

Where do all the motion verbs come from?

The speed of development of manner verbs and path verbs in Indo-European*

Annemarie Verkerk

Max Planck Institute for Psycholinguistics, Reading Evolutionary
Biology Group

The last four decades have seen huge progress in the description and analysis of cross-linguistic diversity in the encoding of motion (Talmy 1985, 1991, Slobin 1996, 2004). Comparisons between satellite-framed and verb-framed languages suggest that satellite-framed languages typically have a larger manner of motion verb lexicon (*swim, dash*), while verb-framed languages typically have a larger path of motion verb lexicon (*enter, cross*) (Slobin 2004, Verkerk 2013, 2014b). This paper investigates how differences between the motion verb lexicons of satellite-framed and verb-framed languages emerge. Phylogenetic comparative methods adopted from biology and an etymological study are used to investigate manner verb lexicons and path verb lexicons in an Indo-European dataset. I show that manner verbs and path verbs typically have different types of etymological origins and that manner verbs emerge faster in satellite-framed subgroups, while path verbs emerge faster in verb-framed subgroups.

Keywords: motion events, manner verbs, path verbs, Indo-European, etymology, phylogenetic comparative methods, rate of evolutionary change

* This work was supported by the Max Planck Gesellschaft zur Förderung der Wissenschaften through their funding of the Max Planck Research Group “Evolutionary Processes in Language and Culture” headed by Michael Dunn and by the ERC grant ‘MotherTongue’ awarded to Mark Pagel. I am grateful to Simon Branford, Olga Krasnoukhova and three anonymous reviewers for their comments on earlier versions of the manuscript. I am also grateful to Per Ambrosiani, Johnny Cheung, Hrach Martirosyan and Ranko Matasović for answering questions about Russian, Persian, Armenian and Irish etymologies.

1. Introduction

Motion event encoding has been a popular area of investigation in cognitive linguistics and linguistic typology over the last three decades. A major catalyst was the work of Leonard Talmy (1985, 1991), who formulated the opposition between verb-framed and satellite-framed languages: verb-framed languages typically use a construction in which information on the path of motion is encoded on the verb, while satellite-framed languages typically use a construction in which information on path is encoded outside the verb on a so-called satellite (1, 2):

(1) Italian (verb-framed)

Alice si inoltr-ò corre-ndo nel bosco e si ferm-ò sotto un albero frondoso
 Alice REFL penetrate-PST.3SG run-PRS.PTCP in.the wood and REFL stop-PST.3SG under a tree leafy
 “Alice entered the wood running and stopped under a leafy tree”

(2) Swedish (satellite-framed)

Alice sprang in i skog-en och ställde sig under ett stort träd
 Alice run.PST into in forest-the and stand.PST REFL under a large tree
 “Alice ran into the forest and placed herself under a large tree”

In the Italian example in (1), the path of motion (from outside the forest to inside it) is encoded on the verb *inoltrarsi* “penetrate”, while the manner of motion (running) is encoded on the participle *correndo* “running”. In the Swedish example in (2), the manner of motion is encoded on the verb *springa* “run”, while the path of motion is encoded on the adverb *in* “into” and the preposition *i* “in”.

These and other syntactic constructions used to encode motion have been described for many languages around the world. Of interest here is the relationship between these syntactic constructions and the motion verb lexicon. Recent studies have proposed a relationship between the commonly used motion event encoding construction and the size of the lexicon of certain types of motion verbs. Languages that commonly make use of the satellite-framed construction, such as Swedish, are claimed to have a larger lexicon of manner of motion verbs such as *walk*, *fly* and *limp* (Slobin 2004). On the other hand, languages that commonly make use of the verb-framed construction, such as Italian, are claimed to have a larger lexicon of path of motion verbs such as *ascend*, *exit* and *cross* (Cifuentes Férrez 2010). Although the opposition between satellite-framed languages and verb-framed languages is not absolute — satellite-framed languages typically also have path verbs and can employ them in a verb-framed construction and vice versa — dependencies between the motion encoding system and the lexicon have emerged. Recent research (Verkerk 2013, 2014b) has provided some evidence that

Slobin's (2004) and Cifuentes F erez (2010) claims are true for a sample of 20 Indo-European (IE) languages.

If such correlations truly exist this means that the history of the motion verb lexicons of the IE languages should reflect these differences. If we use a phylogenetic tree to model the relationships of the IE languages then we would expect manner verbs to emerge faster in the branches of the tree leading to satellite-framed languages, to generate the larger manner verb classes that are found in these languages. Likewise, path verbs should emerge faster in branches that lead to verb-framed languages. Note that modeling verb evolution in this manner implies a whole range of different processes by means of which verbs can enter the lexicon: inheritance from Proto-Indo-European (PIE), semantic conversion of non-motion verbs, borrowing from related and unrelated languages, changes in morphology, etc. If we know the etymology of a given motion verb, we can pinpoint when and how it entered the language. This paper investigates where IE manner verbs and path verbs come from and whether there is evidence for faster or slower evolution of these verb classes in various branches of the IE tree.

IE languages provide an interesting case study because they are quite diverse with respect to the typical motion event encoding constructions (Slobin 2004, 2005, Verkerk 2014a, 2014c). The Germanic and Balto-Slavic languages are mainly satellite-framed, while the Romance languages are mainly verb-framed. Albanian and Modern Greek are also verb-framed. The Indo-Iranian and Celtic languages, as well as Armenian, can be considered to be verb-framed with respect to their frequency of use of the satellite-framed construction, but frequently use deictic verbs instead of path verbs. This diversity allows us to investigate the following questions:

1. Assuming that it is true that the frequent use of the satellite-framed construction correlates with a larger manner verb lexicon, can we find evidence of faster manner verb evolution in the branches leading to the ancestors of satellite-framed languages (Proto-Germanic, Proto-Balto-Slavic) and to the satellite-framed languages themselves (English, Russian, etc.)?
2. Assuming that it is true that the frequent use of the verb-framed construction correlates with a larger path verb lexicon, can we find evidence of faster path verb evolution in the branches leading to the ancestors of verb-framed languages (Proto-Romance, Proto-Indo-Iranian) and to the verb-framed languages themselves (French, Hindi, etc.)?

In the current contribution, the hypothesis is that the answer to these two questions is 'yes'. This hypothesis is investigated by studying the etymological origins of the manner verbs and path verbs in a sample of 21 different Indo-European languages. Phylogenetic comparative methods are used to estimate the rate of change

or rate of evolution (Venditti et al. 2011). The rate of change, which refers to the speed with which a feature such as lexicon size is evolving, is an important variable if we want to model the evolution of linguistic features in a phylogenetic framework. During the history of the IE languages, manner verb lexicons and path verb lexicons might have been expanding or shrinking at different or slower rates given their sizes in different languages. This can be due to socio-historical factors, such as contact with languages with larger manner or path verb lexicons, or linguistic factors, such as changes in morphosyntax. By studying differential rates of change in different parts of the tree, we might find further support of the hypothesis that the growth and decline of the manner verb and the path verb class is correlated with syntactic patterns of motion event encoding.

The current study provides a contribution to lexical typology by investigating diachronic change in two semantic subfields. Aside from many studies on isolated processes of lexical change such as borrowing, compounding, affixation and semantic shift, remarkably few studies of lexical change take into account all word formation processes in a single language, or take into account word formation in a specific semantic domain. Studies focusing on general word formation are Algeo (1980), who studied the origins of a randomly chosen 1000 English words, and Cannon (1978), who studied the origins of Merriam-Webster's 6000 Words, a collection of new words in American English published in 1976. Studies looking at word formation in specific semantic domains include Witkowski et al. (1981), who study words for "tree", and Brown (1983), who studies the etymological origins of cardinal direction terms. Closely related to the topic of this paper, Fanego (2012) studies the etymological origins of manner of motion verbs in English. Like these studies, this paper continues a focus on general etymological or word formation patterns in a given language or semantic domain. Individual etymologies and their implications for theories of semantic change do not play a big role. The emphasis is on cross-linguistic correspondences and differences in the etymological histories of motion verbs and what this can tell us about rates of change of the motion verb lexicon.

2. Manner verbs and path verbs in 21 Indo-European languages

2.1 The parallel corpus

The current dataset is taken from a parallel corpus of descriptions of motion events in 21 IE languages: French, Italian, Portuguese, Romanian (Romance), Dutch, English, German, Swedish (Germanic), Irish (Celtic), Latvian, Lithuanian, Polish, Russian, Serbo-Croatian (Balto-Slavic), Assamese, Hindi, Nepali, Persian

(Indo-Iranian), Albanian, Modern Greek and Armenian. The parallel corpus consists of translations of two novels, Lewis Carroll's *Alice's Adventures in Wonderland* and Paulo Coelho's *O alquimista* ["The Alchemist"], into these 21 languages.

First, all descriptions of motion events were extracted from these two novels. Each of these motion event descriptions was a single sentence, which encoded a single situation in which an entity moves from one place to another place (Özçalışkan & Slobin 2003:259, Berman & Slobin 1994:657). Second, a selection of motion event descriptions was made, as including all descriptions would create a dataset of unmanageable size. By choosing sentences with the largest range of construction types as well as making sure that at least one instance of every attested motion verb was included, a selection of 132 sentences that encode voluntary (non-causative) motion was made. The translations of these motion event descriptions were found, glossed and analyzed with the help of native speakers. The coding of motion verbs into relevant categories was done with the assistance of native speakers as well. In total, the dataset consisted of 2772 motion encoding sentences that feature at least one motion verb. Although the sample of 21 IE languages is limited due to the availability of translations of *Alice's Adventures in Wonderland* and *O alquimista*, my impression is that the sample spread across the different IE subgroups and the variety of motion verbs included allows for a comprehensive overview of the most frequent manner verbs and path verbs in IE.

2.2 Coding of manner verbs and path verbs

The path verbs and manner verbs were coded as follows. If a manner verb encoded a transparent manner of motion and could be used with different paths, such as English *run*, *roll* or *crawl*, it was coded as a manner verb. If a path verb encoded a transparent path of motion and could be used in different manner of motion contexts, such as English *descend*, *enter* and *pass*, it was coded as a path verb. In this way, verbs that encode both path and manner at the same time, such as Greek *skarfalono* "to climb up" and Persian *goriqtan* "to run away", are kept in a separate category, that of manner plus path verbs (see Verkerk 2014c). Phrasal manner verbs, such as *give a leap* and *make a rush*, were not included as manner verbs in this study (see Verkerk 2013). For path verbs consisting of a lexicalized combination of a path prefix and a verb, strict criteria regarding the semantic transparency of the verb complex were maintained (see Verkerk 2014b).

As originally proposed by Slobin (1997), the manner verb lexicon can be divided into two parts: a first tier of more general or neutral verbs and a second tier of more specific and expressive verbs. Satellite-framed languages have extensive second tier manner verb lexicons, while the first tier manner verbs should be more or less the same set in both non-satellite-framed and satellite-framed languages.

