

Diachronic change in Indo-European motion event encoding

Annemarie Verkerk

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands /
International Max Planck Research School for Language Sciences, Radboud
University Nijmegen, The Netherlands

There are many different syntactic constructions that languages can use to encode motion events. In recent decades, great advances have been made in the description and study of these syntactic constructions from languages spoken around the world (Talmy 1985, 1991, Slobin 1996, 2004). However, relatively little attention has been paid to historical change in these systems (exceptions are Vincent 1999, Dufresne, Dupuis & Tremblay 2003, Kopecka 2006 and Peyraube 2006). In this article, diachronic change of motion event encoding systems in Indo-European is investigated using the available historical-comparative data and phylogenetic comparative methods adopted from evolutionary biology. It is argued that Proto-Indo-European was not satellite-framed, as suggested by Talmy (2007) and Acedo Matellán and Mateu (2008), but had a mixed motion event encoding system, as is suggested by the available historical-comparative data.

Keywords: Indo-European, motion event encoding, grammaticalization, phylogenetic comparative methods, preverbs

1. Introduction

The modern cognitive tradition of the study of motion event encoding originates with Talmy (1985, 1991), who postulated the now well-known difference between so-called ‘verb-framed’ and ‘satellite-framed’ languages. Central concepts in Talmy’s (1985, 1991) framework are *PATH* and *MANNER*. *Path* refers to the path or trajectory that a person or an object has while moving, while *manner* refers to the way in which a person or an object moves (for instance, by flying, swimming or walking). *Path* is encoded on the verb in verb-framed constructions as exemplified by the Albanian example in (1), where it is encoded on the verb *kaloj* ‘to pass’. The

manner of motion is not overtly expressed in this sentence. Path is encoded on the satellite in satellite-framed constructions, as shown by the Swedish example in (2), where it is encoded on the preposition *genom* ‘through’. The manner is expressed by the verb *krypa* ‘to crawl’ in this sentence. These two examples are taken from a parallel corpus of *Alice’s Adventures in Wonderland*, which will be introduced below.

(1) Albanian

nëse ende do të vazhdoj të zvogëlohem
 if still FUT FUT continue.PRES.1SG to melt.PRES.1SG
 do provoj të kaloj nën derë.
 FUT try.PRES.1SG to pass.PRES.1SG under door.F.INDF.ACC.SG
 ‘If I continue to melt, I can try to pass under the door’

(2) Swedish

gör den mig mindre kan jag
 make.PRES 3SG.N 1SG.OBJ small.COMP can.PRES.AUX 1SG.SBJ
 krypa genom springa-n under don-en
 crawl.INF through slot-SG.DEF.UT below door-SG.DEF.UT
 ‘And if it makes me grow smaller, I can creep under the door’

Since Talmy’s (1985, 1991) seminal work, many people have investigated the encoding of motion events in languages around the world (Aske 1989, Slobin & Hoiting 1994, Slobin 1996, 2004, Beavers, Levin & Tham 2010, Croft et al. 2010, and others). However, less attention has been paid to diachronic change in motion event encoding. As of yet, little is known about how languages become satellite-framed or verb-framed.

Nevertheless, a few hypotheses on change in motion event encoding have been put forward. Croft et al. (2010: 236), for instance, extensively discuss the two grammaticalization pathways that lead to unified lexical items, most often verbs that express path (PATH VERBS). These are presented in (3). Both pathways are concerned with complex motion events in which both path and manner are expressed.

(3) Grammaticalization pathways presented by Croft et al. (2010)

- a. Coordination > Serialization > Satellite-framed > Verb-Satellite fusion
- b. Coordination > Verb-framed > Verb-Adverb fusion

Croft et al. (2010) describe (3a) as a process in which a complex motion event starts with coordination of two clauses (as in ‘John walked and crossed the street’). This construction gives rise to a serial verb construction (‘John walked crossed the street’). After this step, one of the verbs of the SVC becomes grammaticalized into a satellite (‘John walked across the street’). Then, these satellites fuse again with their verb root to form single roots, most often creating path verbs (‘John crossed

the street’). In (3b), the coordinated construction changes to a verb-framed construction with a subordinate manner element (‘John crossed the street walking’). Then, the verb and the subordinate manner verb or manner adverb merge again into a compound (‘John crossed-walking the street’) and ultimately merge into a single, semantically bleached root (‘John crossed the street’). Both of Croft et al.’s (2010) pathways end in fusion of verb roots and particles, resulting in unified lexical items, most often path verbs. These unified lexical elements can then be coordinated again to express more complex motion events, and the beginnings of the grammaticalization pathways in (3) are reached once more.

Another study that discusses diachronic change in motion event encoding is Wälchli (2009: 183–221). Wälchli studied lexicalization patterns in motion events in a worldwide sample of 117 languages. He focused on the use of path verbs in five different path domains (enter, exit, ascend, descend and pass/cross) in a set of 56 motion clauses in a parallel corpus of the *Gospel according to Mark* (a Bible text). An assessment of whether a path domain is predominantly encoded by path verbs or by other types of verbs (including deictic verbs and manner verbs) was made for each domain. Wälchli found that certain languages encode all path domains predominantly with path verbs, such as French, Italian and Spanish, while others encode all path domains predominantly with other types of verbs, such as Latin, Irish and German. Other languages encode some but not all path domains predominantly with path verbs, such as English, Greek and Hindi.

Wälchli finds that there is little genealogical stability for the choice of verb type in his worldwide sample. The same holds for Indo-European. None of the subgroups of the Indo-European language family behave in a completely unified manner, and the languages of the Indo-Iranian subgroup even range across the entire scale of path lexicalization — some languages use path verbs for all path domains, while others do not use path verbs for any of the path domains. To explain these results, Wälchli proposes the existence of an areal pattern rather than a genealogical one: “Languages in Northern and Central Europe (including Finnic and Hungarian) as well as in the Caucasus (except Armenian) tend to lack route verbs [path verbs, AV]” (p. 214). Even though languages that do not use path verbs are found all around the world, they seem to be most common in the area mentioned above. Wälchli also finds that no language family larger than a subfamily such as Germanic lacks path verbs altogether. He suggests that languages tend to acquire new path verbs with relative ease due to a variety of mechanisms, including the univerbation of adverbial path markers with verb stems (Croft et al. 2010’s first grammaticalization pathway, see 3a) and borrowing.

Croft et al. (2010) indicate that there are clear grammaticalization pathways for diachronic change in motion event encoding constructions. Wälchli (2009) suggests that languages may move through these grammaticalization pathways

at a steady pace, as even closely related languages may behave quite differently. However, none of these claims have been tested on a full-scale study of motion event encoding in a single language family. This article aims to fill that gap.

The current article is an investigation of diachronic change in motion event encoding in the Indo-European language family. Specifically, it investigates whether the application of phylogenetic comparative analysis on contemporary language data can provide support for one of the following two hypotheses: (i) the hypothesis that Proto-Indo-European was satellite-framed, as proposed by Talmy (2007) and Acedo Matellán & Mateu (2008), and (ii) the hypothesis that Proto-Indo-European was typologically mixed, as suggested by a review of the comparative work on the ancient Indo-European languages presented in Section 2. The question of whether Proto-Indo-European was satellite-framed or mixed was investigated using the available historical-comparative data from ancient Indo-European languages such as Latin, Homeric Greek and Vedic Sanskrit as well as using phylogenetic comparative methods that are also employed by evolutionary biologists. In particular, a method called ancestral state estimation is used to infer the behavior of the now extinct ancestors of contemporary languages on the basis of the behavior of modern languages by finding the evolutionary model that has the best fit to the modern data. For the ancestral state reconstruction, comparable usage data on motion event encoding from a sample of twenty contemporary Indo-European languages were used (English, Dutch, German, Swedish from Germanic, Irish from Celtic, Portuguese, French, Italian, Romanian from Romance, Russian, Polish, Serbo-Croatian, Latvian, Lithuanian from Balto-Slavic, Hindi, Nepali, Persian from Indo-Iranian, Albanian, Armenian and Modern Greek). This dataset is taken from a parallel corpus of two novels: *Alice's Adventures in Wonderland* by Lewis Carroll and *O Alquimista* 'The Alchemist' by Paulo Coelho.

Given the availability of data from the ancient Indo-European languages such as Latin, Ancient Greek, Vedic Sanskrit, Old Church Slavonic, Gothic and Classical Armenian, it would have been possible to investigate motion event encoding in Proto-Indo-European using corpora of these ancient languages. The best parallel corpus of ancient Indo-European languages that is currently available is the *New Testament*, which is available in Latin, Ancient Greek, Old Church Slavonic, Gothic and Classical Armenian (Haug & Jøhndal 2008). However, I have chosen not to use this corpus because the *New Testament* is not a particularly rich source for Talmian motion data. The *New Testament* lacks instances of the satellite-framed construction with a manner verb and a path satellite (see again example 2) that is diagnostic for the study of Talmian motion event encoding (see also Beavers, Levin & Tham 2010: 332, Croft et al. 2010: 221). The satellite-framed construction is diagnostic because it is typically translated with a satellite-framed construction in a satellite-framed language, but translated with a verb-framed construction in

a verb-framed language (Slobin 2005). Manner is typically added in about 25% of the constructions when translating a text in a verb-framed language to a satellite-framed language, whereas manner is deleted in about 50% of the constructions when translating a text in a satellite-framed language to a verb-framed language (Slobin 1996:212). Given this discrepancy, instances of both the satellite-framed construction as well as instances of the verb-framed construction are needed for the most optimal characterization of motion event encoding in a given language. The *New Testament* unfortunately cannot offer both. This is in part due to a lack of manner verbs, as well as to a tendency to use manner verbs in descriptions of activities (e.g. ‘The man walked’) instead of descriptions of motion events (e.g. ‘The man walked into the room’). An illustration of these issues in the *New Testament* and a comparison with a corpus of the two English *Alice* novels by Lewis Carroll is given in Appendix 1.

Another problem with the use of Bible texts is that they are written in a religious register that restricts translational freedom to at least some extent. The study of cross-linguistic motion event encoding has been done using parallel corpora in the past, and one of the findings has been that in translations from a verb-framed language to a satellite-framed language, manner information may be added, while in translations from a satellite-framed language to a verb-framed language, manner information is often deleted — this is done in order to approach the native motion encoding patterns (Slobin 1996, Baicchi 2005). The religious convention to translate the Bible in such a way that it stays close to the original text might interfere with the artistic freedom that would be needed to translate motion events in the most natural way. Given these disadvantages associated with the use of the *New Testament* for motion event encoding, the decision was made to use the parallel corpus consisting of the two modern novels mentioned above. The use of contemporary data admittedly does add a further dimension to the analysis in the sense that the contemporary languages are removed further in time from Proto-Indo-European than the ancient Indo-European languages such as Latin, Vedic Sanskrit and Ancient Greek. However, adding this dimension is compensated by the high quality data that becomes accessible by using data from a parallel corpus of contemporary language use.

In the current article, I use ancestral state estimation analyses to infer the motion event encoding system of Proto-Indo-European, the ancestor of all Indo-European languages. The use of ancestral state estimation analyses on data from contemporary languages allows me to draw a comprehensive picture of typological change from, for instance, the reconstructed Proto-Indo-European language to the non-attested reconstructed Proto-Germanic language, and from Proto-Germanic to the contemporary Germanic languages. The information on motion event encoding in the ancient Indo-European languages is incorporated in the current analysis as much as possible. However, note that the information on motion event

encoding in ancient languages is mostly concerned with qualitative descriptions of different motion constructions. The data presented in this article is concerned with quantitative analysis of the use of different motion constructions in a corpus, and therefore this article goes beyond stating which constructions are attested.

The historical–comparative data is presented in Section 2, while the contemporary data from the parallel corpus is presented in Section 3. The phylogenetic comparative methods that are employed are described in Section 4. The results of these analyses are presented in Section 5. A general discussion of the findings is given in Section 6, while future directions and a general conclusion are provided in Section 7.

2. The historical–comparative evidence for diachronic change in Indo-European motion event encoding

Talmy (2007) and Acedo Matellán & Mateu (2008) have proposed that Proto-Indo-European was satellite-framed on the basis of motion event encoding of several ancient Indo-European languages:

For their characteristic representation of Motion events, Latin, Classical Greek and Proto-Germanic all exhibited the presumably Indo-European pattern of using Co-event-conflating verb roots [manner verbs, AV] together with Path satellites that formed prefixes on the verb roots. (Talmy 2007: 154)

Their claim is based on the existence of a specific satellite-framed construction in these ancient or reconstructed Indo-European languages, a construction in which path was encoded on adverbial particles.

The most ancient Indo-European languages (Greek, Hittite, Vedic Sanskrit, Avestan and Old Persian) were characterized by a relatively free word order, in which these adverbial particles moved freely within the sentence as modifiers of verbs and nouns (Delbrück 1888, 1893, Speyer 1896, Whitney 1879, Kuryłowicz 1964, Watkins 1964, Hofmann & Szantyr 1965, Lehmann 1974: 116ff., 212ff., 228ff., 233ff., Penney 1989, Hewson & Bubenik 2006: 358ff., Luraghi 2010). These adverbial particles are called “preverbs” when they function as modifiers of verbs. Examples of their different functions are provided in (4), in which *epí* is a preverb that is separated from its verb (a process called “tmesis”) and in (5), in which *epí* is an adposition, which could have been placed in several different places in the sentence (examples taken from Hewson & Bubenik 2006: 5–6).

