trace plots of the likelihoods and (in the case of posterior tree sample) of the tree mixing.

All analyses are based on the posterior sample of trees but we also compared these results to estimates based on the MCC summary tree (commented out in the code below). There was no noticeable difference.

2.1 Functions

We use wrapper scripts which run `BayesTraitsV2` (Pagel & Meade 2013) in the background and allow various steps of processing the output:

```
fit.mcmc <- function(variable, tree=ie.tree.sum, prior="uniform 0 .01", assumed.mrca=NULL,
                iterations=mcmc.iterations, burn.in=mcmc.burn.in,
                sampling.rate=mcmc.sampling.rate, covarion=F, reuse.tree=T) {
    test.ard <- Discrete(tree,
            ie.data[,c('language', variable)], "Bayesian",
```

```
            pa=prior, it=iterations, bi=burn.in, sa=sampling.rate,
            rm=T, fo=assumed.mrca, cv=covarion, reuse.tree=reuse.tree)
    test.er <-  Discrete(tree,
            ie.data[,c('language', variable)], "Bayesian",
            resall='qNY', it=iterations, pa=prior, bi=burn.in, sa=sampling.rate,
            fo=assumed.mrca, cv=covarion, reuse.tree=reuse.tree)

    PIE.er <- tail(test.er,1)[,7:8]
    names(PIE.er)  <- c("Prob.N","Prob.Y")
    PIE.ard <- tail(test.ard,1)[,7:8]
    names(PIE.ard)  <- c("Prob.N","Prob.Y")
    summary.er <- tail(test.er$qNY,1)
    summary.ard <- tail(test.ard[,c("qNY","qYN")],1)
    bfres <- bftest(test.er, test.ard)
    bfres$Better <- ifelse(bfres$BetterModel %in% 'Model 1', 'ER', 'ARD')

    return(list(ARD=test.ard,
            ER=test.er,
            ER.vs.ARD=bfres[,c(1,3)],
            ER.rate=summary.er,
            ARD.rates=summary.ard,
            ARD.PIE=PIE.er,
            ER.PIE=PIE.ard
            ))
}

fit.ml <- function(variable, tree=ie.tree.sum, assumed.mrca=NULL) {
    # make sure we don't use the wrong trees in case they are recycled in the mcmc runs:
    if (file.exists("./BT.current.tree.nex")) { file.remove('BT.current.tree.nex') }
    test.ard <- Discrete(tree, ie.data[,c('language', variable)], "ML",
            mlt=100, fo=assumed.mrca, reuse.tree=F)
    test.er <-  Discrete(tree, ie.data[,c('language', variable)], "ML",
             resall='qNY', mlt=100, fo=assumed.mrca, reuse.tree=F)
    return(list(ER=test.er, ARD=test.ard))
}
```

Function for tracing the MCMC iterations:

```
trace.plots <- function(mcmc.res, model=c('ARD','ER'), which=c('Likelihood', 'Rates', 'Trees')) {
    switch(model,
        ARD = {df <- do.call(rbind, lapply(seq_along(mcmc.res), function(r) {
               x <- mcmc.res[[r]][['ARD']]
               x$replicate <- r
               return(x)
               }))
              },
        ER  = {df <- do.call(rbind, lapply(seq_along(mcmc.res), function(r) {
               x <- mcmc.res[[r]][['ER']]
               x$replicate <- r
               return(x)
               }))
              }
    )
```

```
    switch(which,
    Likelihood = { ggplot(df, aes(x=Iteration,y=Harmonic.Mean)) + geom_line(size=.2) +
                     facet_wrap(~replicate) + theme_bw(base_size = 12) },
    Rates =      { gather(df, rate, value, qNY:qYN) %>%
                             ggplot(aes(x=Iteration, y=value, colour=rate)) +
                             geom_line(size=.2) + theme(legend.position='top') +
                             guides(colour = guide_legend(override.aes = list(size=2))) +
                             facet_wrap(~replicate) + theme_bw(base_size = 12)
                          },
    Trees =     {  #ggplot(df, aes(x=Iteration, y=Tree.No)) + geom_line(size=.2)
                    ggplot(df, aes(x=Tree.No)) + geom_density(size=.2) +
                facet_wrap(~replicate) + theme_bw(base_size = 12) }
    )
}
```

Function for estimating ARD models and then using the resulting transition rates for Stochastic Character Mappping:

```
years.per.change <- function(variable, trees=ie.trees, sum.tree=ie.tree.sum,
                             tree.simulations=nsim, rate.replications=replications,
                             prior='uniform 0 .01', reuse.tree=T) {
    model <- replicate(rate.replications, Discrete(trees, ie.data[,c('language', variable)],
                        mode="Bayesian", pa=prior, it=mcmc.iterations, bi=mcmc.burn.in,
                        sa=mcmc.sampling.rate, cv=F, reuse.tree=reuse.tree), simplify=F)
    median.rates <- apply(sapply(model, function(x) c(tail(x$qNY,1), tail(x$qYN,1))), 1, median)
    q.mat <- matrix(c(-median.rates[1],  median.rates[1],
                       median.rates[2], -median.rates[2]),
                    byrow=T, ncol=2, dimnames=list(c('N','Y'), c('N','Y')))
    states <-ie.data[, variable]
    names(states) <- ie.data$language
    treemaps <- make.simmap(sum.tree, states, Q=q.mat, pi='estimated',
                            nsim=tree.simulations, message=F)
    n.changes <- median(describe.simmap(treemaps, message=F)$count[,'N'])
    sum(ie.tree.sum$edge.length)/n.changes
    }
```

Some helpers for exploring and plotting the results:

```
model.fits <- function(m) { do.call(rbind,
    lapply(m, function(x) { data.frame("["(x,"ER.vs.ARD")) } )
    )}

rates <- function(m, model=c('ARD','ER')) {
    if(model %in% 'ARD') {mod <- paste0(model,".rates")} else {mod <- paste0(model, '.rate')}
    do.call(rbind, lapply(m, function(x) { data.frame("["(x, mod)) } )
    )}

ml.rates <- function(m) {
    x <- do.call(rbind, sapply(m, rbind))
    x$model <- rep(c('ER', 'ARD'), nrow(x)/2)
    return(x[c('model','qNY','qYN')])
    }
```

```r
reconstructions <- function(m, model=c('ARD','ER')) { do.call(rbind,
    lapply(m, function(x) { data.frame("["(x, paste0(model, '.PIE')))}))}

transition.matrix <- function(all.rates, print="list", prob=T, years=100) {
    report.names <- list(c('N','Y'),c('N','Y'))
    x <- apply(all.rates, 1, function(rates) {
        if (length(rates)==1) { rates <- c(rates, rates)
                     names(rates) <- c('ARD.rates.qNY','ARD.rates.qYN') } else {}
        m <- matrix(c(-rates['ARD.rates.qNY'],  rates['ARD.rates.qNY'],
                      rates['ARD.rates.qYN'], -rates['ARD.rates.qYN']),
                    byrow=T, ncol=2, dimnames=report.names)
        if (prob) {
                m <- matrix(matexpo(years*m), byrow=F, ncol=2, dimnames=report.names)
        } else { m }
        if (print %in% 'list') { list(m) } else { m }
    })
    switch(print,
        list = { return(lapply(x, "[[",1)) },
        mean = { return(matrix(c(mean(x[1,]), mean(x[2,]), mean(x[3,]), mean(x[4,])),
             byrow=F, ncol=2,
             dimnames=report.names)) },
        sd = { return(matrix(c(sd(x[1,]), sd(x[2,]), sd(x[3,]), sd(x[4,])),
             byrow=F, ncol=2,
             dimnames=report.names)) },
        median = { return(matrix(c(median(x[1,]), median(x[2,]), median(x[3,]),
                                        median(x[4,])),
             byrow=F, ncol=2,
             dimnames=report.names)) },
        mad = { return(matrix(c(mad(x[1,]), mad(x[2,]), mad(x[3,]), mad(x[4,])),
             byrow=F, ncol=2,
             dimnames=report.names)) }
        )
    }

plot.matrix <- function(p.matrix, label='X', arrow.size=.6, arrow.width=1,
    scale.factor=5, label.pos=.8) {
    graph <- graph_from_adjacency_matrix(adjmatrix=p.matrix, weighted=TRUE, mode="directed")
    plot.igraph(graph, asp=1,
        ylim=c(-2,0),
        margin=0,
        vertex.shape='square',
        vertex.color=c('black','grey'),
        edge.color='black',
        vertex.size=20,
        vertex.label=NA,
        layout=layout_on_grid,
        edge.curved=.4,
        edge.width=E(graph)$weight*scale.factor,
        edge.arrow.size=arrow.size,
        edge.arrow.width=arrow.width,
        edge.loop.angle=-80)
    mtext(label, side=1, line=label.pos)
}
```

2.2 Parameters

Replications:

```
B=10
```

Replications of covarion models:

```
B.cv=3
```

MCMC iterations:

```
mcmc.iterations=1e+07
```

Burn in:

```
mcmc.burn.in=1e+05
```

Sampling rate:

```
mcmc.sampling.rate=1e+03
```

2.3 Priors

We defined the prior distribution for each model so that the estimated number of changes does note exceed one change in about 100 years. Changing back and forth within shorter time ranges is not compatible with anything we know from attested changes and even changing back and forth once within 100 years is a highly remarkable, limiting-case event.

With this constraint in mind, we explored transition rates obtained with priors of $\mathcal{U}(0, 1)$, $\mathcal{U}(0, .1)$ and $\mathcal{U}(0, .01)$ by performing *Stochastic Character Mapping* (see Section 3 below; Nielsen (2002); Huelsenbeck et al. (2003)) and considering the median estimated time for a change to be expected.[3] For this purpose we focus on models allowing for rate differences (ARD). This allows a larger range of rates to explore. We then also compare the rates obtained with these priors with rates estimated from Maximum-Likelihood models.

Parameters:

Stochastic character mappings:

```
nsim=100
```

Replications of rate estimates:

```
replications=5
```

2.3.1 Trees from Chang et al. (2015)

```
types <- c('G','A','H','P','C')
ypc_0_1 <- sapply(types, function(type) years.per.change(type, prior='uniform 0 1'))
ypc_0_.1 <- sapply(types, function(type) years.per.change(type, prior='uniform 0 .1'))
ypc_0_.01 <- sapply(types, function(type) years.per.change(type, prior='uniform 0 .01'))
```

[3] We also explored priors bounded by 10 or 100, but these led to transition rates far above what is plausible.