For the purposes of this paper, first tier manner verbs are those that denote the manner of motion concepts WALK, STROLL, RUN, FLY, SWIM, ROLL, JUMP, RUSH/HURRY/DASH and WANDER. Second tier manner verbs are those that are more specific and expressive, including among others the concepts MARCH, SNEAK, RIDE, CRAWL and TROT. These classifications are based on Slobin's (1997) discussion and earlier investigations of this dataset (Verkerk 2013).

In addition, a third tier is distinguished that classifies non-motion verbs used as manner verbs. These are verbs that denote manner of motion, such as *wriggle*, *zoom*, *shave* and *struggle*, but that do not denote translational motion in the same sense as manner verbs do. A manner of motion verb such as *run* in *Sally ran* signifies that Sally changed her position through running, whereas a non-motion verb such as *wriggle* in *Sally wriggled* does not. These verbs can however occur in various types of constructions, including the satellite-framed construction (*Jane wriggled out of her sleeping-bag*) and the subordinate construction (*Alice struggled to get up the stairs*) (see Verkerk 2014c on the subordinate construction). Since these third tier manner verbs do not denote translational motion, the manner of motion reading is not an inherent part of the semantics of these verbs. Rather, the manner of motion reading of these verbs emerges from their use in motion constructions, where they are combined with directional phrases. Although they technically cannot be called manner of motion verbs in light of the definition at the beginning of this section, they are included here as manner verbs because they are the result of a mechanism to create manner verbs within specific syntactic contexts. This process could be determined by the type of motion event encoding construction typically used: Slobin (2004) proposed that the attention that speakers of satellite-framed languages pay to the coding of manner information has given rise to an ever-expanding manner verb lexicon. Fanego (2012: 51ff) has shown that the conversion of sound emission verbs, including verbs such as *rustle*, *whizz*, and *zoom*, is an important strategy to derive new manner verbs in Late Modern English (1700–1900). It seems that English is highly prolific in its creation of manner verbs in general (Fanego 2012). As this paper is concerned with the rate of evolution of motion verbs, the employment of non-motion verbs as manner verbs is relevant as it illustrates the strategies languages may employ to generate larger manner verb lexicons. The third tier manner verbs comprise five subclasses: 1. stationary movement verbs (*wriggle*, *flutter*); 2. sound and light emission verbs (*pop*, *rattle*); 3. verbs used in a metaphorical sense (*skim*, Dutch *scheren* lit. “to shave”); 4. transitive manner of motion verbs (Romanian *iuți* “to make something go faster”) and 5. other verbs (*struggle*, *help*).

The class of path verbs can similarly be divided into first tier and second tier path verbs (see also Verkerk 2014b). In the current dataset, certain types of paths seem to be encoded by a verb in almost all languages included in the current paper,

and these are classed as first tier path verbs. This included LEAVE, ARRIVE/REACH, FALL, RISE/GO UP, RETURN and CROSS. Other types of paths, including FOLLOW, APPROACH, EXIT and ADVANCE, among many others, are lexicalized less often and can therefore be said to be more specific or elaborate. In addition, these are typically found in verb-framed languages, and not in satellite-framed languages. These are classed as second tier path verbs. This division between first and second tier path verbs will be helpful to discern differences in etymological origins.

Lists of all the unique types of manner verbs and path verbs included in this study are presented in Appendices 1 and 2 (<http://www.dx.doi.org/10.1075/dia.32.1.03ver.additional>). Table 1 summarizes these Appendices by providing the size of the manner verb lexicon and the path verb lexicon for each language of the sample, separated by tier.

Table 1. Number of types of manner verbs and path verbs taken from the parallel corpus in 21 Indo-European languages.

Language	M 1T	M 2T	M 3T	M tot.	P 1T	P 2T	P tot.
Dutch	12	11	6	29	7	5	12
English	11	10	6	27	8	6	14
German	15	9	5	29	10	4	14
Swedish	12	9	3	24	4	5	9
French	12	4	3	19	13	12	25
Italian	12	3	1	16	12	13	25
Portuguese	11	6	2	19	12	18	30
Romanian	10	6	3	19	13	14	27
Russian	10	8	1	19	7	7	14
Polish	15	10	4	29	11	8	19
Serbo-Croatian	14	6	2	22	17	10	27
Lithuanian	12	10	4	26	9	6	15
Latvian	14	12	4	30	10	6	16
Persian	10	3	0	13	9	8	17
Hindi	11	6	4	21	5	7	12
Nepali	12	2	4	18	9	5	14
Assamese	10	2	1	13	12	14	26
Albanian	10	4	4	18	13	12	25
Greek	13	4	2	19	11	11	22
Armenian	11	3	3	17	8	10	18
Irish	8	6	1	15	8	3	11

M: manner verbs; P: path verbs; 1T: 1st tier; 2T: 2nd tier; 3T: 3rd tier; tot.: total.

Table 1 indicates that languages vary quite widely in their manner verb and path verb lexicon size. Some languages have a large manner verb lexicon and a small path verb lexicon, such as Swedish, while other languages have a small manner verb lexicon and a large path verb lexicon, such as Portuguese, and still other languages have a small manner verb lexicon and a small path verb lexicon, such as Irish. For manner verbs, the differences between languages are largely due to variation in the second and third tier rather than the first tier: languages have a similar number of first tier verbs (range: 8–15) and a more varied number of second and third tier verbs (range: 3–17). Path verb lexicons typically show a more equal distribution of first and second tier verbs.

Given the hypotheses about the evolution of manner verbs and path verbs, we might expect first tier manner verbs and path verbs to emerge early and remain stable through time as they are the most frequently-used motion verbs, while more specific or expressive manner verbs and path verbs emerge later. There is of course no a priori reason why PIE would have had no second tier manner verbs or path verbs. However, given that the first tier motion verbs seem to be the most basic and therefore the most frequently used, second and third tier motion verbs could have been lost more easily as languages shifted their motion event encoding profile from more satellite-framed to more verb-framed, only to be gained once more through a shift from more verb-framed to more satellite-framed. The results of the etymological analysis, including a discussion on differences between first, second and third tier verbs, is presented in the next section.

3. An etymological analysis of manner verbs and path verbs in 21 Indo-European languages

3.1 Etymological classifications

The first sources of information on where manner verbs and path verbs come from are the etymologies of the verbs in question. The etymologies of all attested 442 manner verbs and 392 path verbs were found by consulting etymological dictionaries, other written source materials and, in certain cases, with experts. The source materials used for each language are listed in Appendix 3 (<http://www.dia.doi.org/10.1075/dia.32.1.03ver.additional>).

The etymology of each verb was classified into one of the following categories. Note that these categories are necessarily a simplification, and do not always do justice to the intricacies of the histories of words. However, distinguishing the following classes was required for a cross-linguistic comparison:

1. language-internal development
 - 1a. unspecified / semantic shift
 - 1b. derived from a non-verbal element (noun, adjective, adverb, etc.)
 - 1c. borrowed
2. subgroup-internal development
3. Indo-European-internal development
4. lexicalized preverb-verb combination
 - 4a. inherited lexicalized preverb-verb combination
 - 4b. modern lexicalized preverb-verb combination
5. complex verb
6. unknown etymology

The first category, language-specific development, is divided into three subcategories: 1a. unspecified / semantic shift; 1b. derived from a non-verbal element; 1c. borrowed. All three designate developments ‘within’ the language, i.e., developments that have taken place in the period since the language split away from its closest sister language. An example of the first subcategory, unspecified internal development / semantic shift, is the Albanian manner verb *nxitoj* “to hurry”, which is a reflex of the PIE verb $*g^wej-3$ “to live” (Orel 1998: 303). The manner of motion sense “hurry” is not attested in any of the IE cognates of $*g^wej-3$ except for *nxitoj*, and therefore it seems to represent an independent semantic shift that took place only in Albanian. An example of the second subcategory, derivation from a non-verbal element, is the Italian manner verb *camminare* “to walk”, derived from the noun *cammino* “road” (Dizionari sapere.it 2012). An example of the third subcategory, borrowing, is the Dutch manner verb *spurten* “to scurry”, borrowed from English *spurt* (de Vries 1971).

The second category, subgroup-specific development, is used for verbs that have developed within one of the big IE subgroups: Germanic, Romance, Balto-Slavic and Indo-Iranian. For a verb to be classified in this class, there must exist at least one cognate within the subgroup that is also a manner verb or a path verb, without there being a relevant IE history outside of the subgroup. Examples are the Hindi path verb *girnā* “to fall” and the Nepali path verb *girnu* “to fall”. These two verbs are related to Sanskrit *girati*, which means “devours, swallows” rather than “falls”, and ultimately to PIE $*g^wo/erh_3-o/h_2-$ “devouring, devoured” (Turner 1962–1985, Uhlenbeck 1899: 80, de Vaan 2008: 690). None of the cognates outside of Indo-Iranian are therefore path verbs, suggesting that the change in meaning from “devour” to “fall” has happened only within the Indo-Iranian subgroup. If the cognate verbs are related to a non-verbal element in a related ancient language, such as the French path verb *monter* “to ascend” and the Catalan path verb *montar* “to ascend”, which are ultimately related to Latin *mons, mont-* “mountain” (von

Wartburg 1922–2003, vol. 6.3: 106), the verbs are classified as subgroup specific developments.

The third category, Indo-European-wide development, includes verbs with cognates that are also manner or path verbs in at least one other IE branch. An example is the Polish manner verb *pływać* “to swim” < *pleh₃(u) “to flee, run, flow, swim”, cognate with Sanskrit *plávate* “to swim, float”, and with English *float* (Derksen 2008: 407, Uhlenbeck 1899: 181).

The fourth category contains those verbs that are (ancient or modern) lexicalizations of preverbs and verbs. As is well known, PIE and the ancient IE languages had a system of path-denoting particles that could be positioned anywhere in the sentence (Delbrück 1888, 1893, Hewson & Bubenik 2006, Hofmann & Szantyr 1965, Lehmann 1974). This system was still in place in Homeric Greek (3, 4). In (3), the path particle *epi* functions as a preposition, and in (4), *epi* functions as a preverb (although it is separated from its verb, a process called ‘tmesis’).

- (3) *epi hoí kalésas... subóten*
 to him.DAT call.AOR.PART swineherd.ACC
 “having called the swineherd to him” (Hewson & Bubenik 2006: 4)
- (4) *kai epi knéphas hierón éltḥēy*
 and on darkness sacred come.3SG
 “and the sacred darkness closes in” (Hewson & Bubenik 2006: 6)

At later stages, namely in Classical Sanskrit, Classical Latin and Classical Greek, these particles became more closely associated with nouns and verbs. The path particles that were closely associated with nouns formed the adpositional systems of contemporary IE languages. The path particles that belonged with verbs formed the preverb systems that later became spatial prefixes. In the transition from Latin to the modern Romance languages, as well as from Sanskrit to the modern Indo-Aryan languages, these spatial prefixes merged with the verb completely. In Balto-Slavic, the system of spatial prefixes is still in place. As a result, lexicalized preverb-verb combinations are an important group of path verbs in many IE languages. Given that the different IE subgroups were all supplied with the same tools to combine preverbs with verbs, there are calques of preverb-verb combinations, both within subgroups and between subgroups. For example Dutch *oversteken* and German *überqueren* both mean “to cross” and are lexicalized combinations of a cognate adverb meaning “across” and a verb. Latin has calqued several preverb-verb combinations from Greek, which is in a different IE subgroup (see for examples Nicolas 2005: 131 and also Adams 2003: 459ff). The extent to which these lexicalized preverb-verb combinations are independent inventions or calques can be hard to assess, especially as most etymological dictionaries think of these as

derivations and often do not pay much attention to highly similar derivations in closely related languages.