(4) Homeric Greek

kai *epi* knéphas hierón élthēi
and on darkness sacred come.3SG
‘and the sacred darkness closes in’ (Il. 1 1.209)

(5) Homeric Greek

pléōn epì oínopa pónton

sailing over wine.dark sea.ACC

'sailing over the wine-dark sea' (Il. 7.88)

Preverbs were used to encode path in satellite-framed constructions such as (4), in which *epi* encodes that the darkness is coming towards a reference point, presumably the speaker. This was true for most ancient Indo-European languages, such as Homeric Greek, Hittite, Vedic Sanskrit, Tocharian, Avestan and Old Persian, and therefore the preverb system is reconstructed to be Proto-Indo-European. Some of these preverbs encoded deictic reference rather than path and were used in the ancient Indo-European languages to encode deixis. Most ancient Indo-European languages (except Hittite) lacked deictic verbs, a situation that can still be observed in modern Balto-Slavic languages.

For these reasons, the changes with regard to the preverb system that occurred as the contemporary Indo-European languages emerged are of great importance if diachronic change in Indo-European motion event encoding is to be understood. Today, none of the contemporary Indo-European languages posses the preverb system as it is attested in the ancient Indo-European languages. However, what was originally the preverb system has morphed into systems of verbal path prefixes, prepositions and postpositions that are attested in the modern languages. For some subgroups of Indo-European, one can see or reconstruct with some confidence how the preverb system has changed over time, giving rise to the motion encoding constructions that we see today. For others, this is more difficult. In this Section, I present an account of changes to the preverb system for the different Indo-European subgroups.

One of the best-documented cases of typological change in motion event encoding is the change of satellite-framed Latin to the verb-framed Modern Romance languages (Acedo Matellán & Mateu 2008, 2010, Vincent 1999, Dufresne, Dupuis & Tremblay 2003). In Latin, the free adverbial particles had been transformed to a system of verbal path prefixes and prepositions (Hofmann & Szantyr 1965: 21ff., Leumann 1977: 557ff.). The modern Romance languages shifted from this satellite-framed system to a verb-framed system. Kopecka (2006, 2009) describes extensively how French became verb-framed: verb stems were fused together with the path prefixes inherited from Latin, and ultimately the path prefixes lost their productivity completely.

However, not all Romance languages seem to be completely verb-framed. Italian is reported to have satellite-framed constructions with a limited set of verbs (Folli & Ramchand 2001, 2005). Italian can also make use of a satellite-framed construction using a set of post-verbal particles that are used both with path verbs and

with manner verbs (Masini 2005, Iacobini & Masini 2006, 2007). Brucale, Iacobini & Mocciaro (2011) and Brucale (2011), who study Classical Latin, show that Latin might not be completely satellite-framed either. Their studies suggest that even though it was possible to use the satellite-framed construction in Classical Latin, it might not have been used very often at all. Ferrari & Mosca (2010:320) seem to support this view when they describe Latin as an “unstable” language in which path is distributed over three linguistic elements (preverbs, prepositions and cases). Even though it is clear that the satellite-framed construction was productive in Latin, corpus studies are needed to tell us how often it was used in order to assess the exact magnitude of the change from Latin to the Romance languages. The same applies to most other ancient Indo-European languages discussed in this article — there exist virtually no quantitative corpus studies of motion event encoding.

A similar change took place in the Indo-Aryan languages: the preverbs that were present in Vedic Sanskrit (for an overview see Danesi 2013) became more fixed and developed into a system of postpositions (Speyer 1896, Bloch 1965). They also became more closely associated with the verb and in the end merged with verb roots altogether: “Preverbs are shown by etymology to exist at the beginning of many modern verbs commencing with *o-* or *u-* (*apa-*, *ava-*, *ud-*) or by *p-* (*pra-*, *prati-*), *v-/b-* (*vi-*), *sam-*” (Bloch 1965:158, italicization mine). Speyer (1896:47) writes that this univerbation process took place during the transition from Vedic Sanskrit to Classical Sanskrit. Hindi, the only Indo-Aryan language to be studied from a Talmian perspective, is verb-framed (Narasimhan 1998).

The same change also took place in the Iranian languages. Western Middle Iranian lost all the case distinctions of Old Iranian, and all the adverbial cases had to be realized using prepositions, giving rise to the Modern Persian system of prepositions (Hewson & Bubenik 2006: 131ff.). The preverbs that were closely associated with verbs became verbal prefixes. Most of these prefixes were used to encode path. Kent (1950) lists thirteen prefixes for Old Persian, while Windfuhr (2009) lists six for Western Middle Iranian and five for Modern Persian. Of these five Modern Persian preverbs, three are related to Proto-Indo-European preverbs: *bar* ‘up’, *bāz* ‘re-, again’ and *farā* ‘forth’. A reduced version of the preverb system is therefore still in place in Modern Persian. However, most spatial preverbs have become obsolete or have merged with verb roots, in some cases having created path verbs such as *āvordan* ‘to bring’. As is well known, the most productive system to create verbs in Modern Persian, including path verbs, is the compounding of nouns, adverbs and prepositions with light verbs to create compound verbs (Lambton 1953: 85ff., Mahootian 1997: 283ff.).

For Homeric Greek, it is clear that the free word order of the path encoding adverbial particles becomes more fixed over time (Schwyzer 1950). This leads to the

development of the prepositional phrase in Classical Greek (Hewson & Bubenik 2006: 59ff). The preverbs also underwent processes of univerbation with the verb root (Skopeteas 2002: 164ff., 349ff.). For Homeric Greek, Imbert (2010) lists fourteen adverbial path particles that were productive as adverbs, preverbs and adpositions (*amphí, aná, apó, diá, eis, ek, en, epí, hupér, hupó, katá, pará, perí* and *pró*). For Modern Greek, Holton, Mackridge & Philippaki-Warburton (1997: 180) list nine of these that may function as prefixes on verbs and as prepositions (*ana, apó, dia, ek, huper, hupo, katá, pará* and *pró*). It is clear that many Modern Greek verbs are the result of univerbation between preverbs and verb roots that took place in the development from Homeric Greek to Modern Greek (e.g. the range of verbs derived from *bállo* ‘to throw, to put’ listed by Holton, Mackridge & Philippaki-Warburton 1997: 180). Many of the preverb-verb combinations no longer have compositional meanings, suggesting that these have become unified lexical elements (Skopeteas 2002). Modern Greek is said to be verb-framed (Papafragou, Massey & Gleitman 2002, 2006) or mixed (Talmy 2007: 105, Skopeteas 2002: 349, Hickmann et al. under review).

Although not much is known on motion event encoding in Hittite, work by Brosch (2013) and Junghänel (in preparation) suggests that Hittite mostly preserved the Proto-Indo-European system of free path particles and thus was satellite-framed. However, Brosch (2013: 442) notes that even though Hittite is mostly satellite-framed, there are at least two path verbs, which may suggest an ongoing change from a satellite-framed to a verb-framed system. In addition, he also claims that Hittite has a very restricted manner verb lexicon, while it has a rich lexicon of verbs that encode deictic information (p. 326–327). Junghänel (in preparation) tentatively notes that Lydian (an Anatolian language attested in 100 BCE) employs a system of path prefixes, suggesting a potential change towards a verb-framed system for the Anatolian languages.

There is some information on Armenian from Wälchli’s (2009: 215) study of lexicalization patterns that suggests that Classical Armenian was already verb-framed to the same extent as the modern Romance languages. This suggests that a change from a hypothetical satellite-framed system in Proto-Indo-European to a verb-framed system must have been completed before the arrival of Classical Armenian. This is also suggested by Schmitt (1981: 86), who points out that preverb-verb combinations only rarely feature in Classical Armenian. Rather than fusing its preverbs with verbs in a later stage, the productive use of preverbs was already obsolete in Classical Armenian.

The shift from a preverb system to a prepositional system is attested in Albanian as well, suggested by several adverbs and prepositions that are cognate with Proto-Indo-European preverbs (*jashtë* ‘out’, *ndë* ‘into’, *nga* ‘from’, *para* ‘before’, *për* ‘for’, etc.). Albanian also has a range of verbal path prefixes that are cognate with the

preverbs recognized in Homeric Greek, Latin and Sanskrit (Orel 2000: 167). Orel (2000) provides evidence that, also in parallel to Modern Greek and the Romance languages, a process of univerbation of verbal prefixes with verb roots has taken place. Motion verbs in which this process can be detected are *ndjek* ‘to follow’, *ngre* ‘to lift’ and *përshkoj* ‘to go through’. Prefixation is still a productive means for verb derivation, although the majority of the prefixes used today are innovations, and none have spatial meanings (Camaj 1984: 208ff.).

Univerbation between preverbs and verb roots seems also to have taken place in Celtic. The preverb system gave rise to both a system of prepositions and verbal prefixes in Old Irish (Pokorny 1925: 98ff.). Pokorny lists 22 preverbs that functioned as verbal prefixes, many of which had cognate prepositions. The prepositional system is still in place in Modern Irish, but none of the verbal prefixes are still productive. The remnants of this system can be found in certain motion verbs where prefixes merged with verb roots, such as Irish *fág* ‘leave’ and Gaelic *fuadaich* ‘drive away’.

For the Balto-Slavic languages, the preverb system has been grammaticalized into a system of prefixes that are inseparable from the verb and that often have corresponding prepositions (Miklosich 1868: 195ff). This system is still in place in the modern Baltic and Slavic languages. The majority of these prefixes are cognate with preverbs in Indo-European languages such as Sanskrit and Homeric Greek. Slobin (2005) and Croft et al. (2010) report that although the Slavic languages are satellite-framed, the path prefixes are merging with verb roots, deriving verb-framed-like patterns or complex verbs that denote manner and path at the same time. The tendency of prefix-verb combinations to grammaticalize into monomorphemic verbs is attested not only for prefixes that denote path, but for all prefixes (Townsend 1968: 116–134, Bielec 1998: 73).

In Germanic, the free adverbial particles developed into a system of prepositions and separable and inseparable verbal elements that is still seen in modern Germanic languages (Roberts 1936, Goetz 2006), which are satellite-framed (Slobin 1996, 2005).

Taking these data on diachronic change in Indo-European languages, it seems clear that the last stage of Croft et al. (2010)’s first grammaticalization pathway (cf. 3a) is attested in all Indo-European subgroups. The preverbs that were used in satellite-framed constructions merged with verb roots to create path verbs. This diachronic change is found in all Indo-European subgroups, albeit to different extents. In the Romance languages and the Indo-Iranian languages, this change has led to the emergence of path verbs and the increased use of the verb-framed construction, while in Germanic and Balto-Slavic it has resulted in some univerbated verbs (such as Dutch *binnen-vallen* ‘lit. inside-fall, i.e. to visit unexpectedly’), but it has not led to a shift away from using the satellite-framed construction.

The data presented in this Section lend support to the claim that Proto-Indo-European could make use of the satellite-framed construction, as there is evidence for the existence of the satellite-framed construction in Latin, Hittite, Ancient Greek, Old Irish and Vedic Sanskrit. This is also why Talmy (2007) and Acedo Matellán & Mateu (2008) have proposed that Proto-Indo-European was satellite-framed. However, it remains unclear whether the satellite-framed construction was the most frequently used motion encoding construction in these languages. As we have seen, Brucale, Iacobini & Mocciaro (2011) and Brucale (2011) point out that in Latin, the satellite-framed construction was not used as often as suspected earlier, and Brosch observes that in Hittite, the manner verb lexicon was quite small (path satellites were most often used in combination with deictic verbs). Therefore, the overview of historical-comparative data on motion in the ancient Indo-European languages presented in the current Section gives rise to the alternative hypothesis that Proto-Indo-European was typologically mixed, rather than strictly satellite-framed.

In the remainder of this article, the behavior of Proto-Indo-European is investigated further using data from contemporary languages and ancient languages. The dataset consists of sophisticated and detailed data on the use of motion event encoding constructions and has been checked by native speakers (a type of data which is unavailable for any of the ancient Indo-European languages). It allows me to provide a first impression of the usage of different motion event encoding constructions in Proto-Indo-European, and thus goes beyond claiming that Proto-Indo-European could employ the satellite-framed construction, which is without any doubt given the historical-comparative evidence. The contemporary dataset is introduced in the next Section.

3. The contemporary dataset on Indo-European motion event encoding

3.1 The parallel corpus

The current dataset is a parallel corpus of motion events in twenty Indo-European languages (English, Dutch, German and Swedish from Germanic, Irish from Celtic, Portuguese, French, Italian and Romanian from Romance, Russian, Polish, Serbo-Croatian, Latvian and Lithuanian from Balto-Slavic, Hindi, Nepali and Persian from Indo-Iranian, Albanian, Armenian and Modern Greek) taken from two novels: *Alice's Adventures in Wonderland* by Lewis Carroll and *O Alquimista* 'The Alchemist' by Paulo Coelho. These two novels were chosen because they have been translated into a wide range of Indo-European languages, and their continuing popularity enabled their easy acquisition.