Median estimated time (in years) for a change to occur per NP type,

given a prior of $\mathcal{U}(0,1)$:

```
ypc_0_1
```

```
##        G        A        H        P        C
## 2.662322 5.996763 9.682158 1.834052 1.597575
```

given a prior of $\mathcal{U}(0,.1)$:

```
ypc_0_.1
```

```
##        G        A        H        P        C
## 40.58483 36.67056 76.95446 21.24735 14.43978
```

given a prior of $\mathcal{U}(0,.01)$:

```
ypc_0_.01
```

```
##        G        A        H        P        C
## 323.4672 469.8602 468.5903 200.9020 215.3769
```

2.3.2 Trees from Bouckaert et al. (2012)

```
# first, make sure we don't inherit the trees that were used in the loops above:
if(file.exists('BT.current.tree.nex')) { file.remove('BT.current.tree.nex') }
ypc_0_1b <- sapply(types, function(type) years.per.change(type, prior='uniform 0 1',
                trees=ie.b.trees, sum.tree=ie.b.tree.sum))
ypc_0_.1b <- sapply(types, function(type) years.per.change(type, prior='uniform 0 .1',
                trees=ie.b.trees, sum.tree=ie.b.tree.sum))
ypc_0_.01b <- sapply(types, function(type) years.per.change(type, prior='uniform 0 .01',
                trees=ie.b.trees, sum.tree=ie.b.tree.sum))
```

Median estimated time (in years) for a change to occur per NP type,

given a prior of $\mathcal{U}(0,1)$:

```
ypc_0_1b
```

```
##        G        A        H        P        C
## 4.919373 2.536218 4.633061 1.632319 1.799072
```

given a prior of $\mathcal{U}(0,.1)$:

```
ypc_0_.1b
```

```
##        G        A        H        P        C
## 19.04629 20.90660 32.85549 13.65507 10.20173
```

given a prior of $\mathcal{U}(0,.01)$:

```
ypc_0_.01b
```

```
##         G         A         H         P         C
## 207.88777 228.73140 378.55545 256.85689  98.73485
```

2.3.3 Maximum Likelihood estimates based on trees from Chang et al. (2015)

```
g.ml <- replicate(B, fit.ml('G', tree=ie.tree.sum), simplify=F)
ml.rates(g.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.001298 0.000363
## 2    ER 0.000246 0.000246
```

```
a.ml <- replicate(B, fit.ml('A'), simplify=F)
ml.rates(a.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY       qYN
##   <chr>    <dbl>     <dbl>
## 1   ARD 0.000203 0.0006555
## 2    ER 0.000181 0.0001810
```

```
h.ml <- replicate(B, fit.ml('H'), simplify=F)
ml.rates(h.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY     qYN
##   <chr>    <dbl>   <dbl>
## 1   ARD 0.198818 1.62368
## 2    ER 0.052750 0.05275
```

```
p.ml <- replicate(B, fit.ml('P'), simplify=F)
ml.rates(p.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.000297 0.000142
## 2    ER 0.000186 0.000186
```

```
c.ml <- replicate(B, fit.ml('C'), simplify=F)
ml.rates(c.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.000339 0.000747
## 2    ER 0.000508 0.000508
```

2.3.4 Maximum Likelihood estimates based on trees from Bouckaert et al. (2012)

```
g.b.ml <- replicate(B, fit.ml('G', tree=ie.b.tree.sum), simplify=F)
ml.rates(g.b.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.001181 0.000325
## 2    ER 0.000142 0.000142
```

```
a.b.ml <- replicate(B, fit.ml('A', tree=ie.b.tree.sum), simplify=F)
ml.rates(a.b.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.000133 0.000416
## 2    ER 0.000101 0.000101
```

```
h.b.ml <- replicate(B, fit.ml('H', tree=ie.b.tree.sum), simplify=F)
ml.rates(h.b.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.001448 0.011820
## 2    ER 0.000072 0.000072
```

```
p.b.ml <- replicate(B, fit.ml('P', tree=ie.b.tree.sum), simplify=F)
ml.rates(p.b.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.000247 0.000107
## 2    ER 0.000144 0.000144
```

```
c.b.ml <- replicate(B, fit.ml('C', tree=ie.b.tree.sum), simplify=F)
ml.rates(c.b.ml) %>% group_by(model) %>% summarise_each(funs(median))
```

```
## # A tibble: 2 × 3
##   model      qNY      qYN
##   <chr>    <dbl>    <dbl>
## 1   ARD 0.000199 0.000497
## 2    ER 0.000383 0.000383
```

2.3.5 Conclusions

The results from Stochastic Character Mapping are largely consistent with those of Maximum Likelihood models and a prior of $\mathcal{U}(0, .01)$ seems to be a realistic choice. The ML estimates of the head marking type are exceptional, especially in the C trees. The estimates here suggest an unrealistically high rate (beyond 1 in the C and beyond .01 in the B trees). Since the Stochastic Character mappings suggest that any rates beyond .1 are unrealistic, we chose a prior of $\mathcal{U}(0, .05)$ as compromise range for the head marking type for both trees.

2.4 Individual analyses based on trees from Chang et al. (2015)

2.4.1 Genitives

```
# g.mcmc.sum <- replicate(B, fit.mcmc('G', tree=ie.tree.sum, prior="uniform 0 .01"), simplify=F)
g.mcmc.cv <- replicate(B.cv, fit.mcmc('G', tree=ie.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
g.mcmc <- replicate(B, fit.mcmc('G', tree=ie.trees, prior="uniform 0 .01"), simplify=F)
```

```
# trace.plots(g.mcmc, model='ARD', which='Trees')
# trace.plots(g.mcmc, model='ARD', which='Likelihood')
# trace.plots(g.mcmc.cv, model='ARD', which='Trees')
# trace.plots(g.mcmc.cv, model='ARD', which='Likelihood')

# model.fits(g.mcmc.sum)
```

Covarion models:

```
model.fits(g.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  11.40848          ARD
```

Simple models:

```
model.fits(g.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                 better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  13.21787          ARD
```

```
transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=F)
```

```
##            N          Y
## N -0.0057305  0.0057305
## Y  0.0014365 -0.0014365
```

```r
transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.5847526 0.4152474
## Y 0.1027190 0.8972810
```

```r
transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.2273411 0.7726589
## Y 0.1944665 0.8055335
```

```r
transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.2085450 0.7914550
## Y 0.2080712 0.7919288
```

2.4.2 Adjectivizers

```r
# a.mcmc.sum <- replicate(B, fit.mcmc('A', tree=ie.tree.sum, prior="uniform 0 .01"), simplify=F)
a.mcmc.cv <- replicate(B.cv, fit.mcmc('A', tree=ie.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
a.mcmc <- replicate(B, fit.mcmc('A', tree=ie.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(a.mcmc, model='ARD', which='Trees')
# trace.plots(a.mcmc, model='ARD', which='Likelihood')
# trace.plots(a.mcmc.cv, model='ARD', which='Trees')
# trace.plots(a.mcmc.cv, model='ARD', which='Likelihood')

# model.fits(a.mcmc.sum)
```

Covarion models:

```r
model.fits(a.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1   6.68102          ARD
```

Simple models:

```r
model.fits(a.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1   7.03483          ARD
```

```r
transition.matrix(rates(a.mcmc, 'ARD'), 'median', prob=F)
```

```
##           N         Y
## N -0.001527  0.001527
## Y  0.006376 -0.006376
```

```r
transition.matrix(rates(a.mcmc, 'ARD'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.8999889 0.1000111
## Y 0.4355788 0.5644212
```

```r
transition.matrix(rates(a.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8009731 0.1990269
## Y 0.7406869 0.2593131
```

```r
transition.matrix(rates(a.mcmc, 'ARD'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.7815324 0.2184676
## Y 0.7725696 0.2274304
```

```r
transition.matrix(rates(a.mcmc, 'ARD'), 'median', prob=T, years=3000)
```

```
##           N         Y
## N 0.7789333 0.2210667
## Y 0.7789272 0.2210728
```

2.4.3 Head marking

```r
# h.mcmc.sum <- replicate(B, fit.mcmc('H', tree=ie.tree.sum, prior="uniform 0 .01"), simplify=F)
h.mcmc.cv <- replicate(B.cv, fit.mcmc('H', tree=ie.trees, prior="uniform 0 .05", covarion=T),
             simplify=F)
h.mcmc <- replicate(B, fit.mcmc('H', tree=ie.trees, prior="uniform 0 .05"), simplify=F)
```

```r
# trace.plots(h.mcmc, model='ARD', which='Trees')
# trace.plots(h.mcmc, model='ARD', which='Likelihood')
# trace.plots(h.mcmc.cv, model='ARD', which='Trees')
# trace.plots(h.mcmc.cv, model='ARD', which='Likelihood')

# model.fits(h.mcmc.sum)
```

Covarion models:

```r
model.fits(h.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  20.42433          ARD
```

Simple models:

```r
model.fits(h.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  19.57874          ARD
```

```r
transition.matrix(rates(h.mcmc, 'ARD'), 'median', prob=F)
```

```
##           N         Y
## N -0.004451  0.004451
## Y  0.023292 -0.023292
```

```r
transition.matrix(rates(h.mcmc, 'ARD'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.8628250 0.1371750
## Y 0.7746346 0.2253654
```

```r
transition.matrix(rates(h.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8410713 0.1589287
## Y 0.8409155 0.1590845
```

2.4.4 Adpositions

```r
# p.mcmc.sum <- replicate(B, fit.mcmc('P', tree=ie.tree.sum, prior="uniform 0 .01"), simplify=F)
p.mcmc.cv <- replicate(B.cv, fit.mcmc('P', tree=ie.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
p.mcmc <- replicate(B, fit.mcmc('P', tree=ie.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(p.mcmc.sum, model='ER', which='Likelihood')
# trace.plots(p.mcmc, model='ER', which='Trees')
# trace.plots(p.mcmc, model='ER', which='Likelihood')

# model.fits(p.mcmc.sum)
```

Covarion models:

```r
model.fits(p.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  2.607434           ER
```

Simple models:

```r
model.fits(p.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  3.526145           ER
```

```r
transition.matrix(rates(p.mcmc, 'ER'), 'median', prob=F)
```

```
##           N         Y
## N -0.000212  0.000212
## Y  0.000212 -0.000212
```

```r
transition.matrix(rates(p.mcmc, 'ER'), 'median', prob=T, years=100)
```

```
##            N          Y
## N 0.97924324 0.02075676
## Y 0.02075676 0.97924324
```

```r
transition.matrix(rates(p.mcmc, 'ER'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.8272178 0.1727822
## Y 0.1727822 0.8272178
```

```r
transition.matrix(rates(p.mcmc, 'ER'), 'median', prob=T, years=6000)
```

```
##           N         Y
## N 0.5393012 0.4606988
## Y 0.4606988 0.5393012
```

2.4.5 Juxtaposition

```r
# c.mcmc.sum <- replicate(B, fit.mcmc('C', tree=ie.tree.sum, prior="uniform 0 .01"), simplify=F)
c.mcmc.cv <- replicate(B.cv, fit.mcmc('C', tree=ie.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
c.mcmc <- replicate(B, fit.mcmc('C', tree=ie.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(c.mcmc.sum, model='ARD', which='Likelihood')
# trace.plots(c.mcmc, model='ARD', which='Trees')
# trace.plots(c.mcmc, model='ER', which='Likelihood')

# model.fits(c.mcmc.sum)
```