The fourth category is divided into two subcategories, of inherited lexicalized preverb-verb combinations and modern lexicalized preverb-verb combinations. An example of an inherited lexicalized preverb-verb combination is the Hindi path verb *nikalnā* “to exit” and the Nepali path verb *niskanu* “to exit” from the Sanskrit verb *niṣkāsayati* “to drive out, away”, a preverb-verb combination with *niṣ* “out, away” (Turner 1962–1985, Uhlenbeck 1899: 149). Verbs in this category must either have contemporary cognates that clearly originate from the same preverb-verb combination, such as Hindi *nikalnā* and Nepali *niskanu*, or the prefixal derivation of the original root verb should be listed in the source material. An example of a modern lexicalized preverb-verb combination is the Portuguese path verb *retornar* “to return”, a combination of the prefix *re-* and the verb *tornar* “to turn” that has become lexicalized since Portuguese split from its nearest relative (Machado 1952, vol. 5: 92). The modern prefix-verb lexicalization should have taken place only for that contemporary language, i.e., at a time after the language has split away from its most closely related sister languages. Note however that since several Romance languages have verbs similar to Portuguese *retornar* “to return” (such as French *retourner*), the Portuguese verb has probably been influenced by such verbs and is at least partly a calque. In addition, verbs in this category must no longer have transparent semantics, i.e., the verb in its contemporary use must be monomorphemic. If prefix-verb combinations were still semantically transparent and non-monomorphemic, the verb and the preverb were analyzed separately, i.e., the preverb was analyzed as a path satellite and the verb (in most cases) as a manner verb or a deictic verb.

The fifth category, complex verbs, contains verbs that are not monomorphemic but that are not phrasal either. Examples are the many Persian light verbs such as *qadam gozāštan* “to step”, *dākel šodan* “to enter” and *tark kardan* “to leave” and similar constructions in other languages, such as Hindi *pār karnā* “to cross”. It is often difficult to assess the etymology of such complex verbs. Information is often given on the non-verbal element and the verb separately, but not on the complex verbs themselves. Therefore these complex verbs have been placed in a separate category of unknown etymologies.

If no etymology has been found, or the etymology is described as unknown or difficult according to reputable source material, the verb is placed in category six, unknown etymology.

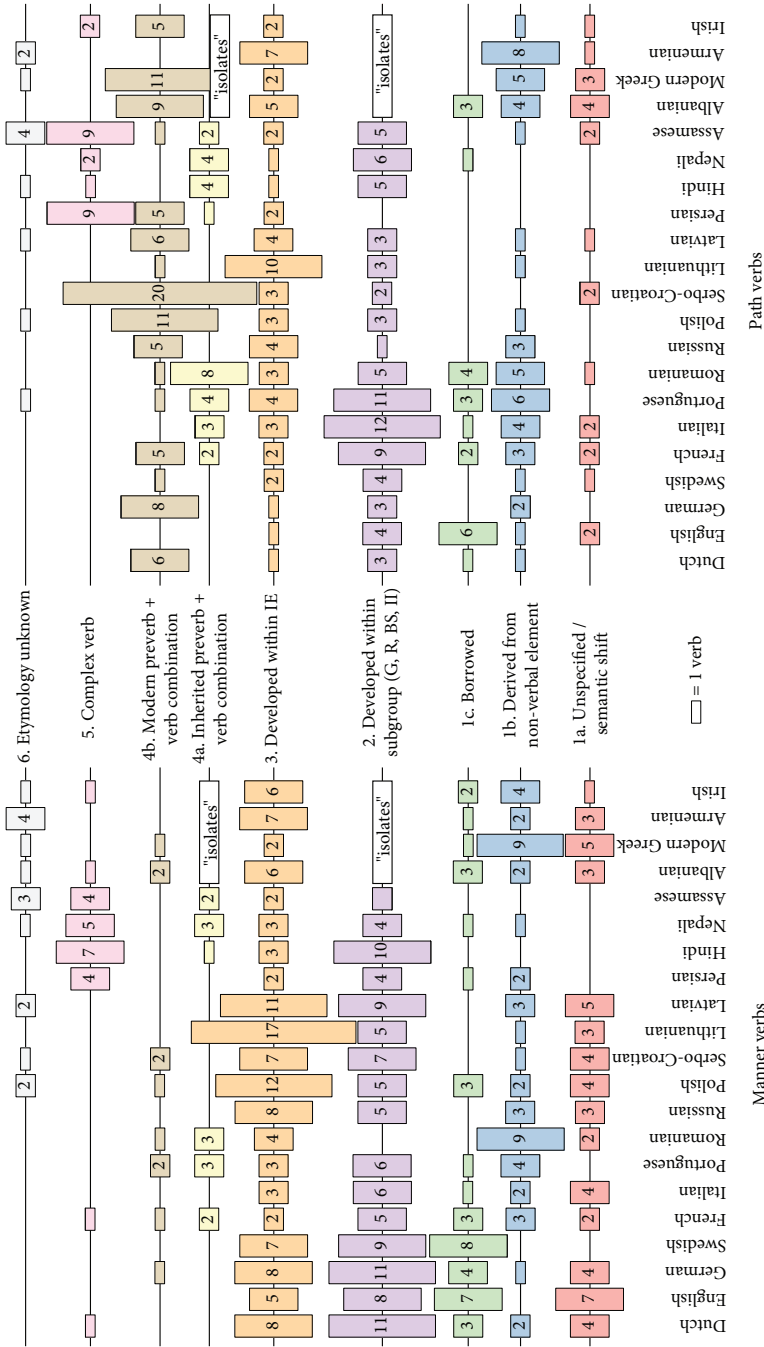


Figure 1. The etymologies of the manner verbs and the path verbs taken from the parallel corpus in 21 Indo-European languages. Isolates are those languages which either form a subgroup by themselves (Albanian, Modern Greek, and Armenian) or which are the only language from a subgroup included in the sample (Irish). For these languages, class 2 and 4a are not relevant.

3.2 Manner verbs and path verbs have different etymologies

The results of the etymological investigation of manner verbs and path verbs are depicted in Figure 1 and Appendices 4–7 (<http://www.dx.doi.org/10.1075/dia.32.1.03ver.additional>). Figure 1 gives an overview of the etymologies of manner and path verbs for the 21 IE languages. Appendices 4 and 5 give summaries of etymology types across the big IE subgroups (Germanic, Romance, Balto-Slavic and Indo-Iranian). Appendix 6 gives a summary of etymology types per tier for both the manner verbs and the path verbs. Appendix 7 presents the results of paired sample t-tests that indicate whether the cross-linguistic distributions of etymological categories of manner verbs and path verbs are significantly different or not.

Figure 1 shows that the most notable differences between manner verbs and path verbs are as follows. Manner verbs often originate at the PIE level and at the subgroup level (see Figure 1). On average, 27% of the manner verbs can be traced back to PIE, compared to only 17% of the path verbs (see Appendices 4 and 5). This difference is statistically significant (see Appendix 7; $t=4.43$, $p\leq 0.001$). In contrast, 23% of the manner verbs can be traced back to the subgroups, compared to 21% for the path verbs (see Appendices 4 and 5). This difference is not statistically significant (see Appendix 7; $t=1.42$, $p=0.17$). It also seems that manner verbs have approximately the same distribution of origin types across languages, while path verbs have much more varied origins from language to language. This suggests that in IE, manner verbs are more stable than path verbs: they are typically older and have similar origins across languages.

There are also differences between the etymological origins of manner verbs across the different languages (see Figure 1) and subgroups (Appendix 5). Germanic has the largest proportion of manner verbs that originate at the subgroup level (36%, compared with 23% overall, see Appendix 5). This can be seen as evidence for a faster rate of change in the creation of manner verbs at the Proto-Germanic level. Proto-Germanic is likely to have been satellite-framed (Verkerk 2014a), which could have created pressure to develop more manner verbs before it split up into the different Germanic language groups. Germanic is the only clade where the derivation of manner verbs from non-verbal elements is not attested at all. This is a fairly common process in verb-framed languages such as Portuguese, Romanian and Modern Greek, and it is also common in Irish. Borrowed manner verbs are most common in Germanic, French and Albanian. Lexicalized preverb-verb combinations are not very common origins for manner verbs, although they are attested in Romanian, Portuguese, Nepali and Assamese. The large number of Lithuanian manner verbs that have an IE origin is probably due to the fact that Lithuanian is very conservative in many other respects as well, and also that

Lithuanian etymologies are the best-studied etymologies of all Balto-Slavic languages.

As for path verbs, their etymologies are generally far more diverse if we compare them across languages (see Figure 1). The formation of path verbs through lexicalized preverb-verb combinations is a very important source of path verbs. Compared to manner verbs, significantly more path verbs are formed through inherited preverb-verb combinations (see Appendix 7; $t = -2.26$, $p \leq 0.05$) and modern preverb-verb combinations (see Appendix 7; $t = -3.93$, $p \leq 0.001$). The formation of path verbs through inherited preverb-verb combinations is attested in all Romance and Indo-Iranian languages. The formation of path verbs through modern preverb-verb combinations is attested in Dutch, German, Swedish, French, Portuguese, all Balto-Slavic languages, Persian, Albanian, Modern Greek and Irish. This indicates that in these languages, some form of the ancient preverb system is still productive or has been productive until recently. Both the Romance and the Indo-Iranian languages have many path verbs with an origin at the subgroup level: 35% of the Romance path verbs and 26% of the Indo-Iranian path verbs originate at the subgroup level, compared with 21% overall (see Appendix 4). This can be regarded as evidence for an increase in the number of path verbs in the Romance and Indo-Iranian subgroups. Germanic also shows an increase in the number of path verbs at the subgroup level: 30% of Germanic path verbs originate at the subgroup level, compared with 21% on average. English has borrowed almost half of its path verbs (Aske 1989). The large number of Lithuanian path verbs that have an IE origin is again likely to be due to the conservative nature of Lithuanian as well as its etymologies being well studied. The lack of preverb-verb combinations in Armenian can be explained by the fact that even in Classical Armenian, preverb-verb combinations were rarely used (Schmitt 1981).

