The correlation between motion event encoding and path verb lexicon size in the Indo-European language family

Annemarie Verkerk
Evolutionary Processes in Language and Culture Group,
Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

There have been opposing views on the possibility of a relationship between motion event encoding and the size of the path verb lexicon. Özçalışkan (2004) has proposed that verb-framed and satellite-framed languages should approximately have the same number of path verbs, whereas a review of some of the literature suggests that verb-framed languages typically have a bigger path verb lexicon than satellite-framed languages. In this article I demonstrate that evidence for this correlation can be found through phylogenetic comparative analysis of parallel corpus data from twenty Indo-European languages.

Keywords: Motion events, path verbs, phylogenetic comparative methods, PGLS, Indo-European

1. Introduction

Since the seminal work of Leonard Talmy (1985, 1991), linguists around the world have asked themselves why some languages often use path of motion verbs such as *enter*, *ascend*, and *pass*, while other languages often use manner of motion verbs such as *run*, *swim*, and *crawl*. Talmy (1985, 1991) formalized this difference when he formulated the terms 'verb-framed' and 'satellite-framed'. A language is verb-framed when it commonly encodes the path of motion on the verb, while a language is satellite-framed when it commonly encodes the path of motion outside the verb, in a directional particle Talmy called 'satellite'. Examples from two languages under study in this article are provided in (1) and (2).

(1) Persian (verb-framed construction)

pādešāh goft-∅   "eškāl-i   na-dār-ad
king   say.PST-3SG   problem-INDF NEG-have-3SG

mi-tavān-i   be-rav-i"   va kolā-be-sar
DUR-can.MOD-2SG NIND-go-2SG.SBJV and hat-to-head

bedun-e   ān=ke   hattā kafš-hā-y-aš rā
without-EZ that=that.REL even shoe-PL-3SG OBJ.DEF

be pā kon-ad   bā 'ajale dād-gāh   rā
to foot do-3SG.SBJV with speed justice-court OBJ.DEF

tark   kard-∅.
leaving do.PST-3SG

'The king said "No problem, you can go," and the Hatter, without even putting his shoes on, **left the court room speedily**.'

(2) Russian (satellite-framed construction)

"Ty   svobod-en"   skaz-a-l   korol'
2SG.SBJ free-2SG.M.NOM say-PFV-PST.3SG.M king.SG.M.NOM

bolvanščik-u.   I   Bolvanščik   vy-bež-a-l
hatter-SG.M.DAT and hatter-SG.M.NOM PFV-run-VF-PST.3SG.M

iz zal-a   sud-a
out hall-SG.M.GEN court-SG.M.GEN

'"You may go," said the king to the Hatter. And the Hatter **ran out of the court room . . .**'

In the Persian example in (1), the path of motion (from inside to outside of the court room) is encoded on the verb *tark kardan* 'to leave', while the manner of motion (speedily, running) is encoded on the adverbial phrase *bā 'ajale* 'speedily'. In contrast, in the Russian example in (2), the path of motion is encoded on the spatial verbal prefix *vy-* 'out', as well as the preposition *iz* 'out', while the manner of motion is encoded on the verb *bežat'* 'to run'. These two types of constructions, the verb-framed construction and the satellite-framed construction, as well as several other different types of motion constructions, have been attested in a range of different languages spoken around the world. Languages typically make use of more than one motion event encoding construction, as has been shown by Beavers et al. (2010), Croft et al. (2010), Slobin (2004), and others. However, linguists usually characterize languages as 'satellite-framed' or 'verb-framed', depending on which strategy is most frequently used in unmarked settings.

Slobin (1997, 2003, 2004, 2006) has proposed that satellite-framed languages typically have a larger class of manner of motion verbs (manner verbs) such as *jog*, *fly*, and *hover*. Verkerk (2013) has demonstrated that there is some support for this hypothesis in a phylogenetic comparative analysis of 20 Indo-European languages. If satellite-framed languages typically have a larger manner verb lexicon as compared with verb-framed languages, it could be the case that verb-framed languages have a larger lexicon of path of motion verbs (path verbs) such as *exit*, *ascend*, and *pass* as compared with satellite-framed languages. On the other hand, however, Özçalışkan (2004: 85) has proposed that both verb-framed and satellite-framed languages should have approximately the same number of path verbs because the path verb lexicon "does not provide many options for elaboration". In a large study of a comparable corpus of metaphorical motion events in English (a satellite-framed language) and Turkish (a verb-framed language), Özçalışkan (2004) finds divergent numbers of manner verbs (95 types in English; 30 types in Turkish), but similar numbers of path verbs (20 types in English; 24 types in Turkish). Özçalışkan (2004) seems to be correct at first glance, as the range of possible purely directional paths of motion is quite small as compared with for instance the range of possible manners of motion: they can relate to an abstract endpoint (*enter*), sourcepoint (*exit*), mid-point (*pass*, *cross*) or to the basic directions up (*ascend*), down (*descend*), forward (*advance*), back (*return*), around (*circle*), behind, or in front.