Covarion models:

```r
model.fits(c.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  1.137996          ARD
```

Simple models:

```r
model.fits(c.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  0.810415          ARD
```

```r
mad(model.fits(c.mcmc)[,1])
```

```
## [1] 0.130881
```

```r
transition.matrix(rates(c.mcmc, 'ER'), 'median', prob=F)
```

```
##            N          Y
## N -0.0055545  0.0055545
## Y  0.0055545 -0.0055545
```

```r
transition.matrix(rates(c.mcmc, 'ER'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.6647098 0.3352902
## Y 0.3352902 0.6647098
```

```r
transition.matrix(rates(c.mcmc, 'ER'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.5000078 0.4999922
## Y 0.4999922 0.5000078
```

2.5 Individual analyses based on trees from Bouckaert et al. (2012)

First, we make sure we don't inherit the trees that we used in the loops above:

```r
if(file.exists('BT.current.tree.nex')) { file.remove('BT.current.tree.nex') }
```

2.5.1 Genitives

```r
g.b.mcmc.cv <- replicate(B.cv, fit.mcmc('G', tree=ie.b.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
g.b.mcmc <- replicate(B, fit.mcmc('G', tree=ie.b.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(g.b.mcmc, model='ARD', which='Trees')
# trace.plots(g.b.mcmc, model='ARD', which='Likelihood')
# trace.plots(g.b.mcmc.cv, model='ARD', which='Trees')
# trace.plots(g.b.mcmc.cv, model='ARD', which='Likelihood')
```

Covarion models:

```r
model.fits(g.b.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                       better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  11.42114          ARD
```

Simple models:

```r
model.fits(g.b.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  13.30115          ARD
```

```r
transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=F)
```

```
##           N         Y
## N -0.007961  0.007961
## Y  0.002643 -0.002643
```

```r
transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.5099191 0.4900809
## Y 0.1575147 0.8424853
```

```r
transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.2427598 0.7572402
## Y 0.2322844 0.7677156
```

```r
transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.2376483 0.7623517
## Y 0.2329996 0.7670004
```

2.5.2 Adjectivizers

```r
a.b.mcmc.cv <- replicate(B.cv, fit.mcmc('A', tree=ie.b.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
a.b.mcmc <- replicate(B, fit.mcmc('A', tree=ie.b.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(a.b.mcmc, model='ARD', which='Trees')
# trace.plots(a.b.mcmc, model='ARD', which='Likelihood')
# trace.plots(a.b.mcmc.cv, model='ARD', which='Trees')
# trace.plots(a.b.mcmc.cv, model='ARD', which='Likelihood')
```

Covarion models:

```r
model.fits(a.b.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  4.618912          ARD
```

Simple models:

```r
model.fits(a.b.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                    better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  5.357128          ARD
```

```r
transition.matrix(rates(a.b.mcmc, 'ARD'), 'median', prob=F)
```

```
##            N          Y
## N -0.0018135  0.0018135
## Y  0.0078790 -0.0078790
```

```r
transition.matrix(rates(a.b.mcmc, 'ARD'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.8796127 0.1203873
## Y 0.5015512 0.4984488
```

```r
transition.matrix(rates(a.b.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8154268 0.1845732
## Y 0.7819336 0.2180664
```

```r
transition.matrix(rates(a.b.mcmc, 'ARD'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.7881179 0.2118821
## Y 0.7881120 0.2118880
```

```r
transition.matrix(rates(a.b.mcmc, 'ARD'), 'median', prob=T, years=3000)
```

```
##           N         Y
## N 0.7881167 0.2118833
## Y 0.7881167 0.2118833
```

2.5.3 Head marking

```r
h.b.mcmc.cv <- replicate(B.cv, fit.mcmc('H', tree=ie.b.trees, prior="uniform 0 .05", covarion=T),
             simplify=F)
h.b.mcmc <- replicate(B, fit.mcmc('H', tree=ie.b.trees, prior="uniform 0 .05"), simplify=F)
```

```r
# trace.plots(h.b.mcmc, model='ARD', which='Trees')
# trace.plots(h.b.mcmc, model='ARD', which='Likelihood')
# trace.plots(h.b.mcmc.cv, model='ARD', which='Trees')
# trace.plots(h.b.mcmc.cv, model='ARD', which='Likelihood')
```

Covarion models:

```r
model.fits(h.b.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                      better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  22.96714          ARD
```

Simple models:

```r
model.fits(h.b.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                   better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  19.18678          ARD
```

```r
transition.matrix(rates(h.b.mcmc, 'ARD'), 'median', prob=F)
```

```
##             N          Y
## N -0.0067385  0.0067385
## Y  0.0377760 -0.0377760
```

```r
transition.matrix(rates(h.b.mcmc, 'ARD'), 'median', prob=T, years=100)
```

```
##            N         Y
## N 0.8495886 0.1504114
## Y 0.8342935 0.1657065
```

```r
transition.matrix(rates(h.b.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##            N         Y
## N 0.8481452 0.1518548
## Y 0.8481452 0.1518548
```

2.5.4 Adpositions

```r
p.b.mcmc.cv <- replicate(B.cv, fit.mcmc('P', tree=ie.b.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
p.b.mcmc <- replicate(B, fit.mcmc('P', tree=ie.b.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(p.b.mcmc, model='ARD', which='Trees')
# trace.plots(p.b.mcmc, model='ARD', which='Likelihood')
# trace.plots(p.b.mcmc.cv, model='ARD', which='Trees')
# trace.plots(p.b.mcmc.cv, model='ARD', which='Likelihood')
```

Covarion models:

```r
model.fits(p.b.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                      better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1   2.25831           ER
```

Simple models:

```r
model.fits(p.b.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                      better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  2.963341           ER
```

```r
transition.matrix(rates(p.b.mcmc, 'ER'), 'median', prob=F)
```

```
##           N         Y
## N -0.000169  0.000169
## Y  0.000169 -0.000169
```

```r
transition.matrix(rates(p.b.mcmc, 'ER'), 'median', prob=T, years=100)
```

```
##            N          Y
## N 0.98338339 0.01661661
## Y 0.01661661 0.98338339
```

```r
transition.matrix(rates(p.b.mcmc, 'ER'), 'median', prob=T, years=1000)
```

```
##          N        Y
## N 0.856669 0.143331
## Y 0.143331 0.856669
```

```r
transition.matrix(rates(p.b.mcmc, 'ER'), 'median', prob=T, years=6000)
```

```
##           N         Y
## N 0.5662736 0.4337264
## Y 0.4337264 0.5662736
```

2.5.5 Juxtaposition

```r
c.b.mcmc.cv <- replicate(B.cv, fit.mcmc('C', tree=ie.b.trees, prior="uniform 0 .01", covarion=T),
             simplify=F)
c.b.mcmc <- replicate(B, fit.mcmc('C', tree=ie.b.trees, prior="uniform 0 .01"), simplify=F)
```

```r
# trace.plots(c.b.mcmc, model='ARD', which='Trees')
# trace.plots(c.b.mcmc, model='ARD', which='Likelihood')
# trace.plots(c.b.mcmc.cv, model='ARD', which='Trees')
# trace.plots(c.b.mcmc.cv, model='ARD', which='Likelihood')
```

Covarion models:

```r
model.fits(c.b.mcmc.cv) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                      better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  0.270724          ARD
```

Simple models:

```r
model.fits(c.b.mcmc) %>% summarize(median.BF=median(ER.vs.ARD.BayesFactor),
                                   better.model=paste(unique(ER.vs.ARD.Better), collapse=', '))
```

```
##   median.BF better.model
## 1  0.623957      ARD, ER
```

```r
mad(model.fits(c.b.mcmc)[,1])
```

```
## [1] 0.3287028
```

```r
transition.matrix(rates(c.b.mcmc, 'ER'), 'median', prob=F)
```

```
##           N         Y
## N -0.004057  0.004057
## Y  0.004057 -0.004057
```

```r
transition.matrix(rates(c.b.mcmc, 'ER'), 'median', prob=T, years=100)
```

```
##           N         Y
## N 0.7221273 0.2778727
## Y 0.2778727 0.7221273
```

```r
transition.matrix(rates(c.b.mcmc, 'ER'), 'median', prob=T, years=1000)
```

```
##           N         Y
## N 0.5001503 0.4998497
## Y 0.4998497 0.5001503
```

2.6 Summary of results

Table 1 summarizes the models across types and trees. C models are fitted on Chang et al.'s (2015) trees; B models are fitted on Bouckaert et al.'s (2012) trees. Columns: 'best' reports the best fitting model, 'median' the median BF, 'mad' the median absolute deviation of the BF across replications, and 'bias' the states towards which the transition rates are highest in the best-fitting model.

Table 1: Summary of models across types.

Type	C: best	C: median	C: mad	C: bias	B: best	B: median	B: mad	B: bias
G	ARD	13.22	0.18	Y	ARD	13.30	0.41	Y
A	ARD	7.03	0.49	N	ARD	5.36	0.62	N
H	ARD	19.58	2.25	N	ARD	19.19	2.46	N
P	ER	3.53	0.39	N	ER	2.96	0.38	N
C	ARD	0.81	0.13	N	ARD	0.62	0.33	N

This suggests that for **genitives**, **adjectivizers** and **head marking**, there is evidence for an evolutionary bias, although the evidence is slightly weaker for adjectivizers modeled on Bouckaert et al's (2012) trees. The following reports transition probabilities of the three biased types after 500 years (cf. Figure 7). Not that only genitives are preferred; adjectivizers and head marking is dispreferred.

Genitives:

```
transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.2273411 0.7726589
## Y 0.1944665 0.8055335
```

```
transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.2427598 0.7572402
## Y 0.2322844 0.7677156
```

Adjectivizers:

```
transition.matrix(rates(a.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8009731 0.1990269
## Y 0.7406869 0.2593131
```

```
transition.matrix(rates(a.b.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8154268 0.1845732
## Y 0.7819336 0.2180664
```

Head marking:

```
transition.matrix(rates(h.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8410713 0.1589287
## Y 0.8409155 0.1590845
```

```
transition.matrix(rates(h.b.mcmc, 'ARD'), 'median', prob=T, years=500)
```

```
##           N         Y
## N 0.8481452 0.1518548
## Y 0.8481452 0.1518548
```

Again, the results differ only minimally across trees.