The current analysis only provides a tentative answer to the question whether second tier verbs are typically newer and whether first tier verbs are typically older. Since satellite-framed languages have few path verbs and verb-framed languages have few manner verbs, the two groups are difficult to compare as the total number of verbs is not similar. Even so, a summary of etymologies separated by tier is presented in Appendix 6. Across the whole IE family, 32% of the first tier manner verbs emerge at the IE level, compared to 25% of the second tier manner verbs and 22% of the third tier manner verbs. Second and third tier manner verbs are slightly more often derived from a non-verbal element (2nd tier: 14%; 3rd tier: 19%) or borrowed (2nd tier: 14%; 3rd tier 17%) than first tier verbs (derived from a non-verbal element: 9%; borrowed: 7%). Comparing the big subgroups, it is certainly not true that second tier manner verbs emerge on the subgroup level and that first tier manner verbs emerge at the IE level in Germanic and Balto-Slavic languages. Germanic and Balto-Slavic second tier manner verbs are just as likely to originate

at the IE level as they are to originate at the subgroup level. However, all Romance manner verbs that emerge at the IE level (namely, Portuguese *voar*, Italian *volare* and French *voler* “to fly”, Italian *saltare*, Portuguese *saltar*, French *saute* and Romanian *sări* “to jump”, Portuguese *nadar*, Italian *nuotare* and Romanian *înota* “to swim”, as well as Romanian *umbla* “to wander”), except for Romanian *fugi* “to flee”, are first tier manner verbs.

For the Indo-European path verb lexicons, no evidence can be found for the early emergence of first tier path verbs and the later emergence of second tier path verbs. Appendix 6 indicates that slightly more first tier path verbs have an IE origin (18%) as compared with second tier path verbs (15%). But this difference is very small. And in contrast, 31% of first tier verbs are modern preverb-verb lexicalizations, as opposed to 17% of second tier verbs. The only evidence pointing in the direction of the late emergence of second tier verbs is that they are more often derived from a non-verbal element (18% of second tier as opposed to 7% of first tier path verbs). Note, however, that the division of first and second tier path verbs has been made only on the basis of path verbs attested in the current study, and it might not be a good representation of internal path verb class divisions. Lexicalization of preverb-verb combinations has been so common that it has affected all IE languages and has given rise to widespread parallel evolution and calquing of path verbs. The various IE verbs meaning “to return”, for instance, are sometimes similar sounding but actually partly parallel developments and partly calques: French *retourner*, Portuguese *retornar*, Dutch *terugkeren*, German *zurückkehren*, Swedish *återvända*, Greek *epistrefo*, Persian *bāzgaštan* and Russian *vozvrašat'sja* are all based on the combination of a preverbal element with a verb meaning “turn”.

The results suggest that manner verbs and path verbs typically have quite different histories. The question addressed in the next section is whether inferring the rates of evolutionary change can inform us further on the evolutionary processes that underlie the verb histories.

4. Rates of evolutionary change for manner verb and path verb lexicon size

This section investigates the hypotheses that manner verb lexicon size evolves faster in satellite-framed languages and path verb lexicon size evolves faster in verb-framed languages. These hypotheses were already supported by the large number of manner verbs that arise at the subgroup level for Germanic languages, and the large number of path verbs that arise at the subgroup level for Romance and Indo-Iranian languages, as presented in §3.2. In this section, it is shown that

similar patterns emerge when the rates of evolutionary change are investigated. The phylogenetic tree that is vital to this type of analysis is presented in §4.1. The optimal rates of change of manner verb lexicon size and path verb lexicon size are inferred in §4.2.

4.1 The phylogenetic tree

A phylogenetic tree is a representation of the genetic descent of languages: it shows to which languages a particular language is conceived of as being most closely related to, and how the different clusters of closely related languages are grouped in relation to one another. In this paper, a phylogenetic tree is transformed in various ways, in order to see which model of evolution provides the best fit to the data. The sample of phylogenetic trees that was used for this task comes from Bouckaert et al. (2012), whose tree set was built using the cognate sets proposed by Dyen et al. (1992) and Ringe et al. (2002), which were thoroughly checked and updated for the purposes of their study. Bouckaert et al.'s (2012) tree set therefore is one of the latest proposals for IE based on the best dataset prepared so far. The tree set conforms well with what we know about the IE family from other phylogenetic studies, such as Nakhleh et al. (2005). All subgroups (Anatolian, Tocharian, Armenian, Hellenic, Albanian, Indo-Iranian, Balto-Slavic, Germanic, Italic and Celtic) are recovered. In addition, the first split in the tree is between Anatolian and the other IE languages, and the second split is between Tocharian and the rest of the family, which is in accordance with what we know about the higher order subgrouping of Indo-European (Clackson 2007: 13, Beekes 2011: 30–31).

Bouckaert et al.'s (2012) trees were built using data from 103 IE languages, consisting of lists of lexical items coded for cognacy. These lists were recoded in a binary fashion, resulting in a matrix that indicated whether each language had a reflex in each of the 5047 individual cognate sets. A distribution of phylogenetic trees was then estimated using a Bayesian Markov Chain Monte Carlo approach (Huelsenbeck et al. 2001), available in the software BEAST (Drummond et al. 2012). The trees were time-calibrated by constraining lineage divergence times and by using probability distributions of ancient language ages based on historical sources. 12,500 trees with high likelihoods were randomly selected from the posterior tree distribution.

The application of phylogenetic methods adopted from biology to study genealogical relationships between languages and diachronic change of language features has been a topic of debate in historical linguistics and typology (see for instance *Linguistic Typology* vol. 15(2)). The trees proposed by Bouckaert et al. (2012) likewise have been controversial. Criticism has focused on the use of cognate data rather than regular sound changes, problems with the dates used for

some of the calibration points and with the geographic location of certain languages. However, the current paper is not concerned with Bouckaert et al.'s (2012) specific hypotheses regarding the geographical expansion of IE through space and time. Rather it uses the structure of their phylogenetic trees to study the rates of change of motion verbs. Since Bouckaert et al.'s (2012) trees are the only published trees built using the improved Dyen et al. (1992) dataset (Dunn et al. forthcoming), they are the best choice at this time.

For this paper, I made a further random subselection of 1000 trees from Bouckaert et al.'s (2012) sample of 12,500 trees. This random subselection was pruned so that the trees included only the 21 IE languages featured in this paper. Then, a maximum clade credibility tree of the sample was calculated using TreeAnnotator v.1.6.1 (Drummond et al. 2012). This maximum clade credibility tree, which has median branch lengths, is presented in Figure 2. This maximum clade credibility tree is used in the further phylogenetic analysis discussed below. The support values presented over each internal node of the tree give an idea of the structure of the trees in the tree sample. An internal node that is attested in all trees in the sample is marked with 1 and the number is less than one if this internal node is only attested in a subset of the trees in the sample. The internal node that leads to the Romance, Celtic, Germanic and Balto-Slavic subgroups has a support value of 0.99, indicating that it appears in 99% of the trees in the sample.

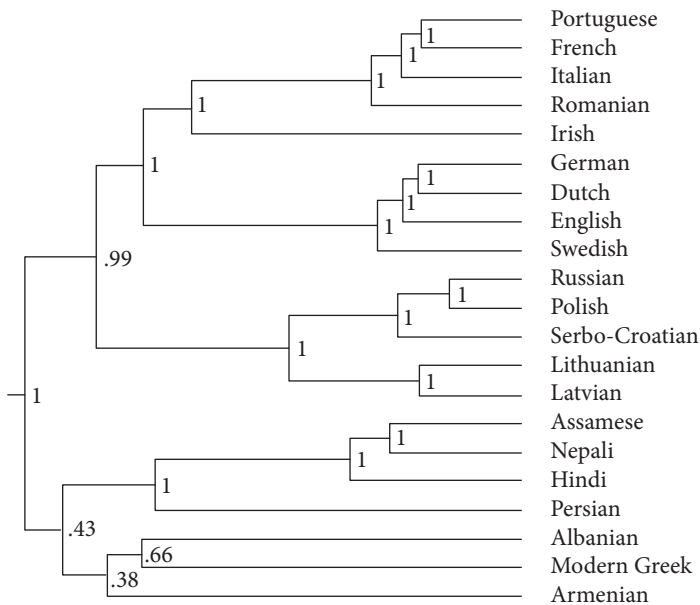


Figure 2. The maximum clade credibility tree of the tree sample from Bouckaert et al. (2012).

Note that the support values for the internal nodes that connect Indo-Iranian, Albanian, Armenian and Modern Greek are low, suggesting that the position of these languages within the tree is uncertain. Note also that Albanian, Armenian and Modern Greek are not a subgroup like Germanic, as indicated by the long branches leading to these three languages. Since the trees were time-calibrated but no ancient languages are included in the current sample, the trees are ultrametric, i.e., the distance from the root to the tip of each individual branch is exactly the same.

4.2 Testing for differential rates of change

The standard model of evolution of continuous features (i.e., features that range on a scale rather than in a number of separate classes, such as lexicon size) is the constant-variance random walk model of evolution (Pagel 1999), also sometimes called ‘Brownian motion’. In this model, the constant rate of change is a measure of how much a feature changes instantaneously at each moment of time. For a given data set and phylogenetic tree, the rate of change is integrated over evolutionary time as represented by branch length. Traditional ‘standard’ phylogenetic analyses estimate a single constant rate of change, while taking into account all sub-branches and data points (Pagel 1999). This is simply in accordance with the simplest model of evolution. This paper is concerned with more advanced phylogenetic analysis that allows the rate of change to differ in different parts of the tree. But before introducing this type of analysis, I illustrate what the rate of change on a phylogenetic tree signifies.

In Figure 3, an ultrametric phylogenetic tree is depicted that represents the evolution of manner verb class size in four hypothetical languages called A, B, C and D. Manner verb class size for each of these languages is listed to the right of

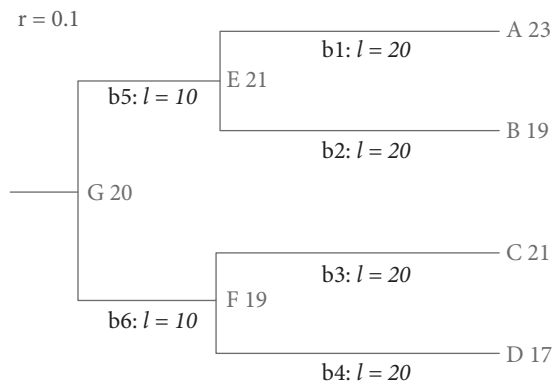


Figure 3. A single rate of change of manner verb lexicon size in a hypothetical data set.

the names A, B, C and D. Languages A and B are descendants of language E, while C and D are descendants of F, which are marked by 'E' and 'F' on the respective internal nodes of the tree. All languages descend from language G, which is also marked on the root of the tree. Let us assume that we know the ancestral states of the languages E, F and G — the numbers to the right of the names E, F and G give the ancestral manner verb lexicon size of each of these ancestral languages. The sub-branches are labeled b1 through b6. Since the (sub-)branches are a representation of the evolutionary process through which languages A, B, C and D emerged, branch length is a representation of the evolutionary timeline. Evolutionary time can be measured in different ways. For phylogenetic trees which have not been time-calibrated it is typically a representation of the amount of linguistic change, i.e., the number of cognate replacements. For time-calibrated trees evolutionary time typically is given in years. In Figure 3, however, the unit of evolutionary time is a generation, and the length of each sub-branch is given under each sub-branch in generations (a generation being approximately 20 years). Given that we know the contemporary states, the ancestral states and the length of each sub-branch in generations, the rate of change in each sub-branch that is needed to generate the distribution of contemporary manner verb lexicon sizes is 0.1. For example, for language A to have 23 manner verbs, it must have gained 0.1 manner verb per generation (or 1 verb per 10 generations) since its ancestor language G, 30 generations ago.