However, despite this limited set of abstract directions, there are two reasons why the class of path verbs can be large in some languages. The first is that many languages have path verbs that refer to a far more varied set of more or less abstract reference points for motion placed within an environment. Jahai, a Mon-Khmer language spoken in the Malay Peninsula, features verbs that refer to the flow of the water in rivers and their tributaries, distinguishing verbs such as *rkruk* 'to move along the main river (in both upstream and downstream direction)', *piris* 'to move across the flow of water', *dey* 'to move upstream on a tributary' and *hɔc* 'to move downstream on a tributary' (Levinson & Burenhult 2009). A similar system exists for other geographical features such as mountain ridges, which "distinguish 'motion lengthwise on mountain ridge' vs. 'motion across mountain ridge', as well as 'motion lengthwise on mountain side' vs. 'motion up on mountain side'/'motion down on mountain side' "(Levinson & Burenhult 2009: 161).

Second, languages with sizable path verb lexicons may also feature a range of verbs that have highly similar but still distinct semantics. Cifuentes Férez (2010) lists 14 Spanish path verbs that relate movement away from a sourcepoint, out of which seven are highly similar (original glosses given): *apartarse*, *distanciarse*, *ladearse* 'to move away from'; *largarse*, *marcharse*, *partir* 'to leave'; *pirarse* 'to go away (informal)'. However, these verbs are not simply synonyms as they are used in different types of context, providing a degree of semantic granularity that is not captured by English *go away* and *leave*. It is clear from this that languages can have big path verb lexicons even by encoding just the most basic set of directional paths, which goes against Özçalışkan's (2004) position.

A review of the literature also suggests that there is support for the hypothesis that verb-framed languages have larger path verb lexicons than satellite-framed languages. An overview of path verb lexicon sizes reported in the literature has been provided in Tables 1 and 2. Table 1 provides an overview of different types of studies (experimental, corpus-based, or dictionary-based) that provide information on path verb lexicon size for English and German, both satellite-framed languages, in comparison with different verb-framed languages. Table 2 presents path verb lexicon size data from a set of different languages. Note that these numbers are not comparable across languages, as they have not used the same methodology.

Table 1 gives an overview of the path verb lexicon sizes in studies that compared two languages, in this case mostly English, being a satellite-framed language, with a verb-framed language. The range of reported path verb vocabulary sizes for English is very large since this table compiles experimental studies in which the number of stimuli was very limited

Table 1. Comparisons of path verb lexicon size in satellite-framed and verb-framed languages from the motion event encoding literature

Satellite-framed language	No. of path verbs	Verb-framed language	No. of path verbs	Source
English	14	Basque	37	Ibarretxe-Antuñano (2004)
English	24	Hindi	14	Narasimhan (2003)
English	4	Spanish	8	Naigles et al. (1998)
English	44	Spanish	65	Cifuentes Férez (2010)
English	20	Turkish	24	Özçalışkan (2004)
English	7	Turkish	13	Özçalışkan (2009)
German	36	French	37	Berthele (2006)

Table 2. Reported sizes of path verb lexicons for individual languages in the motion event encoding literature

Language	No. of path verbs	Source
Chinese (contested)	13	Chen and Guo (2009)
German (satellite-framed)	'virtually none'	Wienold (1995)
Indonesian (verb-framed)	15	Wienold (1995)
Japanese (verb-framed)	33	Matsumoto (1997) as cited in Matsumoto (2003)
Japanese (verb-framed)	36	Wienold (1995)
Korean (verb-framed)	35	Wienold (1995)
Sidaama (verb-framed)	13	Kawachi (2011)
Thai (equipollently-framed)	19	Wienold (1995)

(Naigles et al. 1998), with dictionary studies that list a far more comprehensive range of path verbs (Narasimhan 2003; Cifuentes Férez 2010). The reported path verb lexicon sizes are therefore largely dependent on the methodology that is employed. Table 1 shows that although English has a large class of path verbs based on a dictionary consultation, as done by Narasimhan (2003) and Cifuentes Férez (2010), in other types of studies English has a smaller class of path verbs as the verb-framed languages to which it is compared. Table 2 gives an overview of some reported path verb lexicons for several individual languages. In Table 2 German, the only satellite-framed language for which data on path verb lexicon size could be found, was reported to have the smallest path verb class by Wienold (1995), although Berthele (2006) in Table 1 has a very different opinion.

The differences between satellite-framed and verb-framed languages as presented in Table 1 and 2 seems striking at first, but Matsumoto (2003) writes that the size of the path verb and manner verb lexicon is at least to some extent independent from the strategy that is typically used to encode motion events. For instance, Table 2 shows that Japanese has around 33–36 path verbs, while Sidaama has 13, even though both are verb-framed (see Table 2). It is therefore likely that there are many different factors that determine how many path verbs a language has.

This article investigates whether a correlation exists between the common use of the verb-framed motion event encoding construction and a larger lexicon of path verbs. A dataset based on a sample of 20 Indo-European languages was used. This dataset is discussed in Section 2. To investigate the correlation, phylogenetic comparative methods adopted from biology were used. These methods are discussed in Section 3. The

results are presented in Section 4. Although the presentation of the results from phylogenetic comparative analysis is the main focus of this article, these results are interpreted in light of a much broader perspective, including morphosyntactic change and the role of contact. This wider discussion of results and a general conclusion are provided in Sections 5 and 6.