Genitives reach a stationary distribution after this many years when modeled on Chang et al.'s trees:

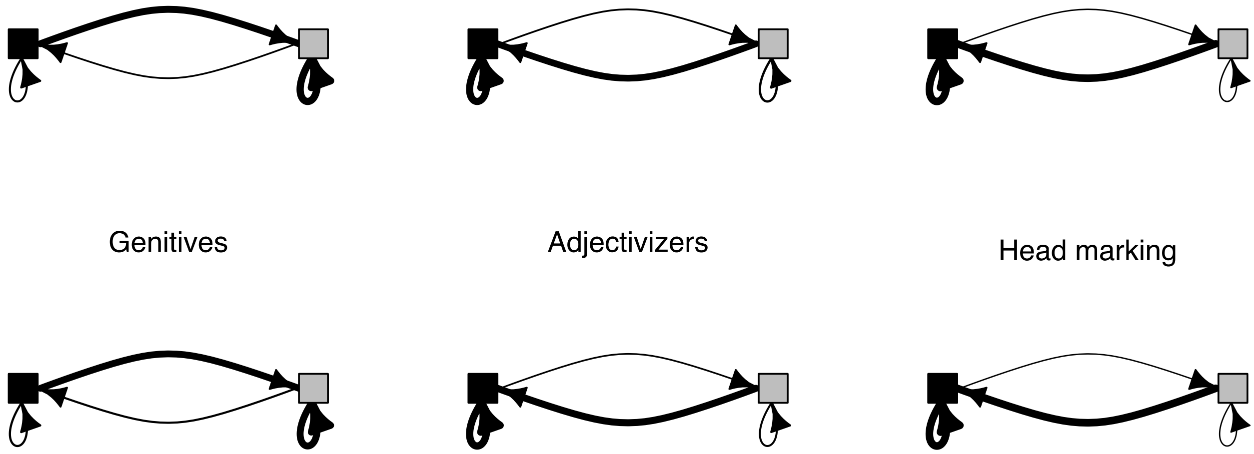

Figure 7: Transition probabilities after 500 years (top row based on Chang et al.'s trees, bottom row based on Bouckaert et al's trees). Black = not available for recursion; Grey = available for recursion.

```
(stationary <- round(optimize(f=function(y) {
                         m <- transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=T, years=y)
                         m[1,1]-m[2,1]
                         },
               interval = c(100,6000), tol=100)$minimum)
)
```

```
## [1] 5661
```

and after this many years when modeled on Bouckaert et al.'s trees:

```
(stationaryb <- round(optimize(f=function(y) {
                         m <- transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=T, years=y)
                         m[1,1]-m[2,1]
                         },
               interval = c(100,6000), tol=100)$minimum)
)
```

```
## [1] 4012
```

Thus, regardless of the trees we assume, the stationary distribution has been reached and looks like this when modeled on Chang et al.'s trees

```
round(transition.matrix(rates(g.mcmc, 'ARD'), 'median', prob=T, years=stationary),2)
```

```
##      N    Y
## N 0.21 0.79
## Y 0.21 0.79
```

and like this when modeled on Bouckaert et al.'s trees

```r
round(transition.matrix(rates(g.b.mcmc, 'ARD'), 'median', prob=T, years=stationary),2)
```

```
##      N    Y
## N 0.23 0.77
## Y 0.23 0.77
```

Comparing the stationary distribution with the synchronic distribution suggests that the synchronic distribution estimates the trend towards genitives reliably in Chang et al.'s trees but slighlty *overestimates* the trend in Bouckaert et al's trees:

```r
round(prop.table(table(ie.data$G)),2)
```

```
##
##   N   Y
## 0.2 0.8
```

The trend would be *underestimated* in both trees if we sampled only extant languages:

```r
round(prop.table(table(subset(ie.data, status %in% 'alive')$G)),2)
```

```
##
##    N    Y
## 0.28 0.72
```

3 Stochastic Character Mapping

Our hypothesis predicts that at any given stage in the history of Indo-European, the probability of at least one NP type being available for recursion is higher than the probability of no type being available for recursion. To test this prediction, we estimate the histories of each type on the tree(s) via *Stochasting Character Mapping* (Nielsen 2002; Huelsenbeck et al. 2003). A stochastic character map assigns states (`Y` = type is available for recursion vs. `N` = not available) to time intervals on each branch of a tree by simulation from a transition matrix. Thus, for example, a branch is in state `Y` for 100 years, then switches to `N` and stays so for 1000 years, then switches back again etc. There are many different such maps that are compatible with a given transition matrix, and so one needs to sample them.

To do so, we use the best-fitting matrices from above and sample B maps from their Bayesian posterior probability distribution, using `phytools::make.simmap` (Revell 2012). We first generate the maps separately for each type. Within each type, we aggregate the sample of maps in form of their posterior probability density over time based on binning the branches into time intervals. Each of the time intervals captures a *diachronic trial* in which we can assess the posterior probability of a type being available or not. Estimating these trials is implemented in the wrapper function `compute.simmap` below, which takes as input an NP type, a tree, a parameter for the resolution of the binning procedure and a parameter for the number of maps to be simulated. The actual work is performed with the algorithm in `phytools::densityMap` (Revell 2013).

In a second step, we combine the probabilities per time interval from the five NP types and compute the probability that there is at least one type available at a given interval. The probability that at least one type T_i is available is the complement of the probability that all types are missing:

$$P(\exists T_i) = 1 - P\left(\bigcap_{i=1}^{n} \bar{T_i}\right) = 1 - \prod_{i=1}^{n} [1 - P(T_i)]$$

Our hypothesis is supported to the extent that $P(\exists T_i)$ is significantly higher than .5 across all time intervals (diachronic trials).

We perform these analysis on the posterior tree sample (`ie.trees` and `ie.b.trees`) in order to allow for phylogenetic uncertainty. However, for visualizing the maps on a tree, we also perform analyses on the Maximum Clade Credibility (MCC) summary tree (`ie.tree.sum` and `ie.b.tree.sum`).

A full analysis on the posterior tree sample is computationally extremely expensive (B simulations $\times$ 6 types $\times$ 20,001 trees). In response to this, we analyze only a random sub-sample of trees. Trying different such sub-samples never changed our results.

3.1 Functions

Wrapper function for estimating the posterior probability density over time, binned into time intervals. The stochastic mapping is always based on the transition matrix derived from the posterior tree sample (contained in the `.*mcmc` models from above):

```
compute.simmap <- function(variable, tree=ie.tree.sum, model=c('ER','ARD'),
    iterations=B, resolution=res) {
    # first make sure that we retrieve the right model (ie.tree vs ie.b.tree):
    if(grepl('b', deparse(substitute(tree)))) {tree.type='.b'} else {tree.type=''}
    mcmc.results <- get(paste0(tolower(variable), tree.type, '.mcmc'))
    Q.matrix <- transition.matrix(rates(mcmc.results, model), 'median', prob=F)
    states <-ie.data[, variable]
    names(states) <- ie.data$language
    treemaps <- make.simmap(tree, states, Q=Q.matrix, pi='estimated', nsim=iterations)
    densityMap(treemaps, plot=F, res=resolution)
    }
```

Function for visualizing the posterior probability density over time:

```
plot.it <- function(density.map) {
        density.map$tree <- ladderize.simmap(density.map$tree)
        plot(density.map, outline=T, lwd=c(6,6),
            fsize=c(1,1), ftype='reg',
                legend=2000, leg.txt=c("0",paste0("PostProb(Y)"),"1"))
        # nodelabels()
             }
```

Functions for combining the posterior probabilities per time intervals from all types:

1. Compute the probability of at least one type being available:

    ```
    add.prob <- function(v) {
    v.df <- as.data.frame(v)
    apply(v.df, 1, function(p) 1 - prod(1-p))
    }
    ```

2. Compute this probability within the sample of maps in a tree. The probabilities per time interval are stored as vectors associated with each branch (edge) of the tree, which are in turn stored in a list of branches. To combine these vectors within a tree one by one, we need to extract them elementwise from the list of branches and apply the `add.prob` function:

```r
compute.cum.prob <- function(X) {
  # get the length of the longest vector
  n <- max(sapply(X, length))
  lapply(1:n, function(i) {
    # extract the i-th element from each vector
    lists <- lapply(X, function(list) list[[1+ ((i-1) %% length(list))]])
    # and apply the function of interest elementwise
    add.prob(lists)
  })
}
```

3. This procedure of combining probabilities elementwise in each branch can now be applied to the five NP types, stored as a list that combines the result from `compute.simmap` from each type. `compute.simmap` generates a list containing (among other things) the tree on which the histories are mapped (object `tree`) and as sublist of this, the probabilities per time interval on each branch (list of branches in object `tree$maps`). The probabilities are stored as names of a vector of time intervals. These names range from 0 to 1000 (matching color names in `phytools::plot.densityMap`), so the probabilities can be extracted by dividing by 1000. Since we generate the maps for each type with exactly the same tree, we can pick any `tree` object from the list of NP types:

```r
    aggregate.tree <- function(map.list) {
    # get a tree, anyone will do:
    res <- map.list[[1]]
    # get the list of maps from all NP types:
    map.probs <- lapply(map.list, function(t) {
            maps <- t$tree$maps
            lapply(maps, function(p) as.numeric(names(p))/1000)
    })
    # and replace the maps with the combined maps
    # - as probabilities for the stats:
    res$tree$cum.probs  <- compute.cum.prob(map.probs)
    # - as color names for use in phytools::plot.densityMap
    res$tree$maps <- Map(function(x,y) {names(x) <- as.character(round(y*1000))
                                        return(x)},
                  res$tree$maps, res$tree$cum.probs)
    return(res)
}
```

3.2 Parameters

Number of simulated histories per tree:

```r
B=100
```

Tree sample size:

```r
tree.sample.size=1000
```

Resolution (bins): the number of time intervals (bins) is fixed by the resolution parameter of `phytools::densityMap`, which we set to:

```
res=200
```

Setting the parameter to this value results in bins of about 30 years, roughly reflecting the state at one generation of speakers.