Note that the rate of change does not specify the direction of change: the evolutionary model depicted in Figure 3 specifies an increase of 0.1 per generation along branches b5 and b1 to model the contemporary manner verb lexicon size of language A, while the model relates a decrease of 0.1 per generation along branches b6 and b4 to model the contemporary manner verb lexicon size of language D. Change along the branches of a phylogenetic tree can, at any moment in time, both be a value increase as well as a value decrease, as the lexicon can grow or shrink. The rate of change simply reflects the pace at which such changes are taking place given a unit of evolutionary time.

In this example, a single constant rate of change suffices to explain the different manner verb lexicons of our hypothetical languages A, B, C and D. However, many linguistic features do not evolve at a constant rate. If the comparative dataset has evolved in a particular directional trend, or if the rates of change vary from time period to time period, the fit between the dataset and the phylogenetic tree will not be optimal when only a single constant rate of change is estimated. The fit between the dataset and the tree will typically improve if the phylogenetic tree is adjusted to take into account information on varying rates of change or directional trends. An example of a dataset that can be explained by employing a set of different rates of change is given in Figure 4.

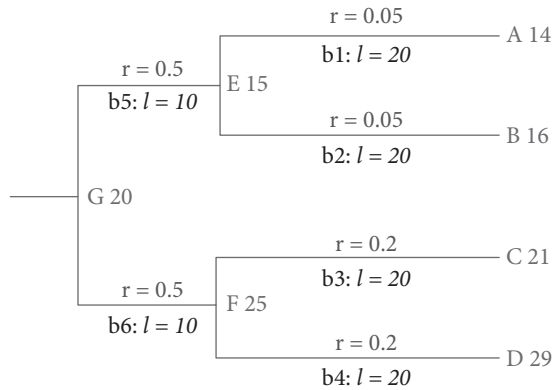


Figure 4. Different rates of change of manner verb lexicon size in a hypothetical data set.

In Figure 4, a different tree and set of languages are depicted, the languages are again named A through G. The manner verb lexicon sizes of these languages vary more widely than those depicted in Figure 3, and different rates of change are required to fit this diversity on the tree. In branches b5 and b6, the rate of change is fast, 0.5 per generation (or 1 verb per 2 generations), whereas the rates of change in branches b1 and b2 are slow, 0.05 per generation (or 1 verb per 20 generations). In the history of language A, it has first lost verbs at a fast rate (0.5) on branch b5, and then at a more slower rate (0.05) on branch b1, in total having lost 6 verbs compared to its ancestor language G.

In the current paper, rates of change of manner verb lexicon size and path verb lexicon size were optimized using the maximum clade credibility tree presented in §4.1 and Figure 2. In this way it was tested whether there was any evidence for larger rates of change of manner verb lexicon size in branches leading to satellite-framed languages versus larger rates of change in path verb lexicon size in branches leading to verb-framed languages. This was achieved by modeling the evolution of the two datasets on lexicon size on the maximum clade credibility tree. The modeling was done using *BayesTraits*, which uses a Phylogenetic Generalized Least Squares model of trait evolution (Pagel 1997, 1999) in a Bayesian reversible-jump Markov chain Monte Carlo framework (Green 1995) that can be used to trace the evolutionary history of shifts in the rate of evolution (Venditti et al. 2011). Given a phylogenetic tree and a dataset, the Bayesian MCMC algorithm allows for the optimization of the rate of change parameters to find the best model of evolution of the data on the phylogenetic tree.

To see how the Bayesian MCMC algorithm accomplishes this, let us look at another hypothetical dataset, this time consisting of five related languages A, B, C, D and E. Imagine that language A has 16 manner verbs, B has 17, C has 20, D has 30 and E has 24. We also know from their phylogeny that languages A, B and C are

closely related, and that D and E are closely related. Thus, the group of languages A, B and C has smaller manner verb lexicons (16–20 verbs) compared to the group of languages D and E (24–30). This may be due to increased or decreased rates of change for one of the two language groups or for the individual languages. In order to test this, the Bayesian MCMC algorithm can be used to find the most optimal shared and non-shared branch lengths of the phylogenetic tree of languages A, B, C, D and E. From these optimized branch lengths, the optimized rates of change can be derived: branches that have been scaled to be longer imply increased rates of change, while branches that have been shrunk to be shorter imply decreased rates of change. Technical aspects are explained below.

In theory, the lengths of the branches of the tree can have an almost infinite number of values. However, some values will be far more likely than others — the aim of the Bayesian MCMC algorithm is to find those values which have the highest likelihood. It searches the so-called parameter space that contains all the possible combinations of values for the different parameters that describe the phylogenetic tree with its branch lengths. The algorithm does this by building a Markov chain, a mathematical device that jumps from state to state. For our hypothetical dataset, a state would be the phylogenetic tree for languages A, B, C, D and E, with some parameter values describing the branch lengths of this tree. This process is illustrated in Figure 5. The dark line in Figure 5 represents the MCMC chain that is jumping from state to state within the parameter space, beginning at the top and moving downwards. Each numbered tree is a state; each change from one

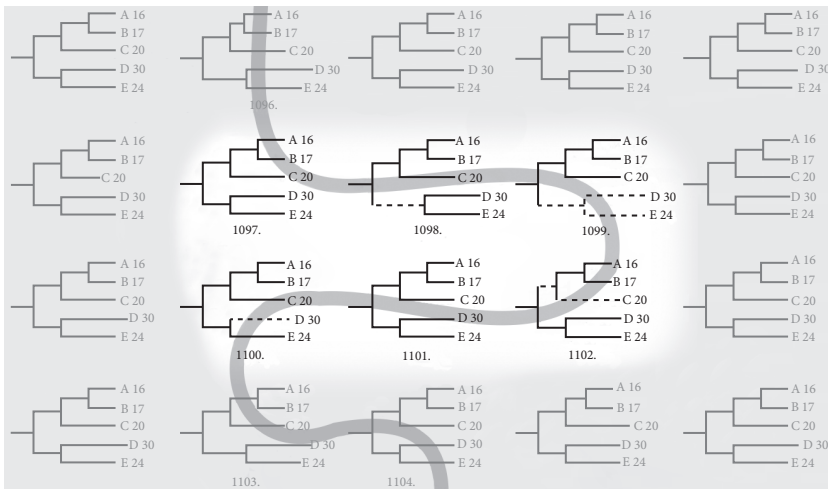


Figure 5. A MCMC chain that searches the parameter space to optimize branch lengths. The dark line represents the MCMC chain; the numbered trees represent the different states on the chain.

numbered tree to the next on the chain represents a jump. The probability of a jump to a new state is determined by how much of an improvement in likelihood the new state is over the current state. New states that have a lower likelihood than the current state may be adopted as well, depending on how much worse it is compared with the current state.

In the parameter space in Figure 5, the trees have the same topological structure, only the length of the various sub-branches is different. The MCMC chain samples the parameter space by making changes in branch length and jumping to a new state (a tree with different branch lengths) when the proposed new state is a sufficient improvement. In Figure 5, changes in branch length are marked by the dashed branches. Branch length changes are made by optimizing two parameters: one that changes the length of a single branch (single branch modification, as in tree no. 1098, 1100 and 1102 in Figure 5), and one that changes the length of a branch and all its descendent branches (clade modification, as in tree no. 1099). Note that the original, unchanged tree is also part of the parameter space, and can be visited by the chain again and again if it has a high likelihood (tree no. 1097 and 1101 have undergone no branch length changes). When the chain is run for an appropriate number of iterations, it becomes stationary. In stationary distribution, the chain no longer moves to different phylogenetic trees with higher and higher likelihoods. The proposed phylogenetic trees all have high likelihood values. It then samples phylogenetic trees in proportion to their frequency of occurrence in the tree space. In this way it constructs a sample of phylogenetic trees that constitute the posterior probability distribution. The result is a sample of trees that have the most optimal branch lengths given the dataset, which in turn relate the optimal rates of change given the dataset. Such a sample is then further summarized into a single tree by taking the mean branch lengths over the posterior tree sample. The resulting tree has optimized branch lengths that reflect the optimal rates of change given the data: if the rate of change in a specific branch has been lower, that branch will have become shorter; if the rate of change was inferred to be higher, that branch will have become longer. For instance, for our hypothetical dataset, branches for languages D and E are scaled to be longer in tree no. 1098, 1099, 1100 in Figure 5, and therefore the mean branch lengths of these languages in the optimized tree will probably become longer. This implies that these languages have an increased rate of change of manner verb lexicon size.

Now let us get back to the real dataset. In order to investigate whether path verb lexicon size and manner verb lexicon size change at different rates, it is necessary to test and compare two models: the null model (constant rate model) and the alternative model (variable rate model). The null model states that the evolution of the size of the manner verb lexicon and the size of the path verb lexicon has taken place at a constant rate throughout the IE family. The alternative model is

that the rate of evolution has varied in different sections of the phylogenetic tree. To investigate whether the alternative model is an improvement on the null model, four analyses were conducted: for manner verb lexicon size, one analysis in which the rate of change was kept constant over the tree (the null model M_f) and one analysis in which it was allowed to vary (the alternative model M_v), and in turn for path verb lexicon size, one analysis in which the rate of change was kept fixed (the null model P_f) and one in which it was allowed to vary (the alternative model P_v). For all four analyses, the MCMC chains were run for 2×10^9 iterations. The phylogenetic trees were sampled every 10^6 iteration. A posterior of 1500 samples was taken from the stationary part of the chain. These were then further summarized by finding the mean values for the adjusted branch lengths and the rate of change parameters on different branches (see below).

To see whether the alternative model is an improvement on the null model, the harmonic means of the Bayesian posterior density of likelihoods are compared using Bayes factors (Pagel et al. 2004, Kass & Raftery 1995). The results are summarized in Table 2, which shows that for the manner verb lexicon size analysis, the variable rate model (M_v) is favored, while for the path verb lexicon size analysis, the fixed rate model (P_f) is favored. However, neither of these results are strong evidence for the existence of such tendencies, as the Bayes factor values are low. As we can see from the right-most column in Table 2, the improvement of the alternative model in which the rate of change is allowed to vary in different parts of the tree over the null model is statistically not relevant for the analysis of manner verb lexicon size (Bayes factor is between -2 and 0). For path verb lexicon size, the null model is preferred over the alternative model, with a Bayes factor between 2 and 6 . This can be considered negative evidence for the existence of variable rates of change for path verb lexicon size. However, since the difference between the two models is still very small (Bayes factor 3.06), the evidence is inconclusive.