All descriptions of motion events were extracted from these two books. Motion events were defined as situations in which an animate or inanimate being moved from one place to another (Özçaliskan & Slobin 2003:259). Each motion extract that was picked constitutes a single sentence in which (approximately) a single situation (event or activity) was described (Berman & Slobin 1994:657). This selection procedure resulted in a full list of all motion event descriptions in the two novels. From this list, a sample of motion sentences was manually selected that included all the variation that was present in the larger collection, i.e. of each manner verb and each path verb that occurred in the larger sample, at least one instance was included in the smaller sample. The smaller set of selected motion sentences amounts to 118 sentences that encode voluntary (non-causative) motion. The total set of data thus consists of 118 original motion event extracts and their translations in a total of twenty languages.

After the sample of motion event descriptions was decided upon, each sentence was glossed and analyzed with the assistance of a native speaker or a language specialist. The Leipzig glossing rules were used as guidelines for the glossing. The categorization of each motion verb as a manner verb, path verb, deictic verb, manner plus path verb, or a neutral verb was carried out with the assistance of a native speaker.

The coding of the verbs was done using the following definitions. Manner verbs refer to the way a person or an object moves, including the rate (*walk*), the rhythm, the motor pattern (*run*), the posture (*crawl*) and any evaluative factors that might be involved with the movement (*sneak*) (Slobin 2004:255). Path verbs refer to the path or trajectory that a person or an object has while moving, typically with reference to people or objects in the environment (*enter*, *cross*). Deictic verbs refer to the trajectory of the person or object as seen from the perspective of the speaker or the addressee. In English, *come* typically signifies movement towards the speaker, while *go* typically signifies movement away from the speaker. Manner plus path verbs refer to manner and path at the same time. Examples are Greek *skarfalóno* 'climb up' and Persian *gorik̄tan* 'run away'. Neutral verbs refer either to movement that does not incorporate manner, path or deixis, such as *move* or *travel*, or that do not refer to movement at all, such as *find oneself* or *continue*.

Other components of motion event encoding that were coded included path satellites and alternative manner expressions. Path satellites are defined as all non-predicative elements that indicate (a part of) the path of the moving person or object. This includes adpositions, adverbs, case markers, verbal prefixes, etc., following the arguments for broadening Talmy (1985:102)'s original definition of 'path satellite' by Filipović (2007:35), Beavers, Levin & Tham (2010:337) and Croft et al. (2010:205–206). Alternative manner expressions can be adverbial expressions that refer to some aspect of the manner of motion or participles of manner

verbs. Examples of both types of alternative manner expressions are included in (7) below.

3.2 The motion event encoding constructions

The coding of the verbs and other motion event components resulted in a set of motion event encoding constructions that were identified in the dataset (see Verkerk 2014 for more discussion on these constructions). Most of these are well known from the motion event literature, although my definitions are generally narrower.

In the satellite-framed construction, a manner verb is used in combination with a path satellite (see example 6). In the verb-framed construction, a path verb is used in combination with an alternative manner expression (see example 7). In the path-only construction, a path verb is used without any indication of manner (see example 8). These three constructions are familiar from the motion event literature (Talmy 1985, 1991, Slobin 1996).

(6) Dutch

en toen ze terug-liep naar het
 and when 3SG.F.SBJ back-walk.PST.SG to DEF.ART
 tafel-tje om het te pakken (...).
 table-DIM in.order 3SG.N to take.INF
 ‘and when she walked back to the small table to get it (...).’

(7) Portuguese

Aproximou-se trotando a toda
 draw.closer.IND.PFV.3SG-REFL trot.PRES.PTCP at all.F
 a brida.
 DEF.ART.F.SG rein.F
 ‘He drew closer trotting at full speed.’

(8) French

Alors je vais faire le tour
 then 1SG.SBJ go.PRES.1SG do.INF DEF.ART.M tour.M
 et passer par la fenêtre.
 and pass.INF through DEF.ART.F window.F
 ‘Then I will go around and pass through the window.’

In certain translations, there is no element that encodes path. In the manner-only construction, just a manner verb is present, without any path information (see example 9).

(9) Armenian

na shtap-shtap t'rrch'kot-el-ov k'ayl-um er
 3SG.SBJ hurriedly-hurriedly trot-INF-INS walk-PRES.PTCP be.AUX.3SG.PST
 'He was hurriedly walking with a trot'

When the only verb in the sentence is a manner plus path verb, and there is no other information on the path of the motion, the manner plus path verb construction is used, as in (10):

(10) Italian

e scavalc-ò con un salto il
 and step.over-3SG.PST with one.M step.M.SG DEF.ART.M.SG
 primo dei sei piccoli ruscelli.
 first.M of.DEF.ART.M.PL six small.M.PL brook.M.PL
 'and stepped over the first of the six small brooks with one step.'

Several studies have pointed out that deictic verbs should not be placed in the category of path verbs, as has been done in the past (Berthele 2006, Wälchli 2009, Verkerk 2014). Deictic verbs refer to motion with respect to a specific deictic center, rather than to motion that has a certain path. In the deictic verb construction, a deictic verb is the only verb in the clause, and no manner information is provided (see example 11). When a deictic verb is used in combination with an alternative manner expression, the deictic verb-framed construction is used (see example 12).

(11) Persian

vali felāmingo-ye u be taraf-e
 however flamingo-of.EZ 3SG to side-of.EZ
 digar-e zamin-e bāzi rafte
 other-of.EZ ground-of.EZ playing go.PERF.3SG
 'however, her flamingo had gone towards the other side of the playing field'

(12) Irish

agus tháinig ag eitilt anuas uirthi
 and come.PST at fly.INF downwards on.3SG.F
 'and [the playing cards] came flying down upon her'

Aside from the motion event encoding constructions in which just one verb is present, there also exist two constructions that employ two verbs. The subordinate construction involves one main verb and one subordinate verb that both encode aspects of the motion that is encoded (see example 13 and Verkerk 2014). The coordinate construction, also discussed by Croft et al. (2010), involves two coordinated motion verbs (see example 14).

(13) Greek

To agóri kai o
 DEF.ART.N.NOM.SG boy.N.NOM.SG and DEF.ART.M.NOM.SG
 Ágglos eíkh-an agorá-sei
 Englishman.M.NOM.SG AUX.PST.PRF-3PL buy-NFIN
 kamēl-es kai duskoleú-tēk-an n' anéb-oun.
 camel-F.ACC.PL and have.difficulty-PST.PFV-3PL to go.up.DEP-3PL
 'The boy and the Englishman had bought camels and had difficulty to ascend them.'

(14) Hindi

sabse āge tair rahī thī
 of.all.SUP in.front.ADV swim-INF PROG.F.SG be.AUX.PST.F.SG
 aur yah pūrī jamāt tālāb
 and 3SG.PROX complete.ADJ.F party.F pond.M
 ke kināre kī taraf baḥ rahī thī
 GEN.M.OBL bank.M.OBL GEN.F side.F proceed-INF PROG.F.PL be.AUX.PST.F.PL
 'she was swimming in front of everyone and the whole party was proceeding to the bank of the pond'

An overview of the frequency of these various motion encoding constructions as they are attested in the 118 sentence sample is given in Figure 1.

Talmy (1985:62) categorizes languages with respect to their most characteristic motion event encoding construction, meaning that the most characteristic construction is the one that is colloquial, most frequent and pervasive. However, recent studies (Beavers, Levin & Tham 2010, Croft et al. 2010) point out that languages often use a range of different means to encode motion events, and reducing this variety to a single classification does not take into account the attested variability in motion event encoding. The same conclusion is evident from Figure 1: most languages make use of a multitude of constructions, but do so to different extents.

The languages in Figure 1 have been ordered so that a cline in the use of the satellite-framed construction becomes evident: the Balto-Slavic (except Serbo-Croatian) and Germanic languages use the satellite-framed construction the most, while Armenian, Albanian and the Indo-Iranian languages Hindi, Nepali and Persian use the satellite-framed construction the least. The Romance languages are in between these two groups. The path-only construction and the verb-framed construction are used most often in Serbo-Croatian, Greek, Albanian and the four Romance languages Italian, Portuguese, Romanian and French. The deictic verb construction is used most often in Irish, the Germanic languages Swedish, English and Dutch and the Indo-Iranian languages Hindi, Nepali and Persian. Most Balto-Slavic languages do not have deictic verbs (see Verkerk 2014 for more discussion

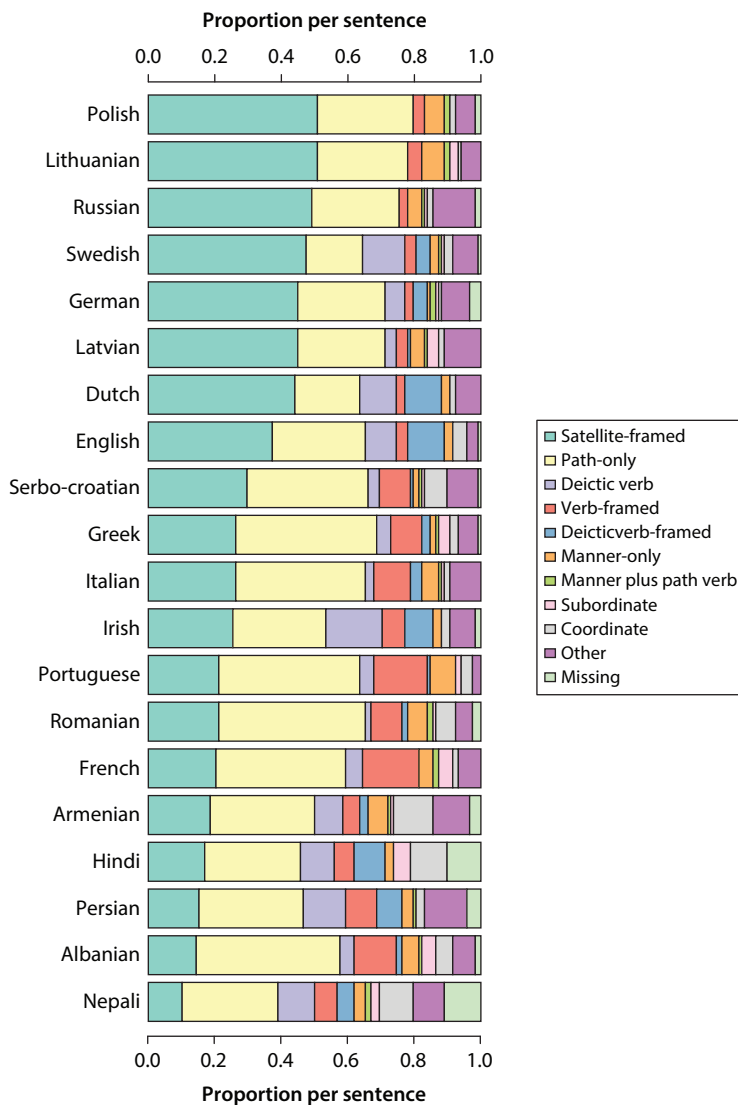


Figure 1. Proportions of the different motion event encoding constructions in twenty Indo-European languages based on the 118 original motion extracts and their translations

on deictic verbs in these languages). The deictic verb-framed construction is only used by languages that also use the deictic verb construction relatively frequently. The coordinate construction and the subordinate construction are used to some extent in Hindi, Nepali and Armenian. The remaining constructions are all quite uncommon throughout the sample. The motion encoding patterns found in

Figure 1 correspond to those that have been described in the literature on motion events for several languages from the sample (Verkerk 2014).

In spite of the variability in the use of motion event encoding constructions evident from Figure 1, it is possible to identify the two traditional Talmyan languages types. On the upper side of the plot we find languages that use the satellite-framed construction often, around 40–50% of the time. In the remainder of this article, I follow Slobin (2004) and call these languages *MANNER-SALIENT*, reflecting a partial habit to encode manner on the main motion verb. Clear manner-salient languages are Russian, Dutch, Polish, Lithuanian, Swedish, German, English and Latvian. On the lower side of the plot we find languages that use the satellite-framed construction less often and that use the path-only construction and verb-framed construction more often. I call these languages *PATH-SALIENT*, reflecting their partial tendency to encode path on the main motion verb. Clear path-salient languages are Greek, Italian, Romanian, French, Portuguese and Albanian. The terms *MANNER-SALIENT* and *PATH-SALIENT* are chosen because it seems clearer to refer to types of languages in this way, rather than by using the name of the construction most commonly used in that language.

However, there are also languages that do not really fit into one of these two traditional classes: Serbo-Croatian, Irish, Hindi, Nepali, Persian and Armenian. The first language, Serbo-Croatian, is currently undergoing a change from manner-salient to path-salient motion event encoding, hence its intermediary position in Figure 1. The latter five languages make relatively much more use of either the deictic verb construction or the coordinate construction, or both. The fact that these languages are difficult to place in one of Talmy's traditional categories, as well as the variation between the languages that can be placed in Talmy's traditional categories, supports the view that motion encoding systems should not simply be generalized in a two-way dichotomy.

4. Methodology

The evolutionary behavior of a range of measures describing motion event encoding in the set of Indo-European languages was investigated. Principal component analysis was used to reduce the number of dimensions present in the dataset (Section 4.1). A phylogenetic measure of relatedness was used in order to reconstruct Proto-Indo-European motion event encoding and identify typological changes in motion event encoding. This measure of relatedness, which takes the form of a set of phylogenetic trees, is introduced in Section 4.2. The phylogenetic comparative methods that were employed to perform ancestral state estimation are introduced in Section 4.3.