3.3 Stochastic character maps on the MCC summary tree from Chang et al. (2015)

3.3.1 Individual types (for illustration)

3.3.1.1 Genitives (Fig. 8)

```
g.simmap.density <- compute.simmap('G', model='ARD')
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##        N        Y
## 0.200433 0.799567
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##             N          Y
## N -0.0057305  0.0057305
## Y  0.0014365 -0.0014365
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##         N         Y
## 0.2004325 0.7995675
```

3.3.1.2 Adjectivizers (Fig. 9)

```
a.simmap.density <- compute.simmap('A', model='ARD')
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##        N        Y
## 0.806782 0.193218
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##            N         Y
## N -0.001527  0.001527
## Y  0.006376 -0.006376
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##         N         Y
## 0.8067822 0.1932178
```

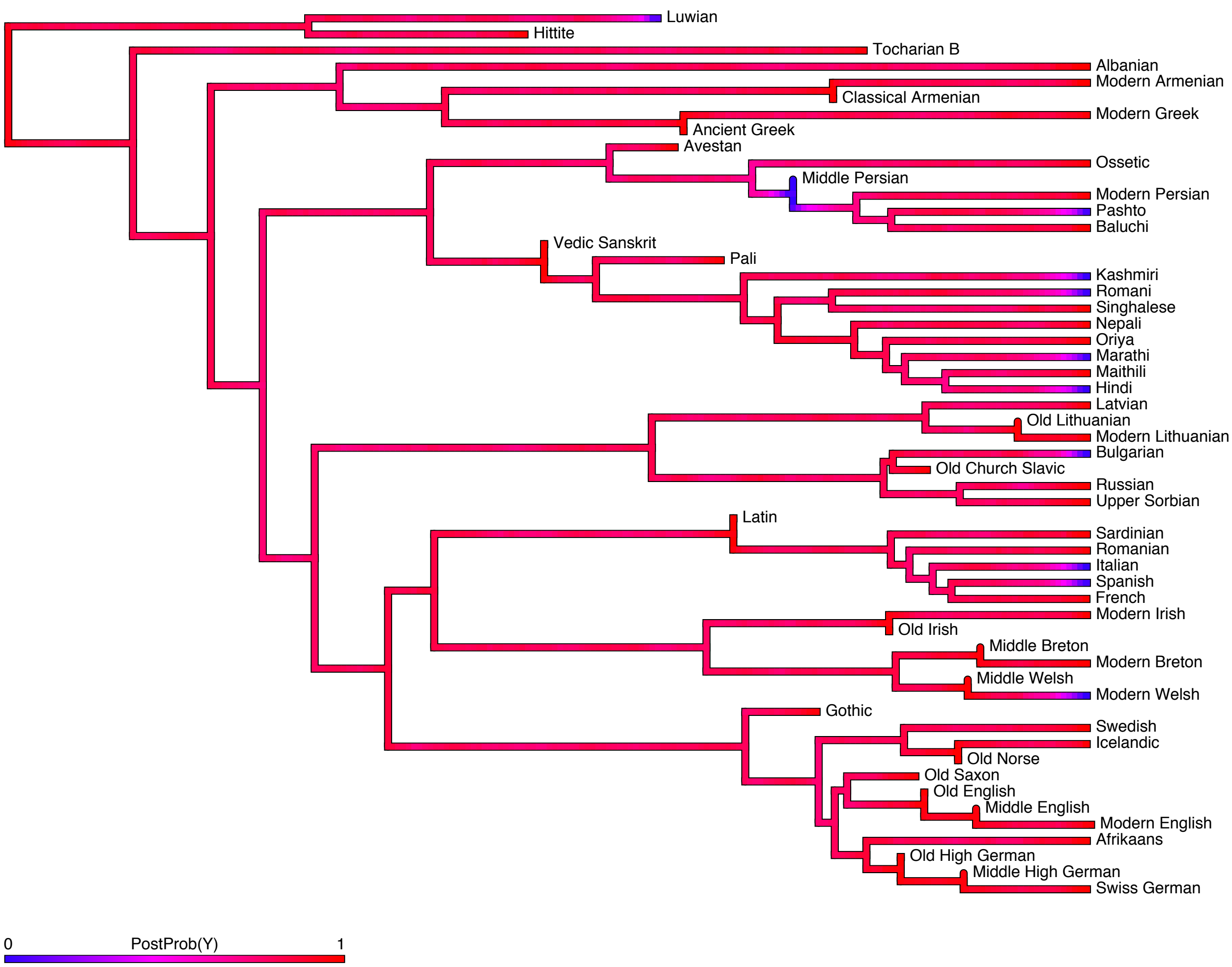

Figure 8: Posterior probability densities of genitives being available for recursion (C tree)

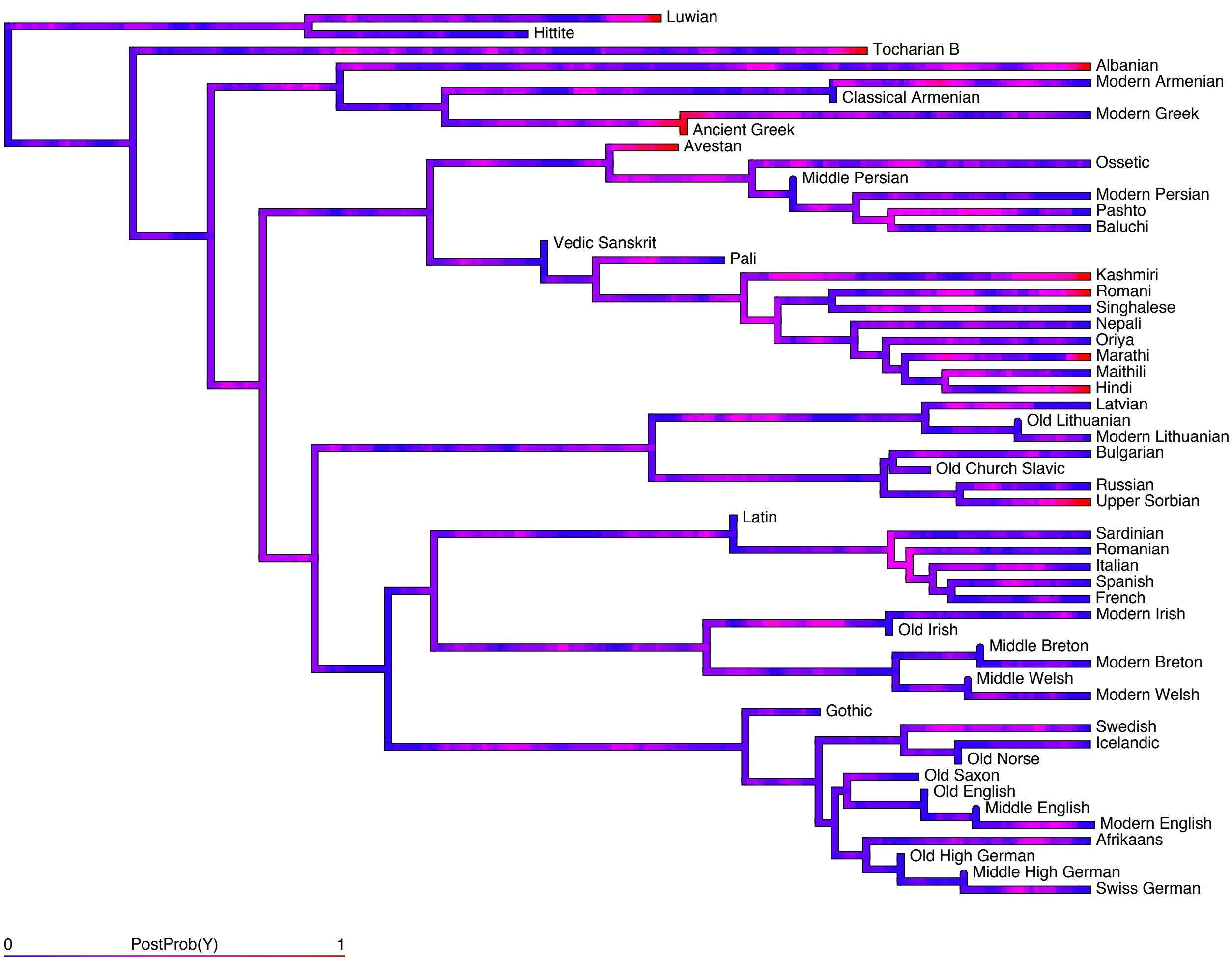

Figure 9: Posterior probability densities of adjectivizers being available for recursion (C tree)

3.3.1.3 Head marking (Fig. 10)

```
h.simmap.density <- compute.simmap('H', model='ARD')
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##        N        Y
## 0.839563 0.160437
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##           N         Y
## N -0.004451  0.004451
## Y  0.023292 -0.023292
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##         N         Y
## 0.8395631 0.1604369
```

3.3.1.4 Adpositions (Fig. 11)

```
p.simmap.density <- compute.simmap('P', model='ER')
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##   N   Y
## 0.5 0.5
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##           N         Y
## N -0.000212  0.000212
## Y  0.000212 -0.000212
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##   N   Y
## 0.5 0.5
```

3.3.1.5 Juxtaposition (Fig. 12)

```
c.simmap.density <- compute.simmap('C', model='ER')
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##   N   Y
```

Figure 10: Posterior probability densities of head marking being available for recursion (C tree)

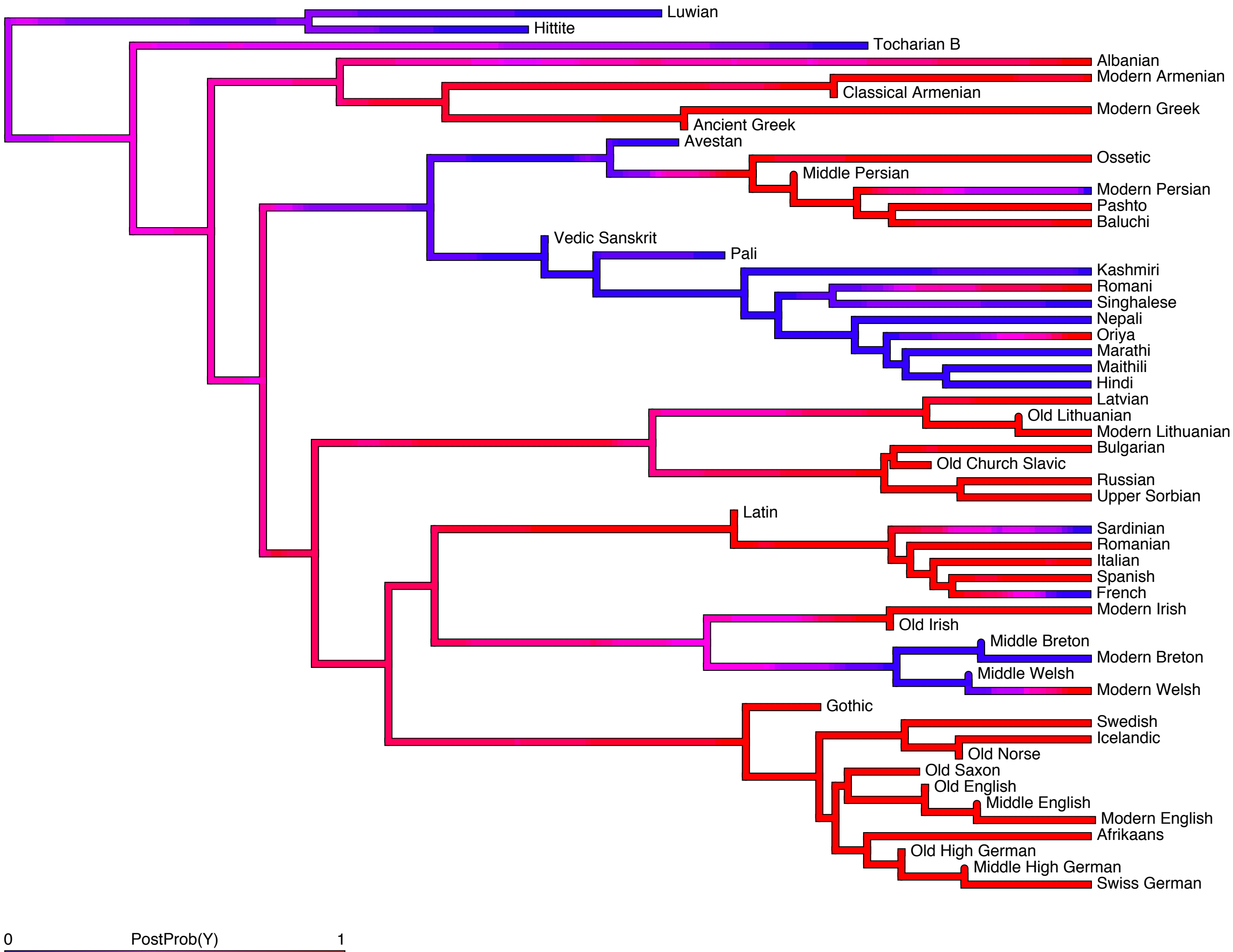

Figure 11: Posterior probability densities of adpositions being available for recursion (C tree)

```
## 0.5 0.5
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##             N          Y
## N -0.0055545  0.0055545
## Y  0.0055545 -0.0055545
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##   N   Y
## 0.5 0.5
```

Figure 12: Posterior probability densities of juxtaposition being available for recursion (C tree)

3.3.2 Probability of at least one type in the MCC summary tree from Chang et al. (2015)

```r
all.maps <- list(g.simmap.density,
                 a.simmap.density,
```

```
                 h.simmap.density,
                 p.simmap.density,
                 c.simmap.density)
aggr.maps <- aggregate.tree(all.maps)
```

Figure 13 aggregates the maps from all types into a single tree and reports the probability of at least one type being available for recursion.