Table 2. Comparison of the constant and variable rate models of evolution of motion verb lexicon size in 21 Indo-European languages.

Dataset	Fixed model	ln Lh ^a	Variable model	ln Lh	BF ^b
Manner verb lexicon size	M_f	15.40	M_v	15.63	-0.46
Path verb lexicon size	P_f	7.28	P_v	5.75	3.06

^a The ln Lh denotes the marginal likelihood of each analysis. This is the harmonic mean of the ln likelihoods of the posterior 1500 samples taken from the stationary part of the chain.

^b Bayes factors were calculated by taking $BF = 2(\ln Lh M_f/P_f - \ln Lh M_v/P_v)$ (Pagel et al. 2004). Bayes factor values indicate which model is statistically favored. If values are positive, the fixed model is better supported. If values are negative, the variable model is better supported. The Bayes factor score is read as follows: $0-2$, barely worth mentioning; $2-6$, positive; $6-10$, strong; >10 , very strong (Kass & Raftery 1995).

Table 3. Optimized rates of change within the main subgroups of Indo-European on the phylogenetic tree by Bouckaert et al. (2012) scaled to reflect manner verb lexicon size and path verb lexicon size evolution.

Subgroup	Manner verbs	Path verbs
Complete tree	1.14	1.01
Romance	1.07	1.15
Germanic	1.09	1.01
Balto-Slavic	1.32	1.05
Indo-Iranian	1.23	1.07

The inconclusive results are likely to be due to the small number of languages used in the current analysis, which unfortunately cannot be improved upon. However, even though there is no strong evidence that the variable rates model provides a significantly better model of evolution for these two datasets, the results do tell us something about different rates of change in different verb classes and why this methodology is encouraging. First of all, it is interesting to see that the path verb lexicon analysis prefers the constant rate model, whereas the manner verb lexicon analysis prefers the variable rate model (even given that this difference is not statistically significant but rather a trend). Secondly, distinct results for manner verb lexicon size and path verb lexicon size were obtained: the phylogenetic trees with optimized branch lengths and the rates of change that can be derived from them do suggest different rates of change for the two different classes. A summary is presented in Tables 3 and 4 and Figure 6. Table 3 gives the average rates of change within each of the four main subgroups of IE (Germanic, Romance, Balto-Slavic and Indo-Iranian) compared with the average rates of change for the whole tree. Figure 6 shows the maximum clade credibility tree presented in Figure 2 scaled to reflect the optimal rates of change of manner verb lexicon size and path verb lexicon size. Table 4 gives an overview of the percentage of scaling of the branches leading to the individual contemporary languages as presented in Figure 6.

Table 3 relates that overall, the rates of change across the whole tree are higher than 1 for the manner verb and the path verb dataset, which indicates that rates of change have been increasing over time. This is in line with expectations, as the adaptations of rates of change are not a zero-sum game, i.e., the total length summed over all the branches does not need to be the same as that of the original tree. Rather, the Brownian motion model of evolution works to make those branches longer for which an increase of rate of change is most likely, and makes those branches shorter for which a decrease of rate of change, or even complete stasis, is most likely. However, apparent slow rate of change or stasis can be explained

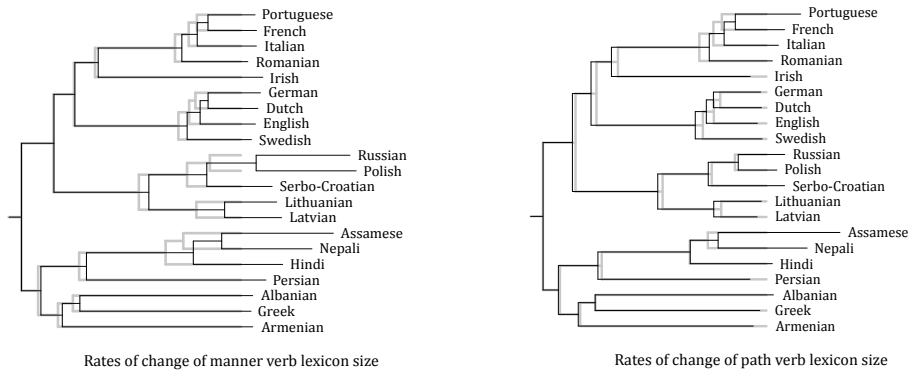


Figure 6. The Indo-European tree by Bouckaert et al. (2012) scaled to reflect manner verb lexicon size evolution (on the left) and path verb lexicon size evolution (on the right). The trees in gray are the original trees; the trees in black are the scaled trees.

on long branches as well as on short branches. Imagine, for instance, that a given language has 16 manner verbs, while we know its ancestor had 15. In such a case, Brownian motion dictates that the language could have lost and gained several verbs, but ultimately was left with one verb more than its ancestor. There is ample time for such accumulated Brownian change on a long branch. The Brownian motion model in this sense ‘prefers’ long branches over short branches, as long branches provide more evolutionary time for changes to occur. However, if subgroups behave highly uniformly (i.e., have very similar manner verb lexicon sizes, for instance), this indicates a decrease in rate or stasis in the whole subgroup, and the branches in that clade may become shorter. For individual branches, a decrease in rate of change will only be found when there is substantial evidence for a slow rate or stasis.

Table 3 and Figure 6 show that in order to optimize the variable rates of change on the phylogenetic tree, all branches are scaled to be longer to some extent for the manner verb lexicon size analysis, while some branches are scaled to be longer and some are scaled to be shorter for the path verb lexicon size analysis. This difference can be related to the comparison of the null model and the alternative model in Table 2: the manner verb lexicon seems to evolve at a higher rate than the phylogenetic tree has predicted and prefers the variable rates model, whereas the path verb lexicon evolves faster in certain parts and slower in other parts, therefore on the whole the constant rate model is the better match.

If we compare the individual branches, such as those leading to the four big IE subgroups (in bold in Table 4) and the individual languages (Figure 6 and Table 4), we find that some branches are optimized to be much longer to account for manner verb lexicon size while others are estimated to be longer to account for path verb lexicon size. Table 4 displays the variability in branch length

Table 4. Transformation of branches of the phylogenetic tree from Bouckaert et al. (2012) scaled to reflect manner verb lexicon size and path verb lexicon size evolution.

Branch leading to	Manner verbs	Path verbs
Proto-Germanic	109%	99%
Dutch	109%	98%
English	108%	97%
German	110%	98%
Swedish	106%	99%
Proto-Romance	106%	121%
French	109%	109%
Italian	108%	109%
Portuguese	108%	116%
Romanian	105%	103%
Proto-Balto-Slavic	112%	99%
Russian	152%	109%
Polish	155%	105%
Serbo-Croatian	116%	109%
Lithuanian	118%	98%
Latvian	120%	113%
Proto-Indo-Iranian	113%	95%
Persian	113%	93%
Hindi	121%	103%
Nepali	134%	119%
Assamese	144%	134%
Albanian	107%	104%
Modern Greek	106%	98%
Armenian	107%	95%
Irish	111%	93%

optimization. The branches of Germanic languages, including the branch leading to Proto-Germanic, are scaled to be longer (Table 4) and have higher average rates of change (Table 3) to account for manner verb lexicon size evolution than for path verb lexicon size evolution. The branches of Romance languages, including the branch leading to Proto-Romance, are scaled to be longer and have higher average rates of change to account for path verb lexicon size evolution than for manner verb lexicon size evolution. The branches of Balto-Slavic languages, including the branch leading to Proto-Balto-Slavic, are scaled to be longer and have higher

average rates of change to account for manner verb lexicon size evolution than for path verb lexicon size evolution. These results suggest that the Germanic and Balto-Slavic satellite-framed languages have higher rates of evolutionary change for manner verbs, while the Romance verb-framed languages have higher rates of evolutionary change for path verbs.

The branches of Indo-Iranian languages, including the branch leading to Proto-Indo-Iranian, are scaled to be longer and have higher rates of change to account for both manner verb lexicon size evolution and path verb lexicon size evolution, although this is more prominent for manner verbs than for path verbs. This seems to be due to the fact that the Indo-Iranian languages have quite different numbers of motion verbs: Hindi has 21 manner verbs, while Assamese and Persian have 13, whereas Hindi has 12 path verbs, while Assamese has 26. The phylogenetic algorithm estimates a large amount of stretching for this subgroup because these languages, although closely related, have quite divergent motion verb lexicon sizes. Note that we see something similar in the manner verb lexicon size analysis for Polish (29 manner verbs) and Russian (19 manner verbs). More divergence within subgroups or between sister languages requires longer optimized branches in that subgroup.

Following the general pattern within the tree, the branches leading to the remaining four languages, Irish, Albanian, Armenian and Modern Greek, are scaled to be slightly longer to account for manner verb lexicon size evolution and slightly shorter to account for path verb lexicon size evolution. For these four languages, the optimized branch lengths imply slightly accelerated evolution of manner verbs as compared with path verbs. The only exception to this pattern is Albanian path verb lexicon size evolution; the branch leading to Albanian is longer in order to account for its larger path verb class.

5. The evolutionary processes that shape the lexicon

The results of the etymological investigation and the optimizations of branch lengths and rates of change reveal several differences between the evolutionary processes that generate manner verbs and those that generate path verbs in IE:

1. Manner verbs typically have longer histories than path verbs;
2. The process by which preverb-verb combinations merged into path verbs is of immense importance for the class of path verbs, but not for the class of manner verbs;
3. There is evidence for higher rates of change in the manner verb lexicon for the Germanic and Balto-Slavic subgroups;

4. There is evidence for higher rates of change in the path verb lexicon for the Romance subgroup.

As discussed in §4.2, the variable rates model did not provide a significant improvement over the constant rate model in the phylogenetic rates of change analysis due to small sample size. However, since the results do indicate that manner verbs and path verbs have different rates of change in different parts of the IE tree, we can still identify emerging trends. The findings provide tentative evidence for an affirmative answer to the hypotheses put forward in §1: yes, there is evidence that manner verb evolution proceeds at a quicker pace in the satellite-framed subgroups (Germanic, Balto-Slavic) and yes, there is evidence that path verb evolution proceeds at a quicker pace in one of the verb-framed subgroups (Romance). Although these results are tentative, the phylogenetic comparative analyses that test for variable rates of change could prove to be useful tools for semantic typology.

The difference in the types of etymological origins of manner verbs and path verbs can be attributed to the fact that path verbs in IE are special, because they are often derived from preverb-verb combinations. The types of etymological origins of manner verbs are expected to be more similar to those of other semantic verb classes, whereas those of path verbs clearly are different across languages. The etymological origins are readily unified with the results of the rate of change optimizations for the big IE subgroups. The higher rates of change for manner verbs in Germanic and Balto-Slavic can be related to the high number of manner verbs that have emerged at the level of the Germanic and Balto-Slavic subgroups (see also Dickey 2010, Greenberg 2010). The higher rate of change for path verbs in Romance can be related to the high number of path verbs that have emerged at the level of the Romance subgroup and through inherited preverb-verb combinations.