2. Dataset

2.1. The parallel corpus

The dataset that was used in this study comes from a parallel corpus. A parallel corpus is a collection of translations of a single text (see Cysouw & Wälchli 2007). The use of a parallel corpus ensures that the data from the different languages in the corpus are translational equivalents, and thus comparable to at least a certain extent. The texts included in the parallel corpus were *Alice's Adventures in Wonderland* (Lewis Carroll), *Through the Looking-Glass and What Alice Found There* (Lewis Carroll), and *O alquimista* (Paulo Coelho). The languages that were included in the corpus are (in the subgroup order in which they appear in Figure 3, below): French, Italian, Portuguese, Romanian [Romance], Irish [Celtic], Dutch, English, German, Swedish [Germanic], Latvian, Lithuanian, Polish, Russian, Serbo-Croatian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Aryan], Albanian, Modern Greek, and Armenian. These languages were chosen as a representative sample of Indo-European languages. Except the two *Alice* novels, all other books were published between 1978 and 2010, making the parallel corpus roughly contemporary.

From the parallel corpus, a selection of motion event descriptions was made that served as the core dataset of this study. The selected motion events describe "situations in which an animate [or inanimate] being moves from one place to another" (Özçaliskan & Slobin 2003: 259). Each selected motion event was a single sentence in which a single situation is described, following Berman & Slobin (1994: 657). These selected motion events were found in the various translations and glossed with the help of native speakers or language experts. The classification of all motion verbs as a path verb, manner verb, etc. was done with the assistance of native speakers.

The final dataset with the selected motion events consisted of 215 different original sentences taken from all three books. 23 of these original

sentences only encode manner, not path, and are therefore excluded from the current analysis. This leaves a total of 192 sentences for the full sample, which is called '192-sentence sample'. This sample is available for 16 of the 20 included languages. Unfortunately, there is currently no translation of *Through the Looking-Glass and what Alice found there* available for four languages: Albanian, Hindi, Nepali, and Persian. A smaller sample of 118 sentences, which is called the '118-sentence sample' was created using only selected motion events from *Alice's Adventures in Wonderland* and *O alquimista*. This sample is available for all 20 languages.

The use of a parallel corpus has many benefits, the most important one being how it enables one to draw inferences across a set of translational equivalents. However, there are drawbacks too, most importantly the risk of translation effects from the source text on the target text and the use of potentially marked speech styles by individual translators. Nevertheless, these potential problems have been restricted by using three source texts from two different languages (English and Portuguese), which reduces the chance of translation and translator effects. Of course, a wider sample of novels from more source languages would have made this chance even smaller, but it is not an easy task to collect these novels in a diverse set of Indo-European languages, let alone analyze them. In addition, my personal observation of motion event encoding in novels, both based on the current materials as well as on materials discussed elsewhere, is that motion is a domain that typically does not suffer from translation effects very much – as far as I have been able to judge by informal discussions with the native speakers and language experts, the translation of motion typically matches native patterns. Table A1 in Appendix 1 provides a classification of languages based on a review of the Talmian literature, which can be compared to the overview of motion event encoding systems from the current sample presented in Figure 1, below. From this comparison it becomes evident that motion event encoding in the current sample matches well with the characterizations made in the literature for those languages that have been studied in the Talmian literature.

2.2. Coding

2.2.1. *Motion event encoding constructions*

The complete dataset was coded for relevant features of motion event encoding. This resulted in the following set of motion event encoding

constructions (see Verkerk 2014 for examples and more discussion). Further elaboration on these constructions follows:

(3)   List of motion event encoding constructions

1. satellite-framed construction: manner verb + path satellite
2. verb-framed construction: path verb + manner expression
3. path verb-only construction: path verb (no indication of manner)
4. manner verb-only construction: manner verb (no indication of path)
5. manner-plus-path verb construction: manner-plus-path verb
6. deictic verb-only construction: deictic verb (no indication of manner)
7. deictic verb-framed construction: deictic verb + manner expression
8. subordinate construction: any two motion verbs, one is subordinate
9. coordinate construction: any two motion verbs, coordinated

The satellite-framed and verb-framed construction have already been discussed above. They make use of the two most common verb classes: manner verbs that indicate the manner of motion, such as *fly* and *walk*, and path verbs that indicate the path of motion, such as *enter* and *descend*. The satellite-framed construction combines a manner verb with a path satellite. A path satellite is any non-verbal element that encodes path, which includes adpositions such as *from*, adverbs such as *back*, as well as case markers, verbal prefixes, and other path encoding entities (see Filipović 2007: 35; Beavers et al. 2010: 337; and Croft et al. 2010: 205–206). The verb-framed construction combines a path verb with a manner expression. A manner expression is any adverb, manner verb participle, or adverbial expression that signifies manner. This includes the adverbial phrase *bā 'ajale* 'speedily' in example (1), adverbs such as *slowly* and *hurriedly*, as well as manner verb participles such as *running* and *crawling*.