Figure 13: Posterior probability densities of at least one type being available for recursion (C tree)

The number of sampled time intervals (diachronic trials) in this tree is:

```
sum(sapply(aggr.maps$tree$cum.prob, length))
```

```
## [1] 2777
```

In Figure 14 we dissolve the tree and collect all time intervals from it. We then plot the density distribution of the probabilities per interval.

```
sum.probs <- data.frame(SummedProb=unlist(aggr.maps$tree$cum.probs),
        as.data.frame(sapply(all.maps, function(t)
        as.numeric(names(unlist(t$tree$maps)))/1000))
        ) %>% gather(type, prob, SummedProb:V5)
```

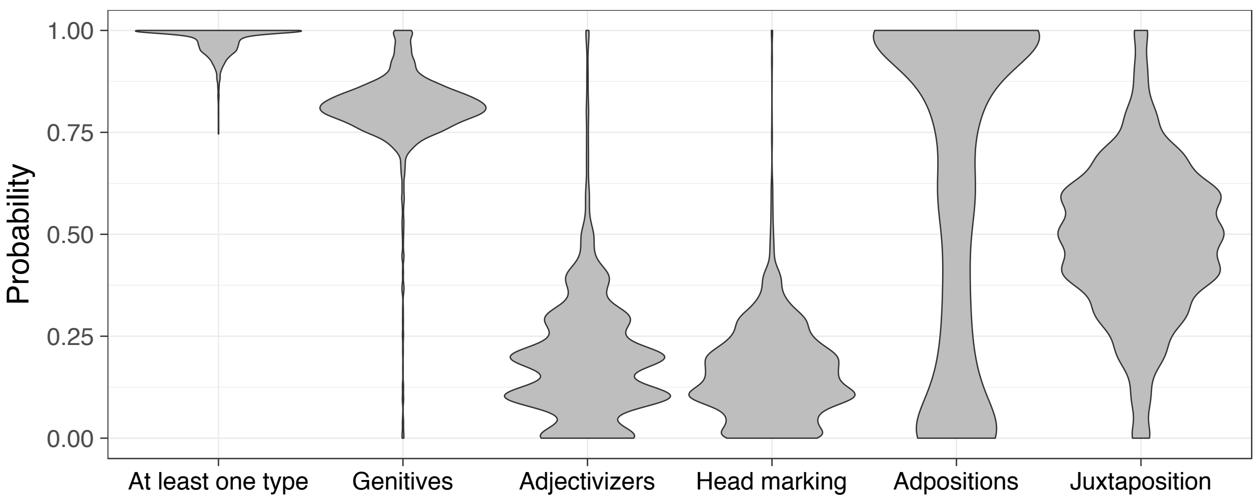

Figure 14: Probabilities of types being available for recursion per time interval (diachronic trial), based on the MCC summary tree from Chang et al. (2015)

The probability of at least one type being available for recursion is significantly higher than 50% (with a 95% CI indicating the lower bound):

```
t.test(subset(sum.probs, type %in% 'SummedProb')$prob, mu=.5, alternative='greater')
```

```
##
##  One Sample t-test
##
## data:  subset(sum.probs, type %in% "SummedProb")$prob
## t = 791.98, df = 2776, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0.5
## 95 percent confidence interval:
##  0.9743513       Inf
## sample estimates:
## mean of x
## 0.9753389
```

3.4 Stochastic character maps on the MCC summary tree from Bouckaert et al. (2012)

3.4.1 Individual types (for illustration)

3.4.1.1 Genitives (Fig. 15)

```
g.b.simmap.density <- compute.simmap('G', model='ARD', tree=ie.b.tree.sum)
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##         N        Y
## 0.249246 0.750754
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##            N         Y
## N -0.007961  0.007961
## Y  0.002643 -0.002643
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##          N         Y
## 0.2492456 0.7507544
```

Figure 15: Posterior probability densities of genitives being available for recursion (B tree)

3.4.1.2 Adjectivizers (Fig. 16)

```
a.b.simmap.density <- compute.simmap('A', model='ARD', tree=ie.b.tree.sum)
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##        N        Y
## 0.812897 0.187103
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##            N          Y
## N -0.0018135  0.0018135
## Y  0.0078790 -0.0078790
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##         N         Y
## 0.8128966 0.1871034
```

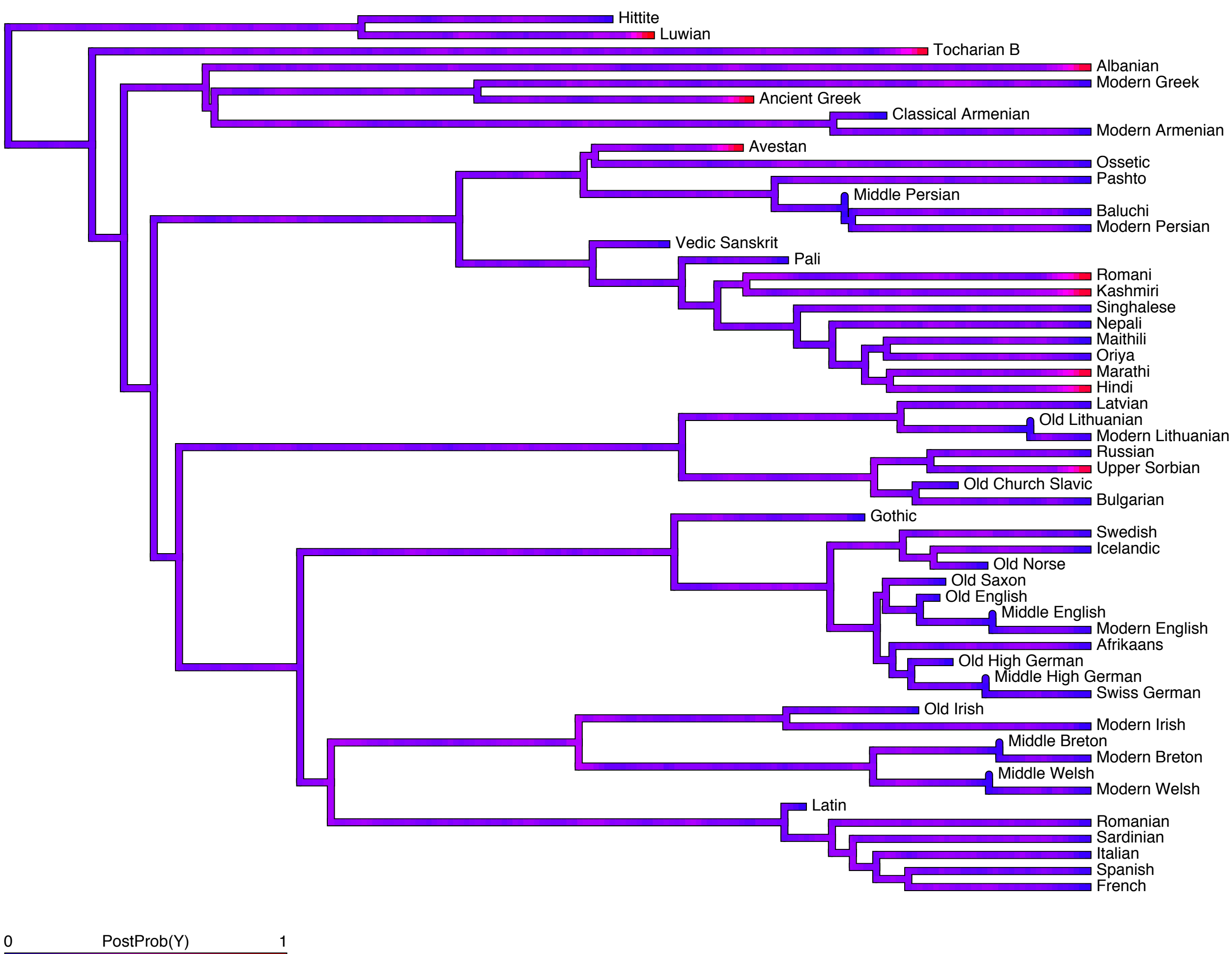

Figure 16: Posterior probability densities of adjectivizers being available for recursion (B tree)

3.4.1.3 Head marking (Fig. 17)

```
h.b.simmap.density <- compute.simmap('H', model='ARD', tree=ie.b.tree.sum)
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##        N        Y
## 0.848622 0.151378
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##            N          Y
## N -0.0067385  0.0067385
## Y  0.0377760 -0.0377760
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##         N         Y
## 0.8486224 0.1513776
```

3.4.1.4 Adpositions (Fig. 18)

```
p.b.simmap.density <- compute.simmap('P', model='ER', tree=ie.b.tree.sum)
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##   N   Y
## 0.5 0.5
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##           N         Y
## N -0.000169  0.000169
## Y  0.000169 -0.000169
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##   N   Y
## 0.5 0.5
```

3.4.1.5 Juxtaposition (Fig. 19)

```
c.b.simmap.density <- compute.simmap('C', model='ER', tree=ie.b.tree.sum)
```

```
## Using pi estimated from the stationary distribution of Q assuming a flat prior.
## pi =
##   N   Y
```

Figure 17: Posterior probability densities of head marking being available for recursion (B tree)