The inferred rates of change for the individual IE languages are more difficult to match with the etymological patterns. In some cases, ancestral states explain the results. Assamese, for example, has a quite large number of path verbs, higher than any of the other Indo-Iranian languages. Assamese also has the longest optimized branch length of all the Indo-Iranian languages in the phylogenetic tree scaled to optimize the evolution of path verb lexicon size. The high rate of change suggested by this scaled branch reflects the fast expansion of the size of the Assamese path verb lexicon. However, not all of the optimizations for individual languages can be explained in this way. And it is impossible to compare rates of change directly to etymological sources: English has the highest number of manner verbs that originate at the language level, but does not have a very long optimized branch length. The reason for this is that the rate of change analysis does not have information on when exactly a verb appears — information that we do have because

of the etymological analysis. The rate of change analysis only has information on the size of the two motion verb classes, and has to infer from this in which parts of the tree the rate of change was higher and in which parts of the tree it was lower. In the future, it might be possible to give the phylogenetic algorithm etymological information, so that the two types of analyses are truly unified and rates of change for individual languages reflect etymological history more closely.

What do these results tell us about the processes that drive change in manner verb lexicon evolution and path verb lexicon evolution? Although I have focused on phylogenetic models throughout, in order to gain a wider perspective on motion verb lexicon evolution, we need to have an areal perspective. Within the last two decades, a picture of a European north-south divide in motion event encoding patterns has emerged: north and central Europe are satellite-framed (Germanic languages, Balto-Slavic languages (on these, see Slobin 2005), Finno-Ugric languages including Estonian, Finnish and Hungarian, Daghestanian languages including Avar, Lezgian and Udi, and Kartvelian languages including Georgian (on these, see Wälchli 2009:214)) whereas south Europe is verb-framed (Romance languages (Slobin 2005), Basque (Ibarretxe-Antuñano 2004), Albanian (Verkerk 2014c) and Turkish (Slobin 2005)), with several languages 'in between' that have mixed encoding (northern Italian dialects (Slobin 2005), Serbo-Croatian (Filipović 2007, Vidaković 2012) and Modern Greek (Hickmann et al. forthcoming)). The proximity of satellite-framed languages in the north is likely to have contributed to the maintenance of satellite-framed encoding and larger manner verb lexicons in the Germanic and Balto-Slavic languages. Likewise, the proximity of verb-framed languages in the south is likely to have contributed to the maintenance of verb-framed encoding and larger path verb lexicons in the Romance languages. The behavior of the Indo-Iranian languages cannot be explained fully in these terms and will be discussed further below.

This areal perspective allows us to unify the results of the etymological investigation and the optimization of rate of change of manner verb lexicon evolution presented here with the claims made in the motion literature. Slobin (2004) proposed that speakers of satellite-framed languages pay more attention to the encoding of manner of motion information, which results in the larger and larger manner verb lexicons of these languages. This is shown by the higher rates of change of manner verb lexicon evolution for Balto-Slavic and Germanic. In addition, there is some evidence for different etymologies for manner verbs from different tiers: first tier manner verbs are more likely to have an IE history, while third tier verbs are more likely to emerge on the level of individual languages (see Appendix 6). On the whole, however, the proportions of different etymology types for satellite-framed and verb-framed languages are far more similar for manner verbs than they are for path verbs (see Appendices 4 and 5). In some sense, then,

satellite-framed languages simply seem to have more manner verbs than verb-framed languages. The Germanic and especially the Balto-Slavic languages retain far more second tier manner verbs with an IE origin than the Romance languages, implying that manner of motion has been a salient concept throughout the history of these languages. It seems highly likely that their proximity to each other, as well as to other satellite-framed languages such as the Finno-Ugric languages, has helped to maintain satellite-framed encoding patterns and large manner verb lexicons. Evidence for this scenario is provided by Fanego (2012: 45ff), who points out that in Middle English, 51 out of 181 new manner of motion verbs were borrowed from Old Norse, Middle Dutch and Middle Low German, and 37 from French. In the current dataset, Germanic languages show 22 borrowed manner verbs: 8 from other Germanic languages, 12 from French and 2 from Italian. Clearly, Germanic languages borrowed manner of motion verbs from each other, while the cultural importance of France during the Middle Ages has left its mark on motion verb vocabulary as well. Future work should focus on the satellite-framed Finno-Ugric languages, to find out more about the proposed areality of satellite-framed motion encoding in north and central Europe, as well as the etymologies and rates of change of Finno-Ugric manner of motion verbs.

The evolution of path verbs in Indo-European is driven to a large extent by the lexicalization of preverb-verb combinations. The availability of preverb-verb combinations throughout most of the history of all IE languages has created an unparalleled source for path verb derivation. Parallel evolution and calquing of path verbs such as the verbs meaning “to return” listed at the end of §3.2 are therefore entirely predicted by the availability of path preverbs. The emergence of path preverbs and the continued use of their descendants in IE languages is a good example of a lineage-specific trend (Dunn et al. 2011) — the resources available in PIE have shaped contemporary motion event encoding in unique ways.

The development of the path verb lexicon and the increase of the use of the verb-framed constructions in Romance and Indo-Iranian is probably a single, unified development. What are now path verbs such as French *descendre* “to descend” and *revenir* “to return” were once satellite-framed constructions with preverbs *de-* and *re-* and verbs *scandere* “to mount” and *venir* “to come” in Latin. As preverb-verb combinations lexicalized into path verbs, the verb-framed construction became more frequent. The data presented in this paper suggest that as a consequence of this development, path verbs become more frequently used and more salient, leading to an increase in the development of path verbs at the subgroup level. In Romance, this takes place mostly through derivation of non-verbal elements, as in French *entrer*, Italian *entrare*, Portuguese *entrar* and Romanian *intra* “to enter” from Latin *intrare* “to enter”, which is derived from the Latin preposition *inter* “among, between”. In Indo-Iranian, this takes place mostly through semantic

shift, as in Hindi *barhnā* and Gujarati *vādhvū* “to advance” from Sanskrit *vārdhatē* “to grow, increase”. The syntactic changes that resulted in the non-productivity of the spatial prefixes in the emergence of the modern Romance and Indo-Iranian languages seem to have triggered a chain reaction in their development of path verbs. It seems that this chain reaction continues in the Romance languages, which have far more language-internal path verb etymologies than the Germanic and Balto-Slavic languages. It is likely that this process has been influenced by contact with southern European verb-framed languages such as Albanian, Basque, Turkish and Greek, the latter of which is mixed but also has many lexicalized pre-verb-verb derivations (Verkerk 2014a).

However, Indo-Iranian languages do not seem to have continued down the same path of path verb creation as the Romance languages, as Hindi, Nepali and Persian hardly have any etymological language-internal developments. Although an overall correlation between the size of the path verb lexicon and the use of path-only and verb-framed motion event encoding constructions has been found for IE (Verkerk 2014b), this correlation does not seem to be present for Indo-Iranian. If the number of manner verbs and path verbs in Persian, Hindi and Nepali (see Table 1) is considered, it seems clear that these three languages have neither large manner verb lexicons nor large path verb lexicons. Assamese is the only Indo-Iranian language with a large path verb lexicon, many of which are complex verbs. In fact, the divergence of Assamese in having both a low number of manner verbs and a high number of path verbs within the Indo-Iranian subgroup has caused the variable rates optimization to find high rates of change in both the manner and the path verb lexicon size analyses.

So although the Indo-Iranian languages have undergone the same changes with regard to the productivity of the preverb system as the Romance languages, three out of four languages in the current sample do not have an equally large path verb lexicon as the Romance languages. One reason why Hindi, Nepali and Persian might not have path verb lexicons similar in size to Romance is that while they use path verbs to encode motion, they use deictic verbs frequently as well (Verkerk 2014c). It might be the case that the frequent use of deictic verbs in these languages has inhibited the emergence of a large class of path verbs. In addition, the existence of many complex path verbs in Persian and Assamese makes it very difficult to see whether these verbs have emerged at the individual language level or the subgroup level, making comparisons across subgroups difficult. If the variability of motion verb lexicon size in the Indo-Iranian subgroup found in this study is a proxy for the diversity of motion verb lexicon sizes in the rest of the Indo-Iranian languages, clearly these languages are more diverse in their motion event encoding patterns than some of the other IE subgroups (see Verkerk 2014c). More investiga-

tion of motion event encoding in the large subgroup of Indo-Iranian languages is needed.

To conclude, the present study is the first cross-linguistic attempt to discover where motion verbs come from, both from an etymological as well as a phylogenetic perspective. It has demonstrated that these two methods of investigation can be used in a complementary manner, to answer different but related types of questions. Phylogenetic comparative methods can be useful in answering questions regarding the rate of change of linguistic variables. The analysis indicates that manner verbs and path verbs typically have dissimilar types of etymological origins and are gained at different rates of change in different languages due to different motion event encoding patterns. An areal perspective on the divergence of northern satellite-framed branches and southern verb-framed branches in IE is helpful in understanding why certain subgroups became more satellite-framed while others became more verb-framed. In the future, I hope that parallel studies of different language families will be undertaken in order to discover more about the origins of motion verbs as well as rates of change in motion verb lexicon size.