The path verb-only and the manner verb-only constructions are constructions in which only a path verb or only a manner verb is used, respectively. The manner-plus-path verb construction employs only a manner-plus-path verb. These are verbs that encode both path and manner. Examples of such verbs include Greek *skarfalono* 'climb up' and Persian *goriḵtan* 'run away'.

Deictic verbs are used in the deictic verb-only construction, in which only a deictic verb is used, and the deictic verb-framed construction, in which a deictic verb and a manner expression is used. Deictic verbs indicate the path of motion as seen from a deictic center, such as *go* and *come*. They are not included in the category of path verbs because of their dif-

ferent semantics and because they are often the most commonly used motion verbs. Including them in the category of path verbs would result in a skewed picture of path verb use.

The subordinate construction and the coordinate construction both make use of two motion verbs. These can be any type of motion verb. The coordinate construction coordinates these two verbs, as in 'Alice ran and entered the forest'. The subordinate construction subordinates one of the two verbs, as in 'Alice ran (in order) to enter the forest'. An overview of the usage of these constructions in the current dataset is discussed in Section 2.3.1.

2.2.2. *Path verbs*

Although the corpus was annotated for a range of motion features, which resulted in the set of constructions discussed in Section 2.2.1, the coding of path verbs deserves more extensive discussion here. The classification of a motion verb as a path verb was done as follows. The semantics of each verb were discussed with a native speaker. In most cases, if a verb encoded the path of motion, this was often clear from the start of the discussion as it was key to the meaning of the verb. If a verb encoded both path and manner it was coded as a manner-plus-path verb (see below). If a verb encoded path and could be used in various different manner contexts, it was coded as a path verb. For instance, the English path verb *enter* can be used both in combination with the adverb *speedily* as well as with the manner verb participle *crawling* and is therefore coded as a path verb.

Verbs that encode both path and manner are not included in the category of path verbs (as described in Section 2.2.1). Examples of manner-plus-path verbs are Greek *koytroyvalo* 'tumble down', Lithuanian *kopti* 'climb up', and Persian *goriḵtan* 'run away'. In the motion event encoding literature, these verbs are sometimes included in the manner verb class, and sometimes in the path verb class. Because this class of verbs is semantically different from both the class of path verbs and the class of manner verbs, as they encode both manner and path, they are not included in the path verb class in the current study.

Several path verbs in the current dataset have originated from merged path prefixes + verbs. This is due to the ancient Indo-European system of spatial verbal prefixes or 'preverbs' that has become unproductive in certain branches of Indo-European, such as Romance and Indo-Aryan, while still being productive in others, such as Balto-Slavic (Watkins 1964;

Kuryłowic 1964; Kopecka 2006, 2013). This type of path verb may be inherited from an older stage of the language, such as French *descendre* 'descend' from Latin *descendere* 'descend', which is a combination of the productive Latin prefix *de-* + verb *scandere*. Or they may be more recent combinations, such as French *retourner* 'return' from French *re-* + *tourner*. In some cases, the original prefix and verb may not have merged in the same way as French *descendre* and *retourner*, note for instance the separable verbs in Dutch (such as *terugkeren* 'return') and German (*zurückkehren* 'return'). Especially for the Slavic languages, many of which still have a productive system of spatial prefixes on verbs, it was sometimes difficult to decide when a prefix + verb combination should be coded as a path verb or not. The criterion used here was that if the verb was a motion verb without the prefix, this was a productive use of that motion verb with an added path satellite in the form of a prefix. However, if the verb without the prefix was not a motion verb, i.e. if the meaning of the path verb could no longer be derived from the meaning of the prefix plus the meaning of the verb, the prefix + verb combination as a whole was classified as a path verb. An example of such a path verb is Polish *przybyć* 'arrive', a derivation with the spatial prefix *przy-* with the verb *być* 'be'. In the discussion (Section 5) I will return to this process, which is of vast importance if the correlation between path verb lexicon size and the use of motion event encoding constructions is to be understood.

In this study, the number of path verbs that are encountered in the 118-sentence sample and the 192-sentence sample for each language are used as a measure of path lexicon size. Since these samples include a limited amount of different paths of motion, this number is not comprehensive. If larger corpora or dictionaries would have been used, a more complete list of path verbs for each language could have been established. However, the current approach has the advantage that only path verbs that are still in contemporary use are included. Of course, the original *Alice's Adventures in Wonderland* and *Through the Looking-Glass and What Alice Found There* were published in 1865 and 1872, although aside from an occasionally distinct choice of words, they can be considered almost contemporary. However, the translations have been published between 1978 and 2010, and represent truly contemporary literary language use. A study based on dictionary entries would not necessarily reflect the contemporary path verb lexicon as it might include verbs that are obsolete. In addition, the 118-sentence sample and the 192-sentence sample were constructed such that they

include at least one instance of each path verb that was encountered in the original texts. Therefore, the most commonly found types of path of motion and thus the most commonly used path verbs are included in the two samples.