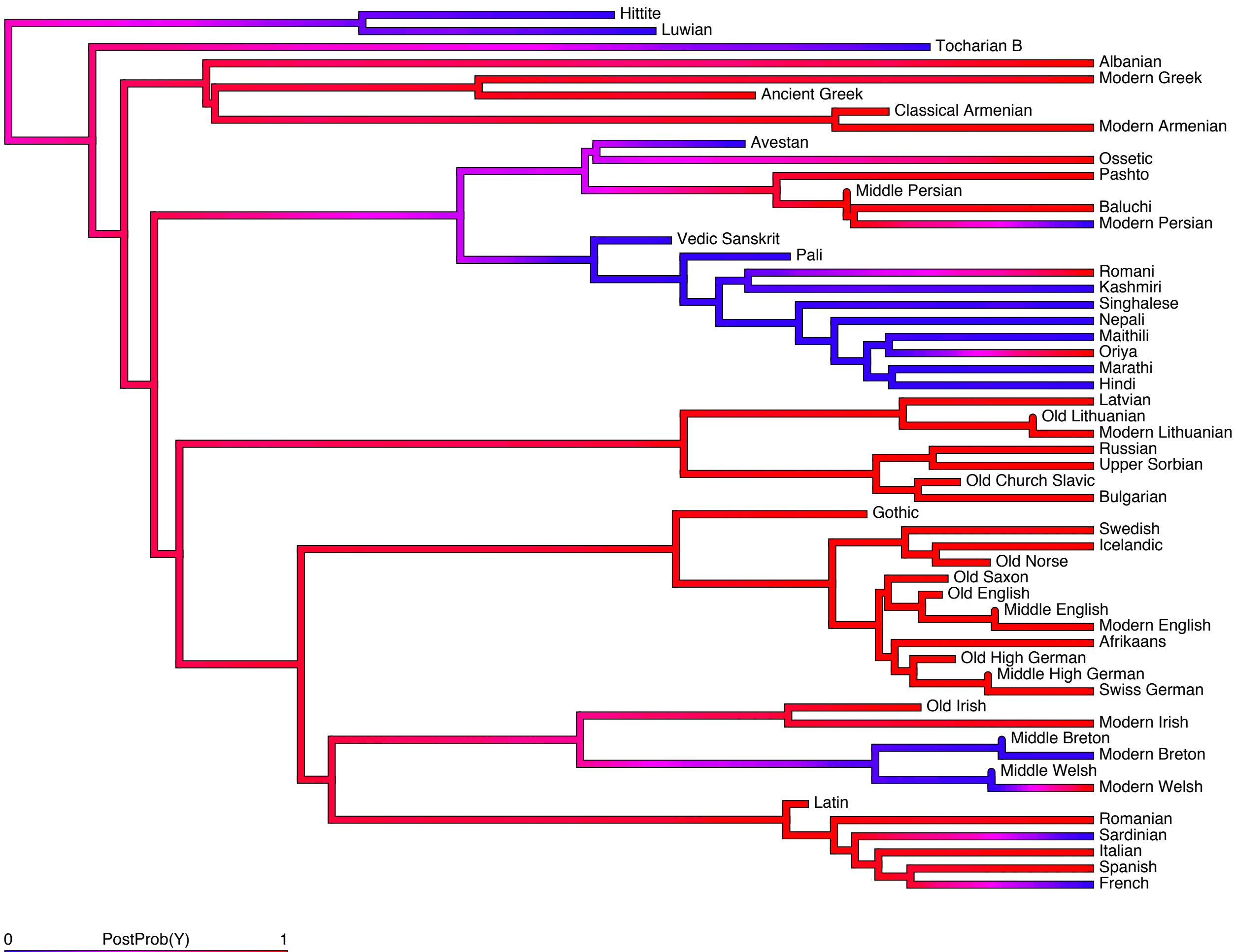

Figure 18: Posterior probability densities of adpositions being available for recursion (B tree)

```
## 0.5 0.5
##
## make.simmap is sampling character histories conditioned on the transition matrix
##
## Q =
##             N          Y
## N -0.004057  0.004057
## Y  0.004057 -0.004057
## (specified by the user);
## and (mean) root node prior probabilities
## pi =
##   N   Y
## 0.5 0.5
```

Figure 19: Posterior probability densities of juxtaposition being available for recursion (B tree)

3.4.2 Probability of at least one type in the MCC summary tree from Bouckaert et al. (2012)

```
all.maps.b <- list(g.b.simmap.density,
                   a.b.simmap.density,
```

```
                    h.b.simmap.density,
                    p.b.simmap.density,
                    c.b.simmap.density)
aggr.maps.b <- aggregate.tree(all.maps.b)
```

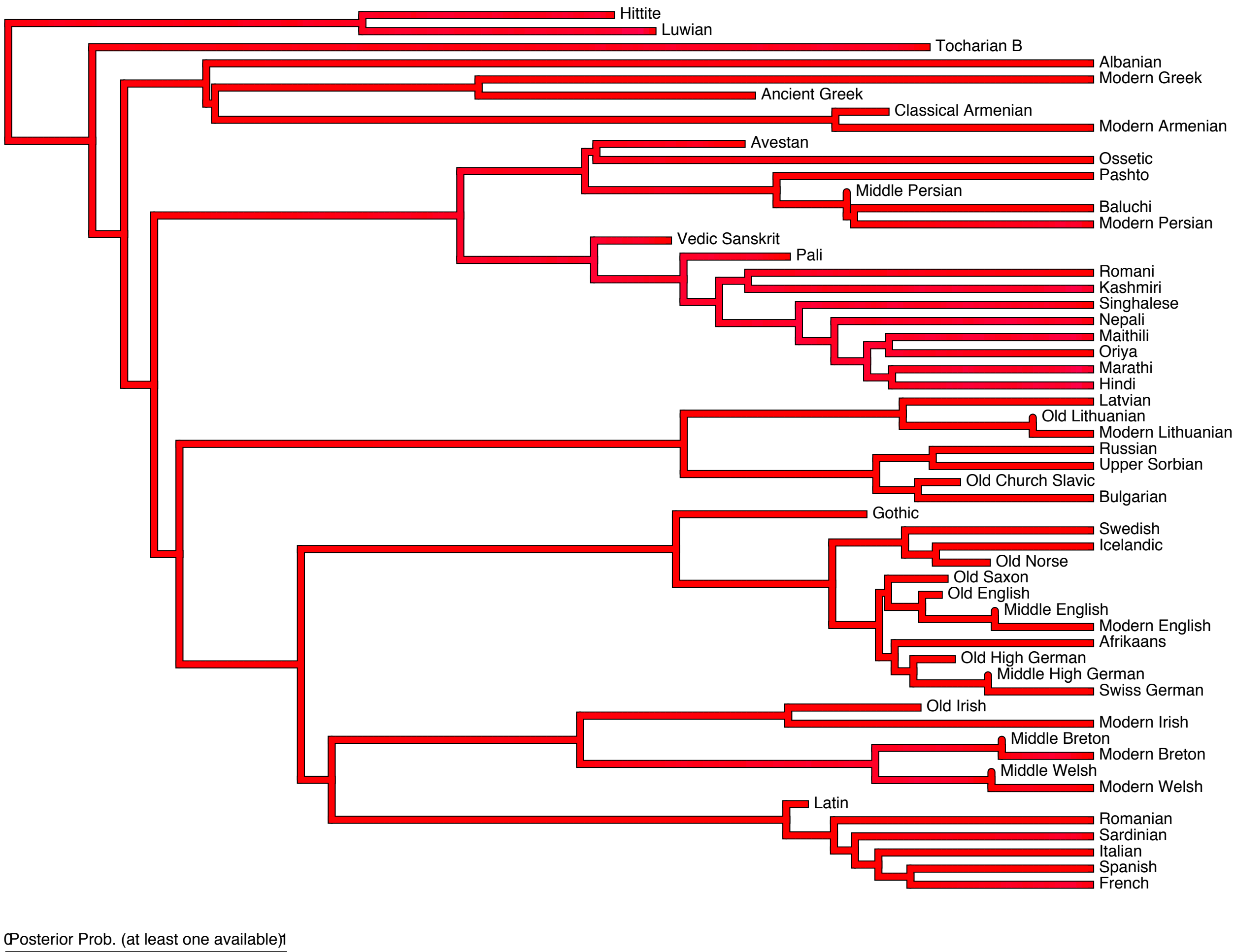

Figure 20: Posterior probability densities of at least one type being available for recursion

The number of sampled time intervals (diachronic trials) in this tree is:

```
sum(sapply(aggr.maps.b$tree$cum.prob, length))
```

```
## [1] 3257
```

In Figure 18 we again dissolve the tree and collect all time intervals from it. We then plot the density distribution of the probabilities per interval.

```
sum.probs.b <- data.frame(SummedProb=unlist(aggr.maps.b$tree$cum.probs),
        as.data.frame(sapply(all.maps.b, function(t)
        as.numeric(names(unlist(t$tree$maps)))/1000))
        ) %>% gather(type, prob, SummedProb:V5)
```

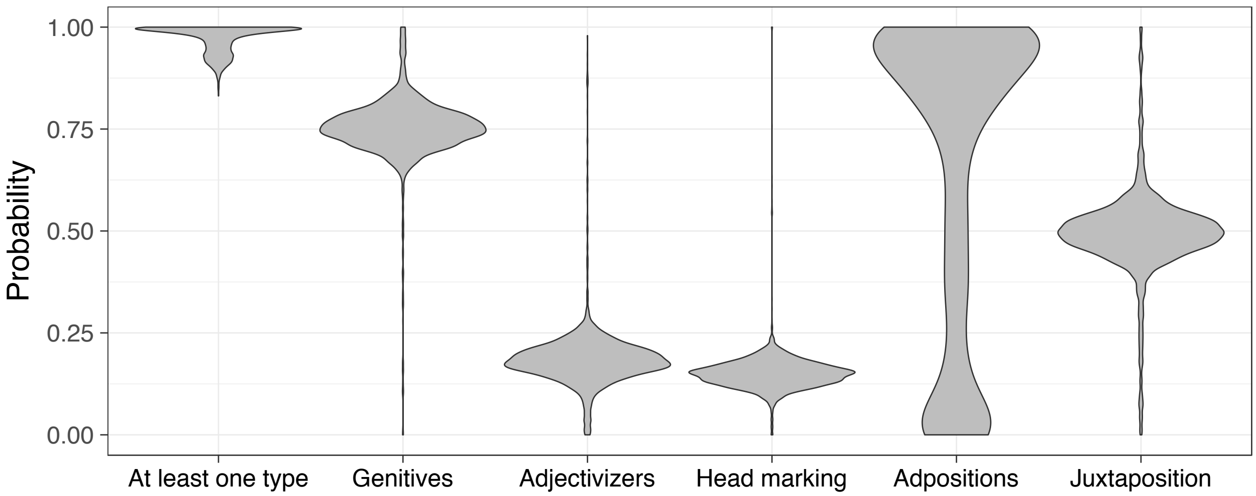

Figure 21: Probabilities of types being available for recursion per time interval (diachronic trial), based on the MCC summary tree from Bouckaert et al. (2012)