4.1 Data reduction

In Section 3.2, it became clear that the attested variation in motion event encoding could not easily be reduced by assigning each language to a single typological class. However, the phylogenetic comparative methods that are employed in this article required the aggregation of the data into a single value for each language. In order to extrapolate a single value that characterizes the motion encoding pattern of each language, a principal components analysis was conducted. Principal components analysis is a ubiquitous data reduction technique that is used for comparative analysis in linguistics (examples are Baayen 1994, Calude & Pagel 2011, Grieve 2014). These studies use principal components analysis as a tool to reduce the dimensionality of their measures and to find a single measure that captures most of the variance. Since the use of the different motion event encoding constructions appeared to covary with one another, a principal components analysis was used to reduce the dimensionality present in the dataset.

The principal components analysis was performed on the proportion of usage of each of the nine motion event encoding constructions as presented in Figure 1. Since the genetic relationships between these languages are likely to explain part of the variance present in the data, I used the phylogenetic principal components analysis proposed by Revell (2009), which removes part of the variance that can be attributed to these relationships. The results are graphically depicted in Figure 2.

Figure 2 gives the scores of the twenty languages on the first principal component (1st PC) on the x-axis and the scores on the second principal component (2nd PC) on the y-axis. The 1st PC accounts for 79.1% of the variance and can be interpreted as analogous to the Talmy typology: languages to the right use the satellite-framed construction more often, while languages to the left use the path-only and verb-framed construction more often. The 2nd PC accounts for 9.8% of the variance and relates the amount of use of the deictic verb construction and the deictic verb-framed construction, with languages that sometimes use these two constructions situated at the bottom of the plot. These two principal components together account for 88.9% of the variance. This indicates that the cline based on the Talmy typology and the use of deictic verbs are the two main parameters that characterize motion event encoding in this sample.

The score of each language on the 1st PC is taken to be the position of that language on a Talmian scale that reaches from a maximally path-salient character on the left side of the scale and a maximally manner-salient character on the right hand of the scale. The score of each language on the 2nd PC is used as a measure of how often deictic verbs are used. These scores, together with the proportions of construction usage as presented in Figure 1, are used for the

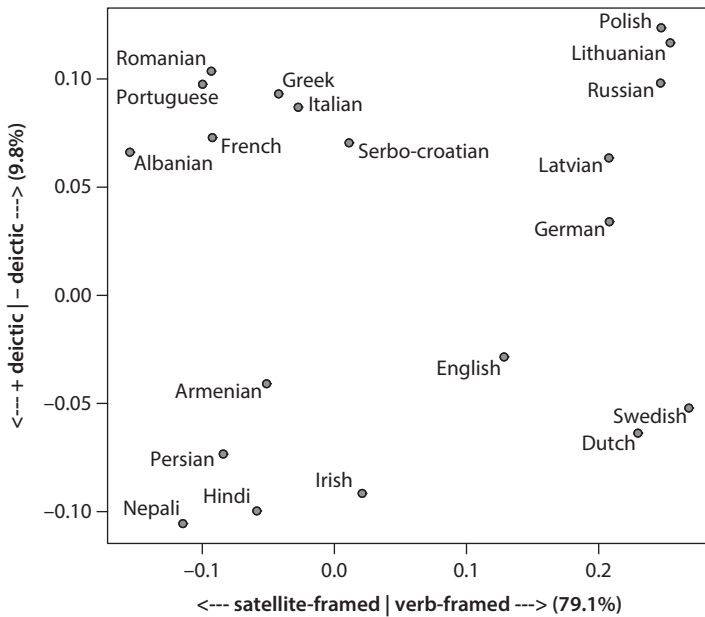


Figure 2. A principal components analysis conducted on the proportion of usage of the nine motion event encodings construction for twenty Indo-European languages

phylogenetic comparative analyses in Section 4.3. Appendix 2 provides a table with the scores on the 1st PC, the 2nd PC and the proportions of construction usage for each language.

4.2 The phylogenetic tree sample

A set of phylogenetic trees was used as a representation of the history of the Indo-European language family. This tree set was taken from Bouckaert et al. (2012) and was constructed on the basis of lexical data. Bouckaert et al. constructed a large dataset of cognate-coded lexical data for 103 languages. This dataset was originally based on the dataset by Dyen, Kruskal & Black (1992) but was highly improved by checking cognate judgments and adding languages. This dataset was then recoded in a binary fashion, in which each language was characterized to have a cognate (1) or not (0) for each of the cognate sets. Subsequently, posterior tree distributions were estimated using a Bayesian Markov Chain Monte Carlo approach (Huelsenbeck et al. 2001) available in the software BEAST (Drummond et al. 2012). Sampling from the posterior distribution of estimated trees generated a sample of 12,500 phylogenetic trees. The complete set of cognates used in this type of analysis is never completely compatible with a single tree, but is typically highly compatible with a range of different trees, which

often partly feature the same clades (subgroups). Using a sample of trees rather than one individual tree allows one to account for the uncertainty that is a part of every phylogenetic estimation.

I took Bouckaert et al.'s (2012) sample of trees and generated a sample of 1,000 trees for the twenty languages by randomly selecting 1,000 phylogenetic trees and removing from this sample the languages not represented in the current dataset. A maximum clade credibility tree was calculated using TreeAnnotator v.1.6.1 (Drummond et al. 2012) and presented in Figure 3, but the phylogenetic comparative analyses were conducted over all 1,000 trees in order to account for uncertainty in the phylogenetic tree model. The numbers on the branches of the phylogenetic tree in Figure 3 are support-values that indicate how often each clade is attested in the tree sample. For instance, the number for the node that relates Albanian, Modern Greek and Armenian, with a value of 0.38, indicates that this clade is attested in 38% of the trees in the tree sample of 1,000 trees. The lengths of the branches of the tree are a measure of the amount of evolution that has taken place. A longer branch indicates more lexical divergence from the root node, while a shorter branch indicates less lexical change away from the root node.

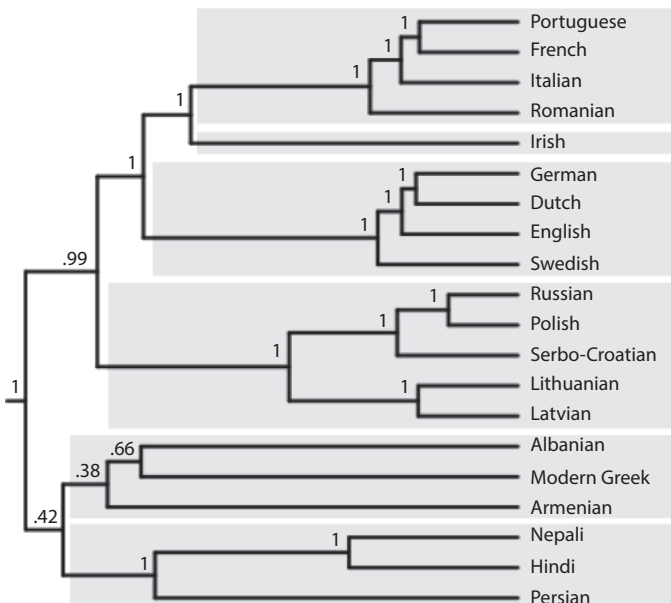


Figure 3. A maximum clade credibility tree for the tree sample taken from Bouckaert et al. (2012)

4.3 Phylogenetic comparative analyses

First, the presence of phylogenetic signal was investigated. When phylogenetic signal is present, this implies that languages behave similarly due to genetic descent. If it is not present, a range of different processes could be involved to generate the data, including lack of variation in the feature, parallel change with other linguistic features, very rapid linguistic change or borrowing.

The presence of phylogenetic signal was tested using the parameter λ (lambda) (Pagel 1999, Freckleton, Harvey & Pagel 2002). The parameter λ estimates to what extent the data are evolving exactly as the phylogeny would predict. A perfect match between data and phylogeny would mean that language A would be exactly as similar to language B, as would be predicted by the distance between language A and B on the phylogeny. A maximum value of λ (which is typically 1 but may be higher due to features of the phylogenetic tree) indicates that a trait is evolving exactly as predicted by the phylogeny under a random walk model of evolution, while a value of $\lambda = 0$ would indicate that the evolution of the trait is taking place entirely independent from the phylogeny, i.e. completely random with respect to history (Freckleton, Harvey & Pagel 2002).

The model that is used to estimate λ is a random walk model of evolution, also sometimes called the ‘Brownian motion process’. This model allows traits to evolve with a mean change of zero and a constant but unknown variance in each unit of evolutionary “time.” Time may be measured in various ways, for instance in terms of lexical divergence. The evolutionary process that takes place at each unit of “time” unfolds independently from former or future evolutionary processes and occurs along the branches of the phylogeny (Pagel 1999).

The parameter λ can be estimated by maximum likelihood approaches implemented in various software. Maximum likelihood approaches to phylogenetics describe the likelihood that a certain evolutionary process has given rise to the observed data as opposed to another process generating the observed data. This is useful because it is possible to compare likelihoods across different analyses and evolutionary models. The estimate of λ can be tested for significance using the log-likelihood ratio test that compares the estimated value of λ to the likelihood of a model of evolution in which λ is 0 or 1 (Freckleton, Harvey & Pagel 2002). If λ is not significantly different from a model of evolution in which λ was set to 1, and significantly different from a model of evolution in which λ was set to 0, it can be argued that phylogenetic signal is clearly present. For the current dataset and the phylogenetic trees described above, λ was estimated using the function `phylosig`, part of the R (R Development Core Team, 2011) package `phytools` (Revell 2012). The estimations of the likelihood of a model in which λ was set to 1 and 0

were conducted using the function `fitContinuous`, part of the R package `GEIGER` (Harmon et al. 2008).

Ancestral state estimation (ASE) was also performed using a maximum likelihood approach, using the random walk model of evolution (Schluter et al. 1997, Webster 2002). The estimation of ancestral states is a two-step problem. First, ancestral states for each internal node in the tree are computed as weighted averages of the values of the languages that are connected by that node. The weights are inversely proportional to the branch length that connects the internal node and the descendent language or node, so that descendant languages or nodes that are connected to the internal node with short branches have a larger weight as descendant languages or nodes that are connected to the internal node with longer branches. Second, an algorithm moves from the root towards the languages at the tips of the tree, adjusting the ancestral states of the internal nodes in such a way that the complete phylogeny is considered in the estimation of the most likely ancestral states. This algorithm minimizes the sum of the square of the weighted differences between ancestral nodes and descendent languages and nodes over the whole tree. ASE was conducted using the function `getAncStates`, part of the R package `geiger` (Harmon et al. 2008). Supporting functions in the R package `ape` (Paradis et al. 2004) were used as well.

4.4 Other analyses

To measure the degree of dependency between languages due to closeness in geographic space, I conducted a partial Mantel test, which is a measure of spatial autocorrelation. The partial Mantel test was used to calculate the correlation between the motion encoding system and geographical distance while taking into account the phylogenetic distance. The score on the 1st PC was used as a characterization of the motion encoding system. The latitudes and longitudes of the capitals of the nations in which the languages in the sample are spoken were used as a measure of geographical distance. The length of the branches between each set of languages on the maximum clade credibility tree presented in Figure 3 was used as a measure of phylogenetic distance.

First, the partial Mantel analysis tests for the correlation between the three different distance matrices. Second, it tests whether this correlation is statistically significant by permuting the matrices and estimating the correlation repeatedly to compare the original test statistic to the distribution of test statistics from the permutations. In this way, a simulated p-value is generated. The Mantel test was conducted using the function `mantel.partial`, part of the R package `vegan` (Oksanen et al. 2012).

5. Results

The test to determine whether a phylogenetic signal was present was conducted on the 1,000 trees sample. The score on the 1st PC was used as a holistic measure of motion event encoding for the twenty languages of the sample. Because a tree sample is used rather than a single tree, we do not have a single estimate of λ , but a range of 1,000 estimates of λ . The parameter λ was estimated to have a range of 0.95–1.28, with a median of 1.16.¹ An overview of the lambda estimates is included in Figure 4. The estimated lambdas were significantly different from 0 for all 1,000 trees (median $p = 4.4 \times 10^{-6}$, range 4.4×10^{-7} –0.0005). Lambda was estimated not to be significantly different from 1 in a large portion of the trees (the p-value ranged between 0.001–0.04 for 914 trees, and ranged between 0.06–1.00 for the other 86 trees).² This indicates that there is a clear phylogenetic signal in these data.

The results of the ancestral state estimation analyses for Proto-Indo-European are presented in Table 1. Ancestral state estimations were made for a range of different measures, including the score on the 1st PC, the score on the 2nd PC, and the proportion of use for the five most frequently used motion encoding constructions: the satellite-framed construction, the path-only construction, the deictic verb construction, the verb-framed construction, and the deictic verb-framed construction. Since the analyses were conducted over a range of 1,000 trees, the estimates again are ranges rather than single numbers. In Table 1, the median of the ancestral state estimation analyses over all 1,000 trees in the sample is provided first, followed by the range of ancestral state estimations between brackets.