2.3. Datasets used for phylogenetic comparative analysis

2.3.1. *Motion event encoding construction measures*

This section discusses the measures that were used for the characterization of languages with regard to the motion event encoding constructions they typically use. An overview of the usage of the motion event encoding constructions in the sample is provided first. Figure 1 (overleaf) gives an overview of the proportions of use of each of the motion event encoding constructions for each language as attested in the 118-sentence sample.

The proportions of use of the different motion event encoding constructions depicted in Figure 1 is not discussed here extensively, because they have been discussed more fully elsewhere (Verkerk 2014). However, it is clear from Figure 1 that the 20 languages all use a range of motion event encoding constructions. The differences between the languages lie in construction usage: each language uses some constructions more often and uses others less often.

In Figure 1, the languages have been ordered in such a way that two clear groups of languages emerge: a 'satellite-framed' group of languages that uses the satellite-framed construction often (Russian, Polish, Lithuanian, Latvian, Swedish, German, Dutch, and English) and a 'verb-framed' group of languages that uses the path verb-only construction and the verb-framed construction often (Italian, Portuguese, Romanian, French, Albanian, and Greek). The remaining six languages (Serbo-Croatian, Irish, Armenian, Hindi, Persian, and Nepali) do not clearly fit the 'satellite-framed' or the 'verb-framed' group and seem to be somewhere in between.

To discover further on which dimensions the 20 languages are differentiated, a principal components analysis (PCA) was conducted. PCA is a data reduction technique that converts a set of observations of possibly correlated variables (such as usage of different motion event encoding constructions) into values of a set of uncorrelated variables, the so-called 'principal components'. Since the use of the different motion event encoding constructions is highly correlated (i.e. if a language uses the satellite-framed
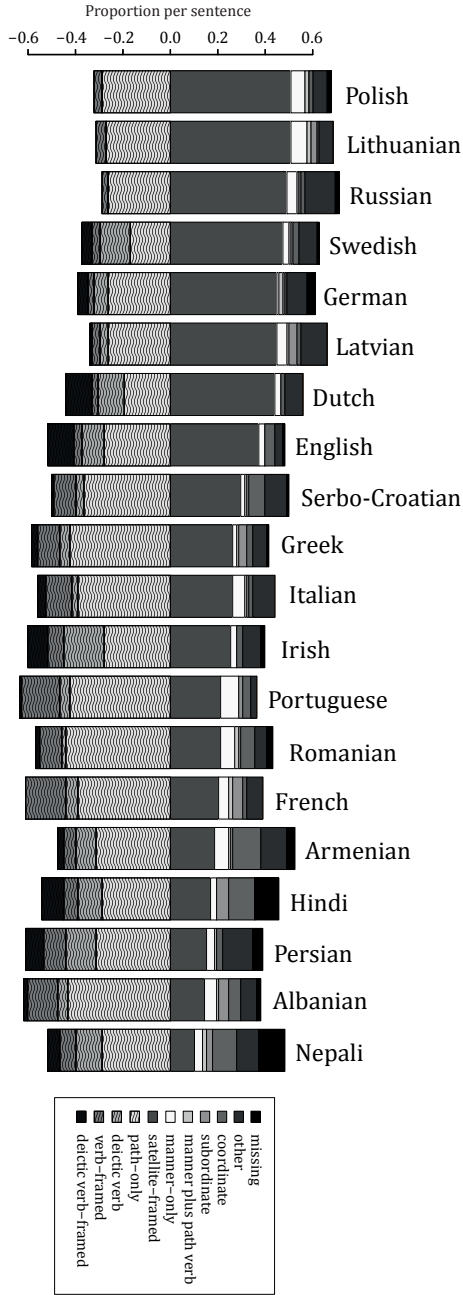
Figure 1. Proportions of the use of different motion event encoding constructions in the 118-sentence sample. Constructions that (may) employ a manner verb as well as unclassified translations ('other') and missing translations are in the positive dimension; whereas constructions that employ a path verb or a deictic verb are in the negative dimension
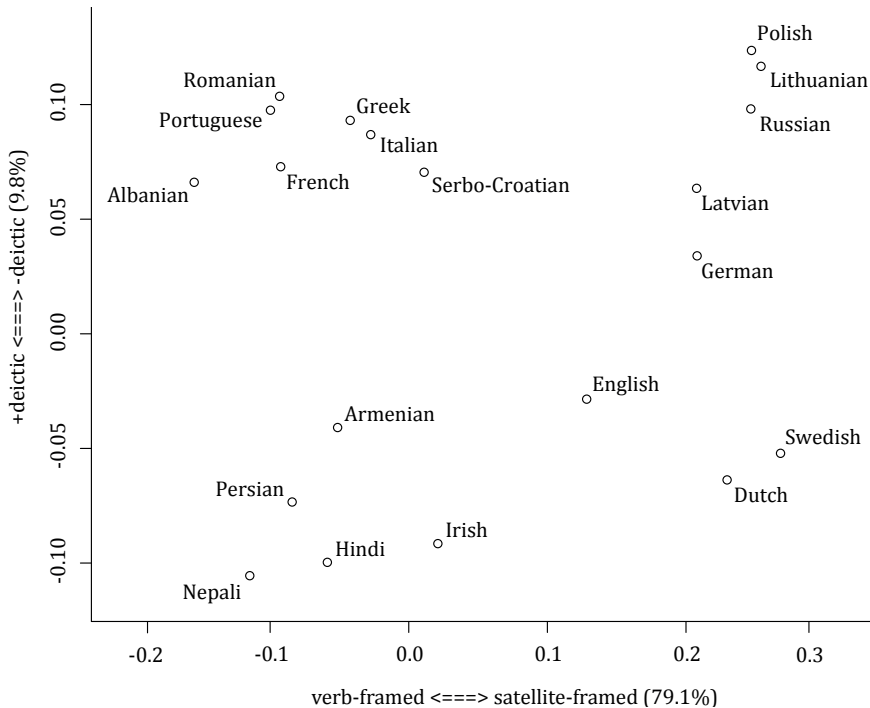
Figure 2. A phylogenetic principal-components analysis conducted on the percentage of usage of each motion event encoding construction in the 118-sentence sample for 20 Indo-European languages