References

- Adams, James Noel. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511482960](https://doi.org/10.1017/CBO9780511482960)
- Algeo, John. 1980. Where do all the new words come from? *American Speech* 55. 264–277. DOI: [10.2307/454567](https://doi.org/10.2307/454567)
- Aske, Jon. 1989. Path predicates in English and Spanish: A closer look. In Kira Hall, Michael Meacham & Richard Shapiro (eds.), *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*. 1–14.
- Beekes, Robert S. P. 2011. *Comparative Indo-European linguistics: An introduction*, 2nd edn. Amsterdam: John Benjamins. DOI: [10.1075/z.172](https://doi.org/10.1075/z.172)
- Berman, Ruth A. & Dan I. Slobin (eds.). 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337. 957–960. DOI: [10.1126/science.1219669](https://doi.org/10.1126/science.1219669)
- Brown, Cecil H. 1983. Where do cardinal direction terms come from? *Anthropological Linguistics* 25(2). 121–161.
- Cannon, Garland. 1978. Statistical etymologies of new words in American English. *Journal of English Linguistics* 12. 12–18. DOI: [10.1177/007542427801200102](https://doi.org/10.1177/007542427801200102)
- Cifuentes F  rez, Paula. 2010. The semantics of the English and the Spanish motion verb lexicons. *Review of Cognitive Linguistics* 8. 233–271. DOI: [10.1075/rcl.8.2.01cif](https://doi.org/10.1075/rcl.8.2.01cif)
- Clackson, James. 2007. *Indo-European linguistics: An introduction*. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511808616](https://doi.org/10.1017/CBO9780511808616)

- Delbrück, Berthold. 1888. *Altindische syntax*. Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.
- Delbrück, Berthold. 1893. *Vergleichende Syntax der indogermanischen Sprachen*. Strasbourg: K.J. Trübner.
- Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.
- Dickey, Stephen M. 2010. Common Slavic “indeterminate” verbs of motion were really manner-of-motion verbs. In Victoria Hasko & Renee Perlmutter (eds.), *New approaches to Slavic verbs of motion*, 67–109. Amsterdam: John Benjamins. DOI: [10.1075/slcs.115.06dic](https://doi.org/10.1075/slcs.115.06dic)
- Dizionario sapere.it. 2012. *Dizionario di Lingua Italiana*. Novara, Italy: De Agostini Editore.
- 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). 1969–1973. DOI: [10.1093/molbev/mss075](https://doi.org/10.1093/molbev/mss075)
- 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](https://doi.org/10.1038/nature09923)
- Dunn, Michael, Kate Bellamy, Chundra Cathcart, Andrew Garrett, Julia Ludewig & Cornelia van Scherpenberg. Forthcoming. IELex: An Indo-European lexical database for phylogenetic analysis. *PLoS One*.
- Dyen, Isodore, Joseph B. Kruskal & Paul Black. 1992. An Indo-European classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82(1). iii–132. DOI: [10.2307/1006517](https://doi.org/10.2307/1006517)
- Fanego, Teresa. 2012. Motion events in English: The emergence and diachrony of manner salience from Old English to Late Modern English. *Folia Linguistica Historica* 33. 29–85.
- Filipović, Luna. 2007. *Talking about motion: A crosslinguistic investigation of lexicalization patterns*. Amsterdam: John Benjamins. DOI: [10.1075/slcs.91](https://doi.org/10.1075/slcs.91)
- Green, Peter J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4). 711–732. DOI: [10.1093/biomet/82.4.711](https://doi.org/10.1093/biomet/82.4.711)
- Greenberg, Marc L. 2010. PIE inheritance and word-formational innovation in Slavic motion verbs in *-i-*. In Victoria Hasko & Renee Perlmutter (eds.), *New approaches to Slavic verbs of motion*, 111–121. Amsterdam: John Benjamins. DOI: [10.1075/slcs.115.07gre](https://doi.org/10.1075/slcs.115.07gre)
- 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](https://doi.org/10.1075/cilt.280)
- Hickmann, Maya, Henriëtte Hendriks, Efstathia Soroli, Tatiana ĭakovleva & Yinglin Ji. Forthcoming. *Space and language typology: Encoding motion across languages*.
- Hofmann, Johann B. & Anton Szantyr. 1965. *Lateinische Syntax und Stilistik*. Munich: C. H. Beck'sche Verlagsbuchhandlung.
- 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](https://doi.org/10.1126/science.1065889)
- Ibarretxe-Antuñano, Iraide. 2004. Motion events in Basque narratives. In Sven Strömquist & Ludo Verhoven (eds.), *Relating events in narrative: Typological and contextual perspectives*, 89–111. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kass, Robert E. & Adrian E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90(430). 773–795. DOI: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572)
- Lehmann, Winfred P. 1974. *Proto-Indo-European syntax*. Austin: University of Texas Press.
- Machado, José Pedro. 1952. *Dicionário etimológico da língua portuguesa*. Lisbon: Livros Horizonte.

- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81(2). 382–420. DOI: [10.1353/lan.2005.0078](https://doi.org/10.1353/lan.2005.0078)
- Nicolas, Christian. 2005. *Sic enim appello...: essai sur l'autonymie terminologique gréco-latine chez Cicéron*. Leuven: Peeters.
- Orel, Vladimir E. 1998. *Albanian etymological dictionary*. Leiden: Brill.
- Özçalışkan, Şeyda & Dan I. Slobin. 2003. Codability effects on the expression of manner of motion in English and Turkish. In Sumru Özsoy, Didar Akar, Mine Nakipoğlu-Demiralp, Eser Erguvanlı-Taylan & Ayhan Aksu-Koç (eds.), *Studies in Turkish linguistics: Proceedings of the Tenth International Conference on Turkish Linguistics*, 259–270. Istanbul: Bogazici University Press.
- Pagel, Mark. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26. 331–348. DOI: [10.1111/j.1463-6409.1997.tb00423.x](https://doi.org/10.1111/j.1463-6409.1997.tb00423.x)
- Pagel, Mark. 1999. Inferring the historical patterns of biological evolution. *Nature* 401. 877–884. DOI: [10.1038/44766](https://doi.org/10.1038/44766)
- Pagel, Mark, Andrew Meade & Daniel Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53(5). 673–684. DOI: [10.1080/10635150490522232](https://doi.org/10.1080/10635150490522232)
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129. DOI: [10.1111/1467-968X.00091](https://doi.org/10.1111/1467-968X.00091)
- Schmitt, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Slobin, Dan I. 1996. Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibitani & Sandra A. Thompson (eds.), *Grammatical constructions: Their form and meaning*, 195–219. Oxford: Clarendon Press.
- Slobin, Dan I. 1997. Mind, code, and text. In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type: Dedicated to T. Givón*, 437–467. Amsterdam: John Benjamins. DOI: [10.1075/z.82.24slo](https://doi.org/10.1075/z.82.24slo)
- Slobin, Dan I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömquist & Ludo Verhoven (eds.), *Relating events in narrative: Typological and contextual perspectives*, 219–257. Mahwah, NJ: Lawrence Erlbaum Associates.
- Slobin, Dan I. 2005. Relating narrative events in translation. In Dorit Diskin Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman*, 115–129. Dordrecht, The Netherlands: Kluwer. DOI: [10.1007/1-4020-7911-7_10](https://doi.org/10.1007/1-4020-7911-7_10)
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. Path to realization: A typology of event conflation. In Christopher Johnson, Laurel A. Sutton & Ruth Shields (eds.), *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 480–519. Berkeley: University of California.
- Turner, Ralph Lilley. 1962–1985. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.
- Uhlenbeck, Christiaan Cornelis. 1899. *Kurzgefasstes etymologisches Wörterbuch der altindischen Sprache*. Amsterdam: Müller.
- Vaan, Michiel de. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.

- Venditti, Chris, Andrew Meade & Mark Pagel. 2011. Multiple routes to mammalian diversity. *Nature* 479. 393–396. DOI: [10.1038/nature10516](https://doi.org/10.1038/nature10516)
- Verkerk, Annemarie. 2013. Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. *Language Dynamics and Change* 3(2): 169–217. Special issue, “Phylogeny and beyond: Quantitative diachronic approaches to language diversity”, edited by Michael Dunn. DOI: [10.1163/22105832-13030202](https://doi.org/10.1163/22105832-13030202)
- Verkerk, Annemarie. 2014a. Diachronic change in motion event encoding. *Journal of Historical Linguistics* 4(1). 40–83. DOI: [10.1075/jhl.4.1.02ver](https://doi.org/10.1075/jhl.4.1.02ver)
- Verkerk, Annemarie. 2014b. The correlation between motion event encoding and path verb lexicon size in Indo-European. *Folia Linguistica Historica* 35. 307–358.
- Verkerk, Annemarie. 2014c. Where Alice fell into: Motion events in a parallel corpus. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*, 324–354. Berlin: Walter de Gruyter.
- Vidaković, Ivana. 2012. He walked up the pole with arms and legs: Typology in second language acquisition. In Luna Filipović & Kasia M. Jaszczolt (eds.), *Space and time in languages and cultures: Linguistic diversity*, 233–262. Amsterdam: John Benjamins. DOI: [10.1075/hcp.36.14vid](https://doi.org/10.1075/hcp.36.14vid)
- Vries, Jan de. 1971. *Nederlands etymologisch woordenboek*. Leiden: Brill.
- Wälchli, Bernhard. 2009. *Motion events in parallel texts: A study in primary-data typology*. Bern: Universität Bern Habilitationsschrift.
- Wartburg, Walther von. 1922–2003. *Französisches etymologisches Wörterbuch*. Bonn: Klopp.
- Witkowski, Stanley R., Cecil H. Brown & Paul K. Chase. 1981. Where do tree terms come from? *Man* 16(1). 1–14. DOI: [10.2307/2801972](https://doi.org/10.2307/2801972)

Résumé

Durant les quatre dernières décennies, de nombreux progrès ont été effectués dans le domaine de l’expression du mouvement (Talmy 1985, 1991, Slobin 1996, 2004). La comparaison entre les langues à satellites et les langues à cadrage verbal montrent que ces premières encodent le mouvement et la manière dans le verbe (nager, partir) tandis que les langues à cadrage verbal encodent le mouvement et la trajectoire dans ce dernier (entrer, traverser) (Slobin 2004, Verkerk 2013, 2014b). Cet article se propose d’étudier les différences qui émergent entre les langues à satellites et les langues à cadrage verbal. Nous utiliserons les méthodes phylogénétiques comparatives (empruntées à la biologie) que nous appliquerons à une étude étymologique afin d’étudier les verbes qui lexicalisent le mouvement et ceux qui lexicalisent la trajectoire : tous ces verbes proviennent d’une base de données indo-européenne. Nous montrerons que ces deux types de verbes ont des étymons bien distincts. En effet, les verbes encodant la manière sont apparus dans les langues à satellites tandis que les verbes encodant la trajectoire sont apparus dans les langues à cadrage verbal.

Zusammenfassung

Über die vergangenen vier Jahrzehnte wurden deutliche Fortschritte in der Beschreibung und Analyse der sprachenübergreifenden Vielfalt im Ausdruck von Bewegung erlangt (Talmy 1985, 1991, Slobin 1996, 2004). Vergleiche zwischen sogenannten *satellite-framed languages* und *verb-framed languages* deuten darauf hin, dass *satellite-framed languages* generell über einen größeren Wortschatz von Verben, die die Art der Bewegung ausdrücken (*swim, dash*; ‚Artverben‘), verfügen, während *verb-framed languages* generell über einen größeren Wortschatz von Verben, die den Weg der Bewegung ausdrücken (*enter, cross*; ‚Wegverben‘), verfügen (Slobin 2004, Verkerk 2013, 2014b). Diese Studie untersucht, wie sich Unterschiede zwischen den Wortschätzen der Bewegungsverben in *satellite-framed* und *verb-framed languages* entwickeln. Um Art- und Wegverbwortschätze in einem indoeuropäischen Datenkorpus zu vergleichen, wurden aus der Biologie stammende phylogenetische komparative Methoden sowie ein etymologischer Ansatz angewendet. Dieser Artikel zeigt, dass Art- und Wegverben generell unterschiedliche etymologische Ursprünge haben, und dass sich Artverben in *satellite-framed languages* und Wegverben in *verb-framed languages* schneller entwickeln.

Author's address

Annemarie Verkerk
Reading Evolutionary Biology Group
School of Biological Sciences
University of Reading
READING
RG6 6AS
United Kingdom
a.verkerk@reading.ac.uk