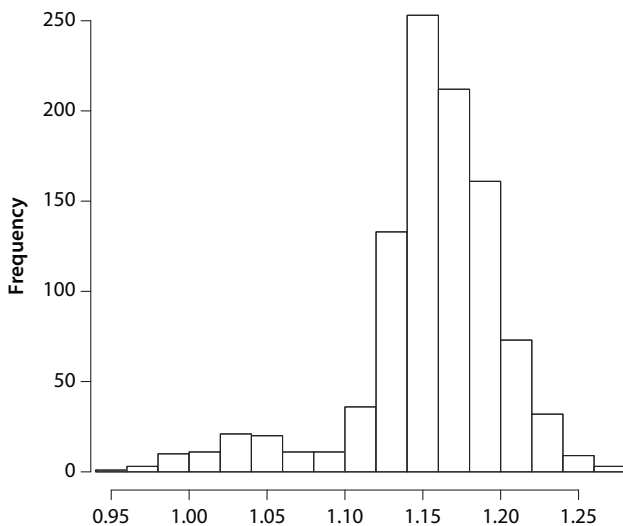


Figure 4. Estimated λ values for the scores on the 1st PC for the 1,000 tree sample

Table 1. Estimated ancestral states for Proto-Indo-European principal component scores and proportions of construction usage

Measure	Median
Score on the 1st PC	-0.001 (-0.03–0.02)
Score on the 2nd PC	0.004 (-0.01–0.01)
Proportion of use of the satellite-framed construction	0.26 (0.23–0.28)
Proportion of use of the path-only construction	0.33 (0.32–0.34)
Proportion of use of the deictic verb construction	0.07 (0.07–0.08)
Proportion of use of the verb-framed construction	0.07 (0.07–0.08)
Proportion of use of the deictic verb-framed construction	0.04 (0.04–0.05)

The results in Table 1 indicate that Proto-Indo-European is estimated to be a type of language that is intermediary on the Talmian scale: the ancestral states inferred for the proportion of use of the different motion event encoding constructions are in all cases quite close to the mean values presented in Appendix 2. There seems to be a slight preference for the path-salient side of the scale, as the use of the satellite-framed construction (0.26) is estimated to be slightly lower as the mean value (0.31), and the use of the path-only construction is estimated to be slightly higher (0.33 compared with 0.32). The score on the first principal component (-0.001 compared with 0.05) is also directed in favor of the path-salient side of the scale. The language that seems to come as close as possible to the estimated values is Greek. Greek is path-salient, but is also very close to the middle of the scale of manner-salient and path-salient languages that was presented in Section 3.

Note that the estimates for the deictic verb construction and the deictic verb-framed construction cannot be interpreted in a meaningful way, as most ancient Indo-European languages and Proto-Indo-European do not have (reconstructed) deictic verbs. The reconstructed Proto-Indo-European verb *ei- ‘to go’, for instance, is not reconstructed to be a deictic verb, but is rather a general motion verb that means ‘to walk, move, go’ (Pokorny 1959–1969:293–297). This indicates that the results of an ancestral state estimation analysis can never be interpreted without the linguistic knowledge that we have about these languages, such as the presence of (reconstructed) deictic verbs that are crucial for these two constructions.

Separate ancestral state estimation analyses for the major Indo-European subgroups Germanic, Romance, Balto-Slavic and Indo-Iranian were conducted as well. The results are presented in Table 2, in which the median of the ancestral state estimation analyses over all 1,000 trees in the sample is provided first, followed by the range of ancestral state estimations in parentheses.

Table 2. Estimated ancestral states for principal component scores and proportion of construction usage of the four most frequently used constructions for the roots of four subgroups: Germanic, Romance, Balto-Slavic and Indo-Iranian.

Subgroup	Germanic	Romance	Balto-Slavic	Indo-Iranian
Score on 1st PC	0.19 (0.17–0.19)	–0.06 (–0.04–0.07)	0.14 (0.11–0.16)	–0.04 (–0.06–0.02)
Score on 2nd PC	–0.03 (–0.03–0.02)	0.07 (0.07–0.08)	0.07 (0.06–0.08)	–0.03 (–0.06–0.02)
Satellite-framed construction	0.41 (0.41–0.42)	0.24 (0.23–0.25)	0.40 (0.37–0.42)	0.21 (0.18–0.23)
Path-only construction	0.24 (0.23–0.24)	0.40 (0.39–0.40)	0.30 (0.29–0.31)	0.32 (0.31–0.33)
Deictic verb construction	0.10 (0.09–0.10)	0.04 (0.04–0.05)	0.03 (0.02–0.04)	0.09 (0.08–0.10)
Verb-framed construction	0.04 (0.03–0.04)	0.12 (0.11–0.12)	0.06 (0.05–0.06)	0.08 (0.07–0.08)

The scores on the 1st PC and the median of the ancestral values estimated for the score on the 1st PC for Germanic, Romance, Balto-Slavic, Indo-Iranian and Proto-Indo-European have been placed upon the maximum clade credibility tree in Figure 5.

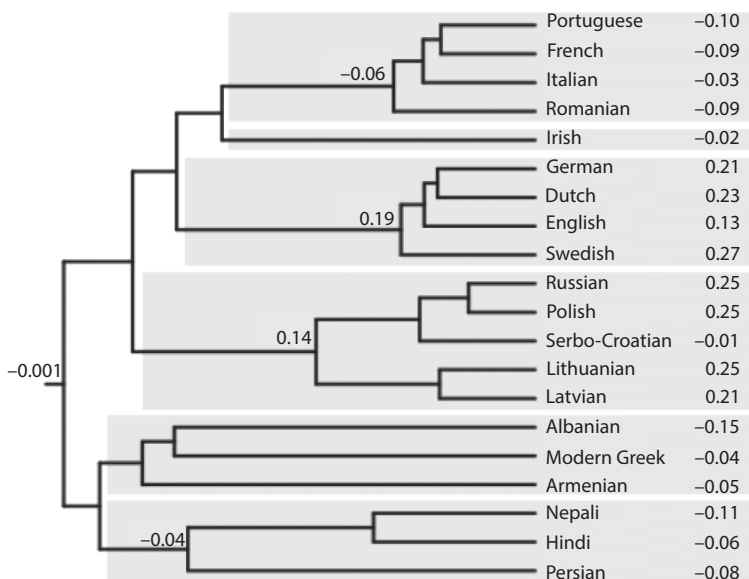


Figure 5. Scores on the 1st PC placed on the maximum clade credibility tree based on Bouckaert et al. (2012). Ancestral state estimates of the score on the 1st PC are added for Proto-Indo-European and the Germanic, Romance, Balto-Slavic and Indo-Iranian subgroups

To test how robust these results are, additional ancestral state estimation analyses were conducted. The first robustness test was performed to investigate whether the estimated ancestral state for Proto-Indo-European would shift when information on ancient languages was incorporated. To this end, the 1,000 phylogenetic tree sample from Bouckaert et al. (2012) was pruned again to include four ancient languages: Classical Armenian, Ancient Greek, Vedic Sanskrit and Latin (see the maximum clade credibility tree in Appendix 3). Since there are no translations of the two novels that form the parallel corpus on which this study is based in these four ancient languages, it is only possible here to provide an approximation of the behavior of these ancient languages on the measures used in this article, which is based on the literature described in Section 2. This was done by assigning each of these ancient languages the scores of one of the modern languages that it is claimed to resemble. This procedure is of course almost entirely arbitrary, as it can only provide a very tentative characterization of motion event encoding in these four ancient Indo-European languages. However, it serves as an illustration of the way in which contemporary and ancient data (when available) may be integrated in future work.

Classical Armenian has been described as verb-framed (Wälchli 2009:215), and therefore it seemed appropriate to assign Classical Armenian the same values as Modern Armenian. In Section 2, the manner-salient nature of Homeric Greek and Vedic Sanskrit, which featured freely moving preverbs, and Latin, which possessed a fully productive set of path prefixes, has been discussed to some extent. I assigned these three ancient languages the values of German, which is clearly manner-salient, but not as radically as some of the Balto-Slavic languages. The results of the ancestral state estimation analyses on the tree sample that incorporated these four ancient Indo-European languages are presented in Table 3.

Table 3. Estimated ancestral states for Proto-Indo-European incorporating information on four ancient Indo-European languages

Measure	Median
Score on the 1st PC	0.06 (0.03–0.08)
Score on the 2nd PC	0.02 (0.003–0.03)
Proportion of use of the satellite-framed construction	0.31 (0.29–0.33)
Proportion of use of the path-only construction	0.31 (0.30–0.33)
Proportion of use of the deictic verb construction	0.07 (0.06–0.08)
Proportion of use of the verb-framed construction	0.06 (0.06–0.07)
Proportion of use of the deictic verb-framed construction	0.04 (0.04–0.05)

Table 3 indicates that the estimates for Proto-Indo-European, compared with those in Table 1, become oriented slightly towards the manner-salient end of the

Talmian scale. The satellite-framed construction is used slightly more often (0.31 compared with 0.26 in Table 1) and the score on the 1st PC is slightly higher (0.06 compared with -0.001 in Table 1). However, compared with the mean values in Appendix 2, the estimation is still intermediary: the score on the 1st PC (0.06 compared with 0.05) is still in the middle of the Talmian scale, although it is now on the manner-salient side of the scale. The same applies for the values for the satellite-framed construction and the path-only construction.

A second robustness test was conducted by doing ancestral state estimation on four different phylogenetic trees. This is useful because there is uncertainty about the higher-order subgrouping of Indo-European. Testing whether different higher-order subgroupings affect the ancestral state estimations allows us to assess if the intermediary estimation for Proto-Indo-European is a robust result or whether it is caused by the specific higher-order subgrouping of Indo-European present in the current tree sample.

The following four trees were used for this robustness test. Three of the trees were maximum clade credibility trees, constructed using subsets of the tree sample used by Dunn et al. (2011). This tree sample is based on a slightly older version of the cognate data used by Bouckaert et al. (2012). Since this tree sample includes more varied tree topologies, several maximum clade credibility trees were used for additional ancestral state estimation analyses. These included: (i) the full tree set (built from the full set of 1,000 trees), (ii) the sample of trees in which Balto-Slavic and Germanic form a clade (built from the 21 trees in which this subgroup existed), and (iii) the sample of trees in which Balto-Slavic and Romance form a clade (built from the 32 trees in which this subgroup existed). The fourth tree is the best phylogenetic tree of the Indo-European language family presented in Nakhleh, Ringe & Warnow (2005), their “tree A”. These four phylogenetic trees are visually presented in Appendix 3.

The results of these analyses are presented in Table 4, which gives an overview of the estimated ancestral states for Proto-Indo-European, using the scores on the 1st PC and the 2nd PC and the proportion of the most frequently used motion encoding constructions as measures. A comparison between the estimates in Table 4 to the ranges presented in Table 1 indicates that the use of these different phylogenetic tree topologies results in an intermediary/manner-salient estimation for Proto-Indo-European. The score on the 1st PC is still intermediary but is now oriented towards the manner-salient end of the scale (0.07–0.10, compared with -0.001 in Table 1). The scores on the satellite-framed construction are also higher (0.33–0.34, compared with 0.26 in Table 1), and the scores on the path-only construction are lower (0.28–0.31, compared with 0.33 in Table 1). However, this is by no means a very radical shift: these values suggest that Proto-Indo-European is positioned somewhere between English, the least radical manner-salient language

in the current sample, and Greek, the least radical path-salient language in the current sample (see also Figure 1). This indicates that the results presented in Table 1 are quite robust: different higher-order subgroupings of Indo-European do not have a large impact on the intermediary ancestral state estimate that was found for Proto-Indo-European.

Table 4. Estimated ancestral states for Proto-Indo-European for different phylogenetic trees.

Phylogenetic tree	Dunn et al.'s (2011) full tree set	Dunn et al.'s (2011) Balto-Slavic-Germanic tree	Dunn et al.'s (2011) Balto-Slavic-Romance tree	Nakhleh, Ringe & Warnow's (2005) tree
Score on 1st PC	0.08	0.09	0.10	0.07
Score on 2nd PC	0.002	0.0005	-0.006	0.03
Satellite-framed construction	0.33	0.34	0.34	0.33
Path-only construction	0.29	0.29	0.28	0.31
Deictic verb construction	0.07	0.07	0.08	0.06
Verb-framed construction	0.06	0.06	0.05	0.07
Deictic verb-framed construction	0.06	0.06	0.06	0.03

To test for areal patterns in the data, a partial Mantel test was carried out. The results of this test suggest that there is no positive correlation between the score on the 1st PC and the distance between the nations' capitals ($R = -0.08$, $p = 0.90$). These findings suggest that languages that are spoken more closely to each other are not more similar because of their spatial proximity.

6. Discussion

6.1 Interpreting the ancestral state estimation analysis

Using maximum likelihood ancestral state estimation analysis, I inferred the motion event encoding patterns of Proto-Indo-European as well as the ancestors of four large subgroups (Proto-Germanic, Proto-Romance, Proto-Balto-Slavic and Proto-Indo-Iranian). Ancestral state estimation of Proto-Indo-European indicates that the root of the Indo-European language family can be placed somewhere in the middle of the scale between path-salient and manner-salient languages. Robustness tests indicated that even though the estimated value is intermediary, it is oriented slightly towards the manner-salient end of the scale. How can this result be assessed in terms of what we know about the history of the preverb systems of these languages and the phylogenetic comparative method that was used?

The link between the scores and inferences presented in Tables 1–4, Figure 5 and the description of changes in the Proto-Indo-European preverb system as presented in Section 2 is clear. Those subgroups in which the Proto-Indo-European preverb system has been retained in the form of a productive set of path prefixes or particles, i.e. Germanic and Balto-Slavic, are manner-salient. Note, however, that the ancestral state estimate for Proto-Balto-Slavic presented in Table 2 and Figure 5 is 0.14, which seems to indicate a less manner-salient character as Proto-Germanic (0.19). This is due to the mixed or even path-salient nature of Serbo-Croatian, which is discussed further below. In contrast, the subgroups for which we have some evidence that this system was lost due to increasing unproductivity and the univerbation of prefixes with verb roots, i.e. Romance, Indo-Iranian, Greek and Albanian, are path-salient.