The probability of at least one type being available for recursion is again significantly higher than 50% (with a 95% CI indicating the lower bound):

```
t.test(subset(sum.probs.b, type %in% 'SummedProb')$prob, mu=.5, alternative='greater')
```

```
##
##  One Sample t-test
##
## data:  subset(sum.probs.b, type %in% "SummedProb")$prob
## t = 811.25, df = 3256, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0.5
## 95 percent confidence interval:
##  0.9700423       Inf
## sample estimates:
## mean of x
## 0.9709975
```

3.5 Stochastic character maps on the posterior tree sample from Chang et et al. (2015)

```
tree.sample <- sample(ie.trees, tree.sample.size)
aggr.per.tree.list <- mclapply(tree.sample, function(t) {
    all.tree.maps <- list(compute.simmap('G', tree=t, model='ARD', iterations=B),
                          compute.simmap('A', tree=t, model='ARD', iterations=B),
                          compute.simmap('H', tree=t, model='ARD', iterations=B),
                          compute.simmap('P', tree=t, model='ER', iterations=B),
                          compute.simmap('C', tree=t, model='ER', iterations=B)
                          )
    aggr.tree <- aggregate.tree(all.tree.maps)
    list(n.trials=sum(sapply(aggr.tree$tree$cum.prob, length)),
         probs=data.frame(SummedProb=unlist(aggr.tree$tree$cum.probs),
```

```
        as.data.frame(sapply(all.tree.maps, function(t)
        as.numeric(names(unlist(t$tree$maps)))/1000))
        ) %>% gather(type, prob, SummedProb:V5)
        )
    }, mc.cores=parallel::detectCores()-2)
```

Number of sampled time intervals (diachronic trials) for which we can evaluate the probability of at least one type being available for recursion:

```
sum(unlist(sapply(aggr.per.tree.list, "[", "n.trials")))
```

```
## [1] 2795738
```

Size of tree sample:

```
tree.sample.size
```

```
## [1] 1000
```

Since here we estimate density maps simultaneously for many different trees, we cannot visualize the combined posterior probability density over time in a single tree. What is of more interest anyway is the overall distribution of probabilities per time interval. Therefore, like in Figures 11 and 21, we disolve all trees, collecting the time intervals from them. We then plot the density distribution of the probabilities per interval.

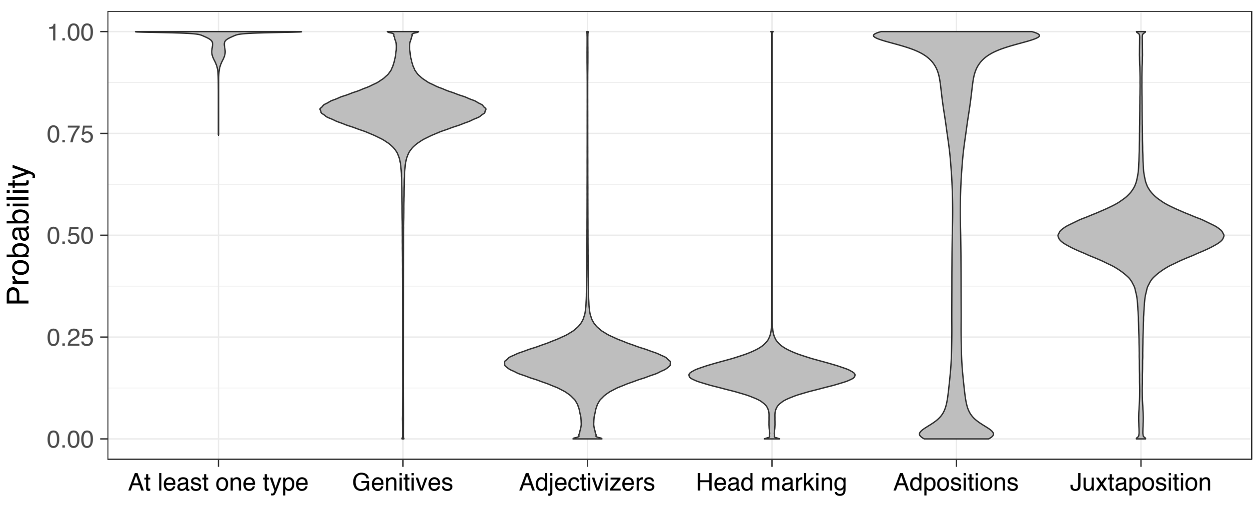

Figure 22: Probabilities of types being available for recursion per time interval (diachronic trial), with phylogenetic uncertainty based on Chang et al.'s trees

The probability of at least one type being available for recursion is significantly higher than 50% (with a 95% CI indicating the lower bound):

```
t.test(subset(all.probs, type %in% 'SummedProb')$prob, mu=.5, alternative='greater')
```

```
##
```

```
##  One Sample t-test
##
## data:  subset(all.probs, type %in% "SummedProb")$prob
## t = 28789, df = 2795700, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0.5
## 95 percent confidence interval:
##  0.9767593       Inf
## sample estimates:
## mean of x
## 0.9767866
```

3.6 Stochastic character maps on the posterior tree sample from Bouckaert et al. (2012)

```
tree.b.sample <- sample(ie.b.trees, tree.sample.size)
aggr.per.tree.list.b <- mclapply(tree.b.sample, function(t) {
    all.tree.maps <- list(compute.simmap('G', tree=t, model='ARD', iterations=B),
                          compute.simmap('A', tree=t, model='ARD', iterations=B),
                          compute.simmap('H', tree=t, model='ARD', iterations=B),
                          compute.simmap('P', tree=t, model='ER', iterations=B),
                          compute.simmap('C', tree=t, model='ER', iterations=B)
                          )
    aggr.tree <- aggregate.tree(all.tree.maps)
    list(n.trials=sum(sapply(aggr.tree$tree$cum.prob, length)),
         probs=data.frame(SummedProb=unlist(aggr.tree$tree$cum.probs),
        as.data.frame(sapply(all.tree.maps, function(t)
        as.numeric(names(unlist(t$tree$maps)))/1000))
        ) %>% gather(type, prob, SummedProb:V5)
        )
    }, mc.cores=parallel::detectCores()-2)
```

Number of sampled time intervals (diachronic trials) for which we can evaluate the probability of at least one type being available for recursion:

```
sum(unlist(sapply(aggr.per.tree.list.b, "[", "n.trials")))
```

```
## [1] 2414439
```

Size of tree sample:

```
tree.sample.size
```

```
## [1] 1000
```

We again plot the density distribution of the probabilities per interval.

The probability of at least one type being available for recursion is significantly higher than 50% (with a 95% CI indicating the lower bound):

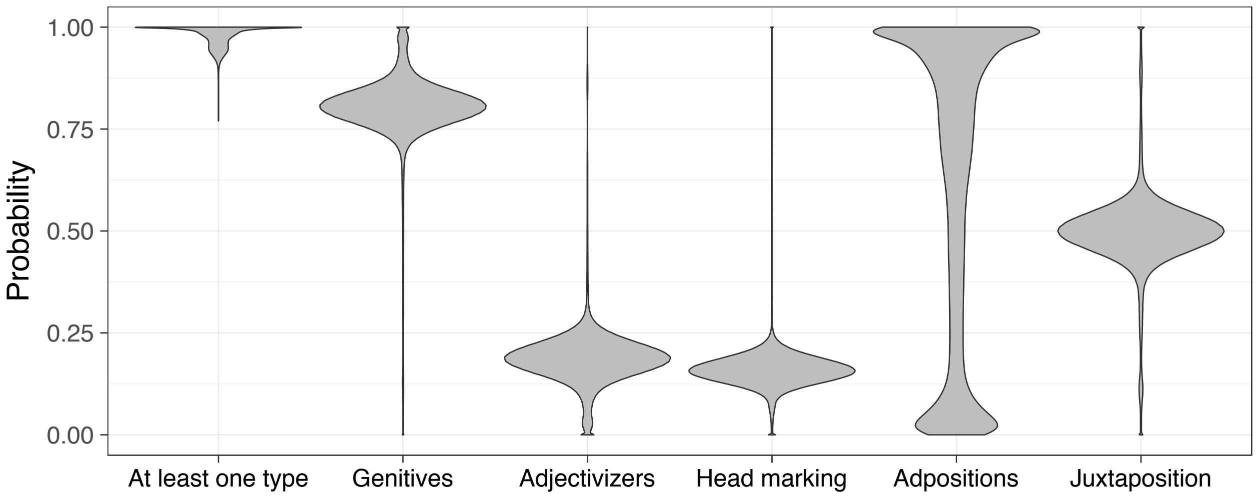

Figure 23: Probabilities of types being available for recursion per time interval (diachronic trial), with phylogenetic uncertainty based on Bouckaert et al.'s trees

```r
t.test(subset(all.probs.b, type %in% 'SummedProb')$prob, mu=.5, alternative='greater')
```

```
##
##  One Sample t-test
##
## data:  subset(all.probs.b, type %in% "SummedProb")$prob
## t = 28104, df = 2414400, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0.5
## 95 percent confidence interval:
##  0.9749544       Inf
## sample estimates:
## mean of x
## 0.9749822
```

References

Bickel, Balthasar, Alena Witzlack-Makarevich, Kamal K Choudhary, Matthias Schlesewsky & Ina Bornkessel-Schlesewsky. 2015. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE* 10. e0132819.

Bouckaert, Remco R & Joseph Heled. 2014. `DensiTree` 2: Seeing trees through the forest. BioRxiv [http://dx.doi.org/10.1101/012401].

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. 957–960.

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

Dunn, Michael J., Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.

Huelsenbeck, John P., Rasmus Nielsen & Jonathan P. Bollback. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52. 131–158.

Levinson, Stephen C., Simon J. Greenhill, Russell D. Gray & Michael J. Dunn. 2011. Universal typological dependencies should be detectable in the history of language families. *Linguistic Typology* 15. 509–534.

Nielsen, Rasmus. 2002. Mapping mutations on phylogenies. *Systematic Biology* 51. 729–732.

Pagel, Mark. 1999. The maximum-likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* 48. 612–622.

Pagel, Mark & Andrew Meade. 2013. `BayesTraits V2`. Software and manual, http://www.evolution.rdg.ac.uk/BayesTraitsV2Beta.html.

Pagel, Mark, Andrew Meade & Daniel Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53. 673–684.

Revell, Liam J. 2012. `phytools`: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3. 217–223.

Revell, Liam J. 2013. Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution* 4. 754–759.

Ronquist, Fredrik. 2004. Bayesian inference of character evolution. *Trends in Ecology and Evolution* 19. 475–481.

Wickham, Hadley. 2009. *`ggplot2`: Elegant graphics for data analysis*. New York: Springer. http://ggplot2.org.

Yu, Guangchuang, David K. Smith, Huachen Zhu, Yi Guan & Tommy Tsan-Yuk Lam. 2016. `ggtree`: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*. 1–9. doi:10.1111/2041-210X.12628.