construction often, it is likely not to use the path verb-only construction very often), PCA is a useful data aggregation technique to reduce the dimensionality present in the two sentence samples.

The PCA was conducted on the proportion of usage of each of the motion event encoding strategies. The phylogenetic principal components analysis developed by Revell (2009) was used in order to remove some of the variance that can be attributed to the genetic relationships between the languages during the PCA.[1] The results of this analysis on the 118-sentence sample are depicted in Figure 2.

---

[1] Note that only a small portion of the variance attributed to phylogenetic relationships is removed by Revell's (2009) phylogenetic principal components analysis, and the data still need to be further analyzed with phylogenetic methods: "phylogenetic size-correction and principal components provide estimates of the allometric coefficient and eigenstructure that will have lower variance relative to nonphylogenetic procedures, thus reducing type I error to its nominal level when residuals and scores are subsequently analyzed using phylogenetic methods" (Revell 2009: 3259).

The results of the PCA indicate that the data can be represented very adequately on two dimensions: the first and second principal component are by far the most important and together, they account for a large proportion of the variance (88.9%). The first principal component (PC1) given on the x-axis accounts for 79.1% of the variance. PC1 can be interpreted to relate to the Talmian scale: languages with a high positive score on this component (situated in the right of Figure 2) are satellite-framed, while languages with a high negative score (situated in the left of Figure 2) are verb-framed. The second principal component (PC2) given on the y-axis accounts for 9.8% of the variance. It can be interpreted to relate to the amount of use of the deictic construction and the deictic verb-framed construction. Languages that use these constructions relatively often have a negative score on this component and are situated in the lower half of Figure 2.

Figure 1 and 2 illustrate the need to classify languages on a scale of motion event encoding construction usage, rather than within a dichotomy, as has also been suggested by Beavers et al. (2010), Croft et al. (2010), and Slobin (2004). Differentiation within the 'satellite-framed' group is mostly due to the diverse use of the deictic construction and the deictic verb-framed construction. English, Dutch, and Swedish use these constructions relatively frequently and they are situated on the lower half of the plot in Figure 2, whereas Russian, Polish and Lithuanian use these relatively infrequently and are situated on the upper half of the plot, with German and Latvian situated in between. The 'verb-framed' group (the Romance languages, Albanian, and Greek) form a relatively well-defined group in the left upper part of Figure 2. Serbo-Croatian is in between the 'verb-framed' group and the 'satellite-framed' group, reflecting its 'mixed' classification (see Filipović 2007; Vidaković 2012). That leaves Armenian, Irish, Hindi, Nepali and Persian, all situated in the left bottom part of Figure 2. Despite appearances, these do not constitute a unified group. Irish, for instance, uses the satellite-framed construction far more often than any of the others. Additional distance-based analyses by Verkerk (2014) have demonstrated that Armenian, Hindi, Nepali and Persian are not actually similar, rather they are driven together here because of their relative disuse of both the satellite-framed and the path verb-only constructions and relatively frequent use of the deictic verb-only strategy. Diversity in motion event encoding (as presented in Figure 1 and 2) is found within classes, across classes, and in-between classes, and illustrates the need for a scalar or continuous measure of motion event encoding.

In order to score each language with respect to their usage of the different motion event encoding constructions, two scalar measures were used. The first measure was simply the proportion of use of the path verb-only construction plus the proportion of use of the verb-framed construction. This measure is a continuous scale that indicates how often a construction with a path verb as the main verb of the sentence is used by each of the languages in the sample. This measure allows me to directly investigate the hypothesis put forward: that 'more verb-framed' languages have a larger path verb lexicon. The second measure are the scores of languages on the first principal component, which seems to relate to the Talmian typology (see again Figure 2). These scores place each language on a Talmian scale, going from maximally verb-framed on the left side of Figure 2, to maximally satellite-framed on the right side of Figure 2. This measure is a more comprehensive characterization of motion event encoding in the 20 languages. Using both these measures rather than just one of them will provide a better perspective on the relationship between motion event encoding and the path verb lexicon. Both these measures were generated for both the 118-sentence sample and the 192-sentence sample. They are used as input for the phylogenetic comparative analyses described in Section 3.