In Section 2, two hypotheses with regard to motion event encoding in Proto-Indo-European were put forward. On the one hand, Talmy (2007) and Acedo Matellán & Mateu (2008) have proposed that Proto-Indo-European was satellite-framed. On the other hand, the review of the historical evidence from ancient Indo-European languages indicated that although all ancient Indo-European languages could use the satellite-framed construction, this strategy might not have been the construction that was used most often (Brucale 2011, Brucale, Iacobini & Mocciaro 2011, Brosch 2013). The review of the historical-comparative data thus suggests that Proto-Indo-European was typologically mixed. Since the ancestral state estimation analyses indicate that Proto-Indo-European was mixed, but placed on the manner-salient side of the Talmian scale, they provide more support for the second hypothesis. Given these findings and the evidence from the historical record of ancient Indo-European languages, it seems clear that Proto-Indo-European was not manner-salient on the same level as the Germanic and most of the Balto-Slavic languages; rather, it had a mixed typological nature.

The estimated mixed motion event encoding system for Proto-Indo-European could in part be ascribed to the method used for ancestral state estimation. The maximum likelihood analysis conducted in this study has a tendency to return intermediate values:

A maximum-likelihood estimate of an ancestor state for a continuous trait is simply a weighted average of the dimensions of the extant species at the tips of the tree (eq. [4]). Reconstructed ancestor states will tend to be intermediate for this reason. [...] This analysis suggests that ancestor reconstructions for continuous traits are often too variable to be of much use, except to place ancestor sizes within broad limits. (Schluter et al. 1997: 1706)

Because of this characteristic of the method that was used, the current ancestral state estimation for Proto-Indo-European may be skewed to the middle of the

Talmian scale. However, the intermediate estimate for Proto-Indo-European is in line with the historical-comparative information reviewed in Section 2. Given the current findings and the historical-comparative record, Proto-Indo-European could in fact be a language that once was highly manner-salient but that was already losing the productivity of the preverb system. Some indications that this might be true have already been given in Section 2. More evidence can be found in the idiomatic meanings of preverb-verb combinations. Whitney (1879: 352) notes that preverb-verb combinations in Vedic Sanskrit may undergo a shift in meaning, although it is never so radical that the change cannot be related to the basic meanings of the preverb and the verb (see also Ringe 2006: 58–59). Danesi (2013: 62) gives several examples of preverb-verb idioms, including the preverb-verb combination *prá vac* ‘to announce, proclaim’, from *prá* ‘forward, fore’ and *vac* ‘to speak’. It seems clear that even though the preverb system was highly productive in the Rigveda, a process of lexicalization or univerbation in which the meaning of the preverb-verb combination was no longer transparent was already taking place.

Other evidence for this claim can be found in the etymology of Romance path verbs. These etymologies indicate that several preverb-verb univerbations were already completed in Latin, and the Romance languages only continued further on this path of univerbation. This seems to have been the case for the most common path verbs, such as French *approcher* ‘approach’, *arriver* ‘arrive’, *descendre* ‘come down’, *entrer* ‘enter’, and *retourner* ‘return’ (Brachet 1882) and their cognates in the other Romance languages. This might indicate that the preverb system that could be used for satellite-framed motion event encoding constructions was very widespread in Proto-Indo-European, but its productivity may already have been declining as Proto-Indo-European split up to give rise to the major Indo-European subgroups.

The hypothesis that Proto-Indo-European was satellite-framed or manner-salient (Talmy 2007, Acedo Matellán & Mateu 2008) is based only on the availability of the satellite-framed construction in all the ancient Indo-European languages. However, only information on how often this construction and other motion event encoding constructions are used can shed light on Proto-Indo-European motion event encoding. Although the choice was made to use contemporary data in combination with phylogenetic comparative methods in this article, corpus studies of individual ancient Indo-European languages and comparative parallel corpus studies should be conducted in the future to shed further light on change in motion event encoding in the Indo-European language family.

6.2 A possible explanation and further theoretical implications

It is clear that there exists a relationship between having manner-salient motion event encoding and having a productive system of adverbial path particles. While there is a potential good explanation for this — as long as the system of path encoding verbal prefixes is in place, the verb is “free” to encode manner — it is not entirely clear why certain Indo-European subgroups have lost this productive verbal prefix system while others have not. Balto-Slavic and Germanic have both retained an unmarked way to productively form the satellite-framed construction: Balto-Slavic through its case system and its system of verbal prefixes and prepositions, and Germanic through its system of verbal prefixes and prepositions. However, this does not explain why the Romance languages do not have a system of verbal prefixes or why Albanian did not simply continue to add new verbal prefixes as the original prefixes began to start to merge with verb roots (as the Germanic languages did, Goetz 2006). Why did Germanic and Balto-Slavic not lose the productivity of the verbal path prefix system, if the univerbation processes that affected Indo-Iranian, Romance, Albanian and Armenian were already attested in Proto-Indo-European?

One potential scenario that could explain this pattern might be contact. The partial Mantel test reported on in Section 5 did not yield any significant results, but this does not exclude the possibility that contact might have affected motion event encoding in the current sample. Wälchli (2009: 214) proposed that the distribution of manner-saliency, investigated by him through verb choice in a parallel corpus of the *Gospel according to Mark*, is mostly limited to North and Central Europe. Manner-saliency is less common than path-saliency in Wälchli’s sample, and this is also evident from the motion event encoding literature as a whole. The manner-salient area in North and Central Europe includes the Germanic languages, the Balto-Slavic languages, several Finno-Ugric languages (Estonian, Finnish, and Hungarian), several Daghestanian languages (including Lezgian) and Georgian. Most of these languages have a case system that is as extensive as that of the Balto-Slavic languages or far more extensive (except the majority of the Germanic languages). These languages all seem to be characterized by a motion encoding system in which path is typically not encoded on the verb but on verbal prefixes and cases. It might be the case that only the close proximity of these languages has allowed them to retain this system of motion event encoding, i.e. that without this areal effect all Indo-European languages would have been path-salient.

The existence of a manner-salient area in North and Central Europe could also potentially explain Irish and Serbo-Croatian motion event encoding. Irish has lost its verbal path prefixes, unlike the majority of the Germanic and Balto-Slavic languages, but is still manner-salient. Its geographical location close to English might

have prevented Irish from becoming more path-salient. To investigate this further we would need to compare Irish with the other Celtic languages.³ Serbo-Croatian is currently undergoing the univertation process: Serbo-Croatian's verbal prefixes have been merging with the deictic verb *ići* 'to go' to create a range of path verbs. Serbo-Croatian therefore seems to be currently in the process of becoming a path-salient language, using the satellite-framed construction less often and the path-only construction and verb-framed construction more often. The reason that Serbo-Croatian is going through this process might be due to its southern location. However, Serbo-Croatian is the only South Slavic language in my sample; a comparison with Macedonian, Bulgarian and Slovenian should be made in order to see whether the same process is currently affecting these languages as well. Further investigation into the similarities and histories of the motion event encoding systems of the languages that are part of this manner-salient area, as well as those that are outside of it, might uncover whether the contact hypothesis is valid or not.

We have seen that change in the motion event encoding system in Indo-European can be explained at least in part by the first of Croft et al. (2010)'s grammaticalization pathways (cf. 3a). Verbs and satellites have merged in all Indo-European subgroups. In addition, coordination is more common in path-salient and non-manner-salient languages as it is in manner-salient languages, suggesting that coordination is a proper mechanism to code manner and path information in one sentence if the satellite-framed construction is not preferred. However, it is also clear that Croft et al.'s (2010) grammaticalization pathways do not describe all the changes that might occur. The two grammaticalization pathways proposed both start with coordination, suggesting that all types of satellites should arise from verbs. This is not true, as path satellites do not necessarily arise from verbs (Lehmann 1985, Stevens 1992), although in some cases they do (Crapo 1970). A full description of all the different grammaticalization pathways that are possible would be needed to characterize change in motion event encoding completely.

7. Conclusion

The ancestral state estimation analysis conducted in this article indicates that Proto-Indo-European motion event encoding can be placed in the middle of the Talmian scale, with a slight tendency towards the manner-salient end of the scale. This result supports the picture of Proto-Indo-European motion event encoding that has emerged from studies of Classical Latin (Brucale 2011, Brucale, Iacobini & Mocciaro 2011) and Hittite (Brosch 2013), idiomatic meanings of preverb-verb combinations in ancient Indo-European languages, etymologies of Romance path verbs, and other historical-comparative evidence.

In the future, different phylogenetic methods that incorporate additional information about change in this domain may be used to further investigate the hypothesis that Proto-Indo-European is manner salient. For instance, an analysis could be used that takes into account information on the attested directionality of linguistic change in the preverb system. A promising method to gather information on the usage of motion event encoding constructions could be original text typology (Wälchli 2009, Haig, Schnell & Wegener 2011), in which linguistic features are investigated using their occurrence and distribution in native texts, instead of in grammars or parallel texts. Such a study could take into account data from contemporary and ancient corpora at the same time.

An analysis that incorporates this additional information may potentially still generate a classification of Proto-Indo-European that is more in line with the claims by Acedo Matellán & Mateu (2008) and Talmy (2007). However, the claim put forward in this article is that the productivity of the preverb system used for satellite-framed motion event encoding constructions was already declining in Proto-Indo-European, i.e. that Proto-Indo-European had an intermediary or mixed system of motion event coding. This seems valid both in light of the ancestral state estimation analysis conducted in this article and in light of several other types of evidence discussed in Sections 2 and 6. In this scenario, Germanic and Balto-Slavic might have been able to become more manner-salient over time due to contact with surrounding manner-salient languages.

Motion event encoding is a very specific semantic and syntactic subdomain. However, the building blocks that are used in the different syntactic constructions are very general: verbs, adverbs, adpositions, case markers, etc. Therefore, it seems clear that change in motion event encoding is dependent on grammatical change in various parts of the grammar. However, the way these different grammatical subsystems interact is unique for motion event encoding. The challenge in discovering the principles that govern change in this domain is to find out how large-scale grammatical changes, such as the merging of preverbs and verb roots or the loss or emergence of case systems, interact to give rise to the diversity of motion event encoding systems. It will be a continuing challenge to identify, describe, and analyze these principles as they are encountered in different languages across the globe.

Notes

1. The maximum possible λ values, given the phylogenetic trees in the current sample, ranged from 1.09 to 1.28.
2. The estimated λ should be tested to be significantly different from a model in which λ is set to have the maximum possible value of λ given the phylogenetic tree, not simply to be significantly

different from a model in which $\lambda=1$. However, the true maximum values could not be used as maximum possible λ values, as they are not accepted by the `corPagel` function from the R package `ape` (Paradis, Claude & Strimmer 2004). This function only accepts fixed values for λ between 0 and 1.

3. The parallel corpus also includes a Breton translation of *Alice's Adventures in Wonderland* that was not included in the current dataset. A comparison of the Breton translation with the other twenty translations indicates that Breton behaves very similar to Irish. This is unexpected, as one would expect Breton to be influenced by French to a much larger extent than is suggested by this comparison. Further investigations may be able to shed more light on potential contact influences on Celtic motion event encoding.

References

- Acedo Matellán, Víctor & Jaume Mateu. 2008. The Path from Satellite-Framed Indo-European to Verb-Framed Romance: A Lexical-Syntactic Account. *Paper presented at the X Diachronic Generative Syntax Conference*. Cornell University, Ithaca, New York.
- Acedo Matellán, Víctor & Jaume Mateu. 2010. From Satellite-Framed Latin to Verb-Framed Romance: A Syntactic Approach. *GGT* 10–01.
- Aske, Jon. 1989. Path Predicates in English and Spanish: A Closer Look. *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* ed. by Kira Hall, Michael Meacham & Richard Shapiro, 1–14. Berkeley: Berkeley Linguistics Society, Inc.
- Baayen, R. Harald. 1994. Derivational Productivity and Text Typology. *Journal of Quantitative Linguistics* 1.16–34. DOI: 10.1080/09296179408589996
- Baicchi, Annalisa. 2005. Translating Phrasal Combinations across the Typological Divide. *Studies in the Semantics of Lexical Combinatory Patterns* ed. by M. Bertuccelli Papi, 487–519. Pisa: Pisa University Press.
- Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The Typology of Motion Expressions Revisited. *Journal of Linguistics* 46.331–377. DOI: 10.1017/S0022226709990272
- Berman, Ruth A. & Dan I. Slobin, eds. 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Berthele, Raphael. 2006. *Ort und Weg. Die sprachliche Raumreferenz in Varietäten des Deutschen, Rätoromanischen und Französischen*. Berlin: De Gruyter. DOI: 10.1515/9783110890464
- Bielec, Dana. 1998. *Polish: An Essential Grammar*. London: Routledge.
- Bloch, Jules. 1965. *Indo-Aryan: From the Vedas to Modern Times*. Paris: Librairie Adrien Maisonneuve.
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russel D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337.957–960. DOI: 10.1126/science.1219669
- Brachet, Auguste. 1882. *An Etymological Dictionary of the French Language* [trans. G. W. Kitchin]. Oxford: Clarendon Press.
- Brosch, Cyril. 2013. *Untersuchungen zur hethitischen Raumgrammatik*. Berlin: de Gruyter.
- Brucale, Luisa. 2011. Manner of Motion Verbs in Latin. Paper at the *Historical-Comparative Linguistics in the 21st Century*. Pavia, Italy, September 2011.