2.3.2. *Path verb lexicon size*

This section discusses the size of the path verb lexicons as they were found in the current data set. The number of path verbs was measured by taking the unique verb types attested for each language in each sentence sample. A full overview of the path verbs encountered in the sample is presented in Table A2 in Appendix 2. The number of unique path verbs is listed in Table 3. Both Table A2 in Appendix 2 as well as Table 3 make clear that there is quite a bit of diversity with regard to the number of path verbs: for the 192-sentence sample, Swedish has the lowest number of path verbs (13), while French has the largest number (33). Some verb-framed languages have quite large path verb lexicons, while satellite-framed languages typically have small path verb lexicons. In addition, languages that are closely related, such as the Romance languages French, Portuguese, Italian and Romanian, have a similar number of path verbs: between 28 and 33 for the 192-sentence sample.

Table A2 in Appendix 2 suggests that the verb-framed languages in the sample, most importantly French, Portuguese, Italian, Romanian, and

Table 3. Number of unique path verb types encountered in the 118-sentence sample and the 192-sentence sample

Language	118-sentence sample	192-sentence sample
French	25	33
Italian	25	30
Portuguese	30	32
Romanian	27	28
Irish	11	14
Dutch	12	17
English	14	14
German	14	18
Swedish	9	13
Latvian	16	20
Lithuanian	15	17
Polish	19	21
Russian	14	15
Serbo-Croatian	27	29
Hindi	12	–
Nepali	14	–
Persian	17	–
Albanian	25	–
M. Greek	22	22
Armenian	17	18

Albanian have the largest number of path verbs. These languages often have more than one verb for a single English verb, see for instance the six French verbs that have been glossed with 'return': *repartir, revenir, retourner, rentrer, rejoindre,* and *regagner.* Some types of paths seem to be encoded by a verb in almost all or all 20 languages, such as movement away from a location (*leave*), movement arriving at a location (*arrive*), unsupported movement downwards (*fall*), and movement back to a location previously abandoned (*return*). Other types of paths are encoded by some but not by others. Some types of paths are only encoded by a verb in verb-framed languages, such as: movement into an enclosure (*enter*), movement out of an enclosure (*exit*) and general movement downwards (*descend*). It might of course be the case that a larger sample or different methodology would find that those languages lacking a verb for certain types of path actually do encode these paths with a verb, which just happened not to be used in the current samples. However, the methodology described in Section 2.2.2

Table 4. Summary of the samples and measures used in the current article

Sample	Motion event encoding construction measure	Path verb lexicon size measure
118-sentence	PC1 score (PC1)	unique path verb count (PV count)
118-sentence	proportion path verb-only + verb-framed (% PVO+VF)	unique path verb count (PV count)
192-sentence	PC1 score (PC1)	unique path verb count (PV count)
192-sentence	proportion path verb-only + verb-framed (% PVO+VF)	unique path verb count (PV count)

ensures that a comparable set of the most common path verbs are featured in the current sample.

2.4. Summary

Table 4 presents a short overview of the measures used for the phylogenetic comparative analyses discussed in Section 3. The abbreviations with which each measure is designated are given in parentheses.

3. Methodology

This article investigates the correlation between the motion event encoding system and the size of the path verb lexicon. Specifically it tests whether languages that use the path verb-only and verb-framed motion event encoding constructions more commonly also have a larger path verb lexicon. This hypothesis has to be investigated using phylogenetic comparative methods, because these methods take into account the genetic relationships between the languages in the sample. If standard statistic analyses were used, which do not account for the historical dependencies that might be present in the data, the possibility exists that a correlation is incorrectly estimated. Such a false positive could be due to closely related languages behaving similarly, which could give the suggestion that two variables are correlated. Phylogenetic comparative methods are needed to identify independent instances of a correlation between two or more features.

In order to conduct these phylogenetic comparative analyses, a meas-

ure of genetic relations in the form of a set of phylogenetic trees is needed. These phylogenetic trees are discussed in 3.1. The phylogenetic comparative analyses themselves are introduced in 3.2.

3.1. Phylogenetic trees

The phylogenetic trees used in the current article were taken from Bouckaert et al. (2012). A sample of trees rather than a single tree was used because a tree sample takes into account some measure of uncertainty. There is always a certain measure of uncertainty present in every phylogenetic estimation because different linguistic features have different histories. Bouckaert et al.'s (2012) set of phylogenetic trees was estimated using cognate-coded lexical data from 103 Indo-European languages. The possibility of cognates to be loan words rather than inherited material was investigated and taken into account. The cognate data was recoded in a binary fashion, so that each language was characterized to have a cognate (1) or not (0) for each of the in total 5,047 cognate sets. Using this dataset, a posterior phylogenetic tree distribution was estimated using a Bayesian Markov Chain Monte Carlo approach (Huelsenbeck et al. 2001) that is available in the software BEAST (Drummond et al. 2012). Random sampling from the posterior tree distribution resulted in a sample of 12,500 estimated trees with high likelihoods. From this 12,500 tree sample, I made a random selection of 1,000 trees to use in the current study in order to make the tree sample more manageable.