- Brucale, Luisa, Claudio Iacobini & Egle Mocciaro. 2011. Typological Change in the Expression of Motion Events from Latin to Romance Languages. Paper at the *44th Annual Meeting of the Societas Linguistica Europaea*, Universidad de la Rioja, September 2011.
- Calude, Andreea S. & Mark Pagel. 2011. How Do We Use Language? Shared Patterns in the Frequency of Word Use across 17 World Languages. *Philosophical Transactions of the Royal Society B* 366.1101–1107. DOI: 10.1098/rstb.2010.0315
- Camaj, Martin. 1984. *Albanian Grammar: With Exercises, Chrestomathy and Glossaries*. [trans. L. Fox] Wiesbaden: Harrassowitz.
- Crapo, Richley. 1970. The Origins of Directional Adverbs in Uto-Aztecan Languages. *International Journal of American Linguistics* 36.181–189. DOI: 10.1086/465109
- Croft, William, Jóhanna Barðdal, Willem Hollman, Violeta Sotirova & Chiaki Taoka. 2010. Revising Talmy's Typological Classification of Complex Event Constructions. *Contrastive Studies in Construction Grammar* ed. by Hans C. Boas, 201–235. Amsterdam: John Benjamins.
- Danesi, Serena. 2013. Particle-Verb Constructions in Vedic: The Case of *ápa*. *Studi e Saggi Linguistici* 51.2: 57–100.
- Delbrück, Berthold. 1888. *Altindische Syntax*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Delbrück, Berthold. 1893. *Vergleichende Syntax der Indogermanischen Sprachen*. Strassburg: K. J. Trübner.
- Drummond, Alexei J., Marc A. Suchard, Dong Xie & Andrew Rambaut. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29.8: 1969–73.
- Dufresne, Monique, Fernande Dupuis & Mireille Tremblay. 2003. Preverbs and Particles in Old French. *Yearbook of Morphology* ed. by Geert Booij & Jaap van Marle, 30–60. Dordrecht: Kluwer.
- Dunn, Michael, 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. DOI: 10.1038/nature09923
- Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indo-European Classification: A Lexicostatistical Experiment. *Transactions of the American Philological Society* 82.1–132.
- Ferrari, Giacomo & Monica Mosca. 2010. Some Constructions of Path: From Italian to Some Classical Languages. *Space in Language: Proceedings of the Pisa International Conference* ed. by Giovanna Marotta, Alessandro Lenci, Linda Meini & Francesco Rovai, 317–338. Pisa: Edizioni ETS.
- Filipović, Luna. 2007. *Talking about Motion: A Crosslinguistic Investigation of Lexicalization Patterns*. Amsterdam: John Benjamins. DOI: 10.1075/slcs.91
- Folli, Raffaella & Gillian Ramchand. 2001. Getting Results: Motion Constructions in Italian and Scottish Gaelic. *Proceedings of the 20th West Coast Conference on Formal Linguistics* ed. by Karine Megerdooomian & Leora Anne Bar-el, 101–114. Somerville, MA: Cascadilla Press.
- Folli, Raffaella & Gillian Ramchand. 2005. Prepositions and Results in Italian and English: An Analysis from Event Decomposition. *Perspectives on Aspect* ed. by Henk J. Verkuyl, Henriette de Swart & Angeliek van Hout, 1–20. Dordrecht: Kluwer.
- Freckleton, Rob P., Paul H. Harvey & Mark Pagel. 2002. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *The American Naturalist* 160.712–726. DOI: 10.1086/343873
- Goetz, Virginia A. 2006. *The Development of Proto-Indo-European Local Adverbs into Germanic Prepositions and Verbal Elements*. University of Connecticut PhD dissertation.

- Grieve, Jack. 2014. A Comparison of Statistical Methods for the Aggregation of Regional Linguistic Variation. *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech* ed. by Benedikt Szmrecsanyi & Bernhard Wälchli. Berlin: Walter de Gruyter.
- Haig, Geoffrey J. L., Stefan Schnell & Claudia Wegener. 2011. Comparing Corpora from Endangered Language Projects: Explorations in Language Typology based on Original Texts. *Documenting Endangered Languages: Achievements and Perspectives* ed. by Geoffrey L. J. Haig, Nicole Nau, Stefan Schnell & Claudia Wegener, 55–86. Berlin: Walter de Gruyter. DOI: 10.1515/9783110260021
- Harmon, Luke J., Jason T. Weir, Chad D. Brock, Richard E. Glor & Wendell Challenger. 2008. GEIGER: Investigating Evolutionary Radiations. *Bioinformatics* 24.129–131. DOI: 10.1093/bioinformatics/btm538
- Haug, Dag T. T. & Marius Jøhndal. 2008. Creating a Parallel Treebank of the old Indo-European Bible Translations. *Proceedings of the Sixth International Language Resources and Evaluation (LREC '08)* ed. by Khalid Choukri, 27–34. European Language Resources Association (ELRA).
- Hewson, John & Vit Bubenik. 2006. *From Case to Adposition: The Development of Configurational Syntax in Indo-European Languages*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.280
- Hickmann, Maya, Henriëtte Hendriks, Efstathia Soroli, Tatiana Iakovleva & Yinglin Ji. Under review. Space and Language Typology: Encoding Motion across Languages.
- Hofmann, Jofmann B. & Anton Szantyr. 1965. *Lateinische Syntax und Stilistik*. München: C. H. Beck'sche Verlagsbuchhandlung.
- Holton, David, Peter Mackridge & Irene Philippaki-Warbuton. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian Inference of Phylogeny and its Impact on Evolutionary Biology. *Science* 294.2310–2314. DOI: 10.1126/science.1065889
- Iacobini, Claudio & Francesca Masini. 2006. The Emergence of Verb-Particle Constructions in Italian: Locative and Actional Meanings. *Morphology* 16.155–188. DOI: 10.1007/s11525-006-9101-7
- Iacobini, Claudio & Francesca Masini. 2007. Verb-Particle Constructions and Prefixed Verbs in Italian: Typology, Diachrony and Semantics. *Proceedings of the Fifth Mediterranean Morphology Meeting* ed. by Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli & Sergio Scalise, 157–184. Bologna: Università degli Studi di Bologna.
- Imbert, Caroline. 2010. Multiple Preverbatation in Homeric Greek: A Typological Insight. *CogniTextes* 4. Available at <http://cognitextes.revues.org/387>.
- Junghänel, Anja. In preparation. Coding Motion Events in Indo-European.
- Kent, Roland G. 1950. *Old Persian*. New Haven: American Oriental Society.
- Kopecka, Anetta. 2006. The Semantic Structure of Motion Verbs in French: Typological Perspectives. *Space in Languages: Linguistic Systems and Cognitive Categories* ed. by Maya Hickmann & Stéphane Robert, 83–102. Amsterdam: John Benjamins.
- Kopecka, Anetta. 2009. Continuity and Change in the Representation of Motion Events in French. *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin* ed. by Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan, 415–426. New York: Psychology Press.
- Kurylowicz, Jerzy. 1964. *The Inflectional Categories of Indo-European*. Heidelberg: Carl Winter Universitätsverlag.

- Lambton, Ann K. S. 1953. *Persian Grammar*. Cambridge: Cambridge University Press.
- Lehmann, Christian. 1985. Grammaticalization: Synchronic Variation and Diachronic Change. *Lingua e Stile* 20.303–318.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Leumann, Manu. 1977. *Lateinische Laut- und Formenlehre*. München: C. H. Beck'sche Verlagsbuchhandlung.
- Luraghi, Silvia. 2010. The Rise (and Possible Downfall) of Configurationality. *A Companion to Historical Linguistics* ed. by Silvia Luraghi & Vit Bubenik, 212–229. London: Continuum Books.
- Mahootian, Shahrzad. 1997. *Persian*. London: Routledge. DOI: 10.4324/9780203192887
- Masini, Francesca. 2005. Multi-Word Expressions between Syntax and the Lexicon: The Case of Italian Verb-Particle Constructions. *SKY Journal of Linguistics* 18.145–173.
- Miklosich, Franz. 1868. *Vergleichende Grammatik der slavischen Sprachen*. Wien: W. Braumüller.
- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81.382–420. DOI: 10.1353/lan.2005.0078
- Narasimhan, Bhuvana. 1998. *The Encoding of Complex Events in Hindi and English*. Boston University PhD dissertation.
- Oksanen, Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, Henry H. Stevens & Helene Wagner. 2012. *vegan: Community Ecology Package*. R Package Version 2.0–4.
- Orel, Vladimir E. 2000. *A Concise Historical Grammar of the Albanian Language: Reconstruction of Proto-Albanian*. Leiden: Brill.
- Özçalışkan, Şeyda & Dan I. Slobin. 2003. Codability Effects on the Expression of Manner of Motion in English and Turkish. *Studies in Turkish Linguistics: Proceedings of the Tenth International Conference on Turkish Linguistics* ed. by Sumru Özsoy, Didar Akar, Mine Nakipoğlu-Demiralp, Eser Erguvanlı-Taylan & Ayhan Aksu-Koç, 259–270. Istanbul: Bogazici University Press.
- Pagel, Mark. 1999. Inferring the Historical Patterns of Biological Evolution. *Nature* 401.877–884. DOI: 10.1038/44766
- Papafragou, Anna, Christine Massey & Lila Gleitman. 2002. Shake, Rattle, 'n' Roll: The Representation of Motion in Language and Cognition. *Cognition* 84.189–219. DOI: 10.1016/S0010-0277(02)00046-X
- Papafragou, Anna, Christine Massey & Lila Gleitman. 2006. When English Proposes What Greek Presupposes: The Cross-Linguistic Encoding of Motion Events. *Cognition* 98.B75–B87. DOI: 10.1016/j.cognition.2005.05.005
- Paradis, Emmanuel, Julien Claude & Korbinian Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* 20.289–290. DOI: 10.1093/bioinformatics/btg412
- Penney, J. H. W. 1989. Preverbs and Postpositions in Tocharian. *Transactions of the Philological Society* 87.54–74. DOI: 10.1111/j.1467-968X.1989.tb00619.x
- Peyraube, Alain. 2006. Motion Events in Chinese: A Diachronic Study of Directional Complements. *Space across Languages: Linguistic Systems and Cognitive Categories* ed. by Maya Hickmann & Stéphane Robert, 121–135. Amsterdam: John Benjamins.
- Pokorny, Julius. 1925. *Altirische Grammatik*. Berlin: Walter de Gruyter.
- Pokorny, Julius. 1959–1969. *Indogermanisches Etymologisches Wörterbuch*. Bern: Francke.

- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at <http://www.r-project.org/>.
- Revell, Liam J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63.3258–3268. DOI: 10.1111/j.1558-5646.2009.00804.x
- Revell, Liam J. 2012. phytools: An R Package for Phylogenetic Comparative Biology (and Other Things). *Methods in Ecology and Evolution* 3.217–223. DOI: 10.1111/j.2041-210X.2011.00169.x
- Ringe, Don. 2006. *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199284139.001.0001
- Roberts, Murat H. 1936. The Antiquity of the Germanic Verb-Adverb Locution. *The Journal of English and Germanic Philology* 35.466–481.
- Schluter, Dolph, Trevor Price, Arne Ø. Mooers & Donald Ludwig. 1997. Likelihood of Ancestor States in Adaptive Radiation. *Evolution* 51.1699–1711. DOI: 10.2307/2410994
- Schmitt, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Schwyzler, Eduard. 1950. *Griechische Grammatik vol. 2: Syntax und syntaktische Stilistik*. München: C. H. Beck'sche Verlagsbuchhandlung.
- Slobin, Dan I. 1996. Two Ways to Travel: Verbs of Motion in English and Spanish. *Grammatical Constructions: Their Form and Meaning* ed. by Masayoshi Shibitani & Sandra A. Thompson, 195–217. Oxford: Clarendon Press.
- Slobin, Dan I. 2004. The Many Ways to Search for a Frog: Linguistic Typology and the Expression of Motion Events. *Relating Events in Narrative: Typological and Contextual Perspectives* ed. by Sven Strömquist & Ludo Verhoven, 219–257. Mahwah, NJ: Lawrence Erlbaum Associates.
- Slobin, Dan I. 2005. Relating Narrative Events in Translation. *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman* ed. by Dorit D. Ravid & Hava Bat-Zeev Shyldkrot, 115–129. Dordrecht: Kluwer. DOI: 10.1007/1-4020-7911-7_10
- Slobin, Dan I. & Nini Hoiting. 1994. Reference to Movement in Spoken and Signed Languages: Typological Considerations. *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, 487–505. Berkeley: Berkeley Linguistics Society.
- Skopeteas, Stavros. 2002. *Lokale Konstruktionen im Griechischen: Sprachwandel in funktionaler Sicht*. Erfurt: Universität Erfurt.
- Speyer, Jacob S. 1896. *Vedische und Sanskrit-Syntax*. Strassburg: K. J. Trübner. DOI: 10.1515/9783111581842
- Stevens, Christopher M. 1992. Grammaticalization in Spatial Deixis: A Case Study. *On Germanic Linguistics: Issues and Methods* ed. by Irmengard Rauch, Gerald F. Carr & Robert L. Kyes, 299–313. Berlin: Mouton de Gruyter.
- Talmy, Leonard. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon* Vol. 3 ed. by Timothy Shopen, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. Path to Realization: A Typology of Event Conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 480–519. Berkeley: Berkeley Linguistics Society.
- Talmy, Leonard. 2007. Lexical Typologies. *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon* Vol. 3 ed. by Timothy Shopen, 66–168. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511618437.002
- Townsend, Charles E. 1968. *Russian Word-Formation*. New York: McGraw-Hill Book Company.

- Verkerk, Annemarie. 2014. Where Alice Fell into: Motion Events from a Parallel Corpus. *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech* ed. by Benedikt Szmrecsanyi & Bernhard Wälchli, 324–354. Berlin: Walter de Gruyter.
- Vincent, Nigel. 1999. The Evolution of C-Structure: Prepositions and PPs from Indo-European to Romance. *Linguistics* 37.1111–1153. DOI: 10.1515/ling.37.6.1111
- Wälchli, Bernhard. 2009. *Motion Events in Parallel Texts: A Study in Primary-Data Typology*. Habilitationsschrift, Universität Bern.
- Watkins, Calvert. 1964. Preliminaries to the Reconstruction of Indo-European Sentence Structure. *Proceedings of the IX International Congress of Linguists* ed. by H. G. Lunt, 1035–1045. The Hague: Mouton.
- Whitney, William D. 1879. *A Sanskrit Grammar*. Leipzig: Breitkopf and Härtel.
- Windfuhr, Gernot L., ed. 2009. *The Iranian Languages*. London: Routledge.

Appendix 1. An overview of manner verbs in two different corpora

This appendix gives an overview of seven different English manner verbs and their Ancient Greek correspondents in two text corpora (first and fourth column). The English numbers are taken from *Alice's Adventures in Wonderland* and *Through the Looking-Glass and What Alice Found There* by Lewis Carroll, consisting of 56,777 words. The Ancient Greek numbers are taken from the *New Testament* (consisting of 179,011 words) available at PROIEL (Haug & Jøhndal 2008). It shows that the *New Testament*, contains approximately the same number of manner verbs (second and fifth column). However, since the *New Testament* contains approximately three times more words as the *Alice* corpus, it contains fewer manner verbs per 10,000 words (third and sixth column).

English	No.	10,000	Ancient Greek	No.	10,000
<i>run</i>	45	7.9	τρέχω 'to run'	19	1.1
<i>walk</i>	35	6.2	(περι)πατέω 'to go about, walk'	85	4.7
<i>jump</i>	20	3.5	πηδαν; ἄλλομαι 'to jump, leap'	3	0.2
<i>swim</i>	7	1.2	νειν; κολυμβάω 'to swim (up and down)'	1	0.06
<i>fly</i>	5	0.9	πέτομαι 'to fly'	3	0.2
<i>crawl</i>	4	0.7	ερπειν 'to crawl'	0	0
<i>rush</i>	2	0.4	ὀρμάω 'to rush'	5	0.3
Total	118	20.8	Total	116	6.56

In addition, many of these manner verbs do not occur in descriptions of motion events, as they do not include a path of motion. A few examples that illustrate the use of manner verbs in the *New Testament* taken from the King James translation are provided below.

- (1) Matthew 11:5: *The blind receive their sight, and the lame walk, the lepers are cleansed, and the deaf hear, the dead are raised up, and the poor have the gospel preached to them.*

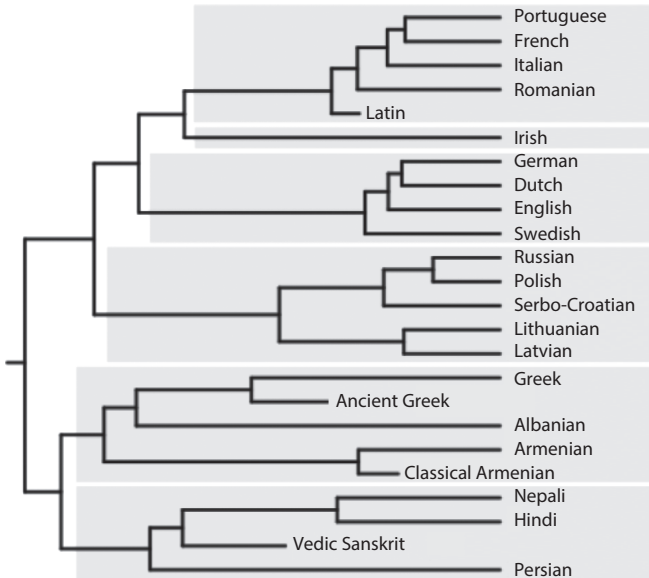
- (2) Galatians 2.2: *And I went up by revelation, and communicated unto them that gospel which I preach among the Gentiles, but privately to them which were of reputation, lest by any means I should run, or had run, in vain.*
- (3) Revelations 4.7: *And the first beast was like a lion, and the second beast like a calf, and the third beast had a face as a man, and the fourth beast was like a flying eagle.*

Appendix 2. Principal component scores and proportions of construction usage for twenty Indo-European languages.

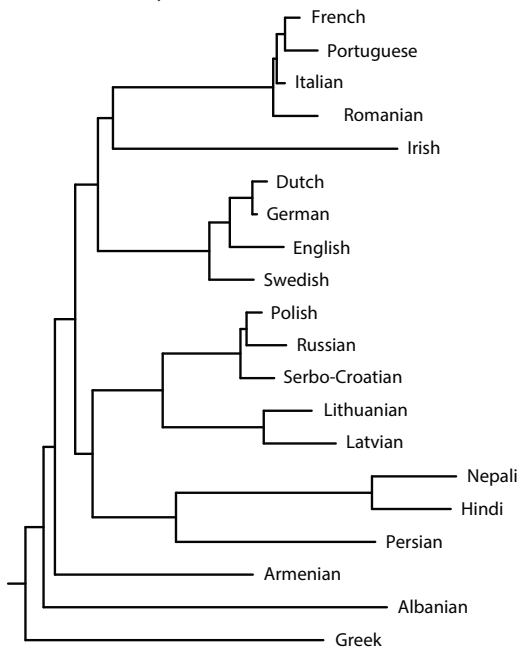
Language	Score on 1st PC	Score on 2nd PC	Satellite-framed construction	Path-only construction	Deictic verb construction	Verb-framed construction	Deictic verb-framed construction	Manner only construction	Manner+path verb construction	Subordinate construction	Coordinate construction
Russian	0.25	0.10	0.49	0.26	0	0.03	0	0.04	0.01	0.01	0.02
Polish	0.25	0.12	0.51	0.29	0	0.03	0	0.06	0.02	0	0.02
Serbo-Croatian	-0.01	0.07	0.30	0.36	0.03	0.09	0.01	0.02	0.01	0.01	0.07
Latvian	0.21	0.06	0.45	0.26	0.03	0.03	0.01	0.04	0.01	0.03	0.02
Lithuanian	0.25	0.12	0.51	0.27	0	0.04	0	0.07	0.02	0.03	0.01
German	0.21	0.03	0.45	0.26	0.06	0.03	0.04	0.01	0.02	0.01	0.01
English	0.13	-0.03	0.37	0.28	0.09	0.03	0.11	0.03	0	0	0.04
Dutch	0.23	-0.06	0.44	0.19	0.11	0.03	0.11	0.03	0	0	0.02
Swedish	0.27	-0.05	0.47	0.17	0.13	0.03	0.04	0.03	0.01	0.01	0.03
French	-0.09	0.07	0.20	0.39	0.05	0.17	0	0.04	0.02	0.04	0.02
Portuguese	-0.10	0.10	0.21	0.42	0.04	0.16	0.01	0.08	0	0.02	0.03
Italian	-0.03	0.09	0.26	0.39	0.03	0.11	0.03	0.05	0.01	0.01	0.02
Romanian	-0.09	0.10	0.21	0.44	0.02	0.09	0.02	0.06	0.02	0.01	0.06
Irish	-0.02	-0.09	0.25	0.28	0.17	0.07	0.08	0.03	0	0	0.03
Greek	-0.04	0.09	0.26	0.42	0.04	0.09	0.03	0.02	0.01	0.03	0.03
Armenian	-0.05	-0.04	0.19	0.31	0.08	0.05	0.03	0.06	0.01	0.01	0.12
Albanian	-0.15	0.07	0.14	0.43	0.04	0.13	0.02	0.05	0.01	0.04	0.05
Persian	-0.08	-0.07	0.15	0.31	0.13	0.09	0.08	0.03	0.01	0	0.03
Nepali	-0.11	-0.11	0.10	0.29	0.11	0.07	0.05	0.03	0.02	0.03	0.10
Hindi	-0.06	-0.10	0.17	0.29	0.10	0.06	0.09	0.03	0	0.05	0.11
Mean	0.05	0.02	0.31	0.32	0.06	0.07	0.04	0.04	0.01	0.02	0.04

Appendix 3. Maximum clade credibility trees used for robustness tests

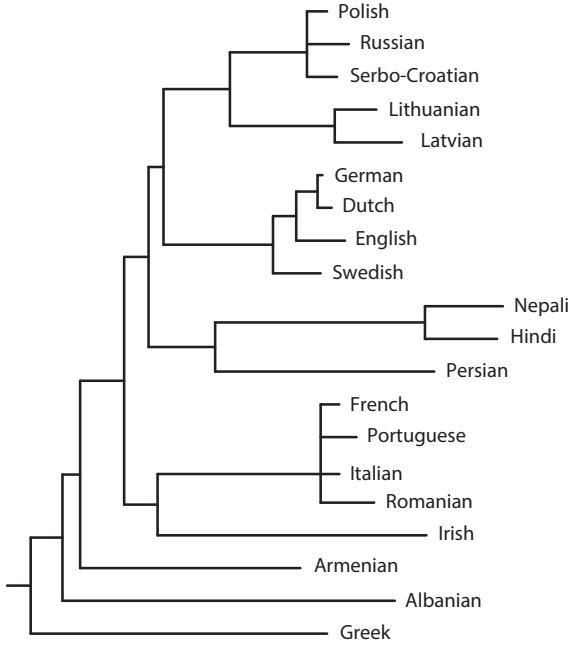
- (1) Maximum clade credibility tree of Bouckaert et al. (2012)'s tree sample including four ancient Indo-European languages:



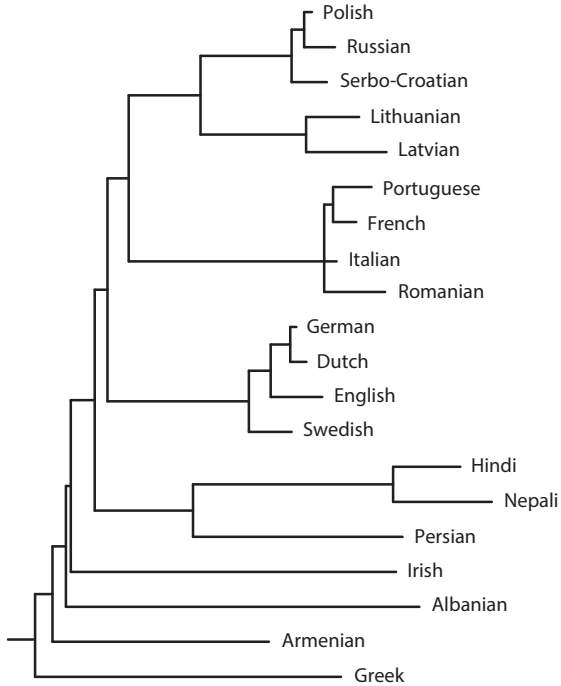
- (2) Maximum clade credibility tree of Dunn et al. (2011)'s full tree set:



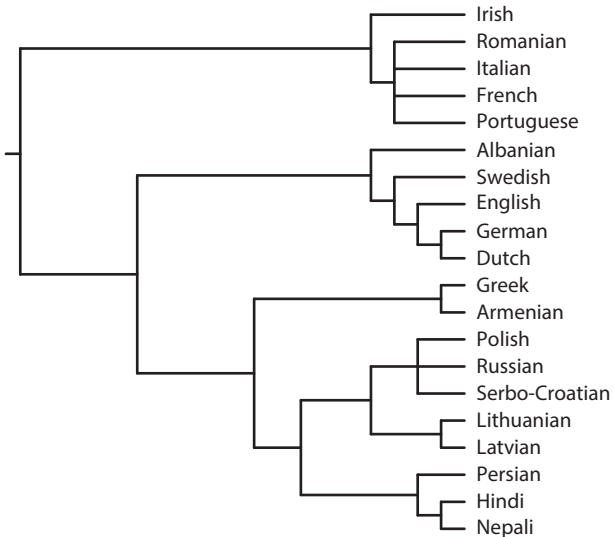
- (3) Maximum clade credibility tree of Dunn et al. (2011)'s trees in which Balto-Slavic and Germanic form a subgroup:



- (4) Maximum clade credibility tree of Dunn et al. (2011)'s trees in which Balto-Slavic and Romance form a subgroup:



- (5) Nakhleh, Ringe & Warnow (2005)'s tree A:



Author's address

Max Planck Institute for Psycholinguistics
Wundtlaan 1
6525XD Nijmegen
The Netherlands
a.verkerk@reading.ac.uk