The trees in the tree sample were pruned to include only those languages the current study presents data on. This results in a sample of 1,000 trees for 20 languages to use with the 118-sentence sample and a sample of the same 1,000 trees for 16 languages to use with the 192-sentence sample. To illustrate what the trees in these tree samples look like, a maximum clade credibility tree of the sample of 1,000 trees for 20 languages was calculated using TreeAnnotator v.1.6.1 (Drummond et al. 2012). This tree is presented in Figure 3. Note, however, that the phylogenetic comparative analyses were not conducted using the tree in Figure 3, rather they were conducted using all trees in both samples of 1,000 phylogenetic trees.

The numbers that are placed with each node on the tree in Figure 3 are so-called support values that indicate how often each subgroup is attested in the sample of 1,000 trees. For instance, the support value given for the subgroup consisting of Albanian, Modern Greek, and Armenian is 0.38. This
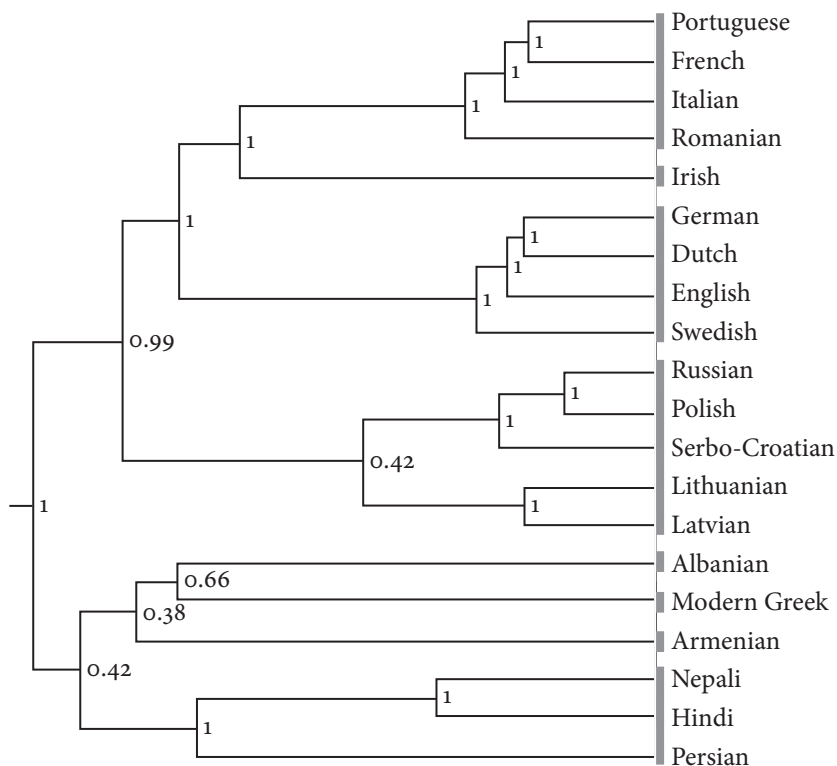
Figure 3. The maximum clade credibility tree of 1,000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 20 languages featured in this article

means that this subgroup is present in 380 out of the total of 1,000 trees. The length of the branches is in proportion to chronological time. As this tree contains only contemporary languages, all branches leading from the root to the tips of the tree have equal length, making this an ultrametric tree.

3.2.   Phylogenetic comparative analyses

3.2.1. *Phylogenetic signal*

For a dataset that includes several (closely and less closely) related languages from a single language family, it is necessary to test whether a phylogenetic

signal is present. If a phylogenetic signal is present within a linguistic feature, this means that languages that are closely related behave similarly with regard to that feature, because they have inherited this feature from a common ancestor. When this is the case, it becomes necessary to use phylogenetic comparative methods to analyze the feature, because these methods can take into account the genetic relationships between the languages. If only conventional methods would be used, a correlation might be inferred that is actually caused by closely related languages that behave similarly, rather than by finding independent instances of co-variance between the features.

In the current study, the estimation of the parameter λ (lambda) was used to test whether a phylogenetic signal was present (Pagel 1999; Freckleton et al. 2002). λ is a scaling parameter that corrects tree branch lengths depending on the amount of shared history between the languages (Pagel 1999). λ assumes an underlying Brownian random-walk model. This means that the data is supposed to evolve in each instance of evolutionary time with a mean change of zero and a constant variance. Evolutionary time is represented by the branch lengths of the tree, which are representative of chronological time in the current set of trees (see Section 3.1). In addition, the Brownian random-walk model of evolution implies that the process of change unfolds independently from previous sections of the tree, implying that the method cannot 'look back' at earlier states of the feature as modeled on the tree (Pagel 1999: 878).

In order to test for the presence of phylogenetic signal, the most likely value of λ was estimated using a maximum likelihood approach, which is done by adjusting the phylogenetic tree by trying out different values for λ until the most likely value is found. Values for λ normally range between 0 and 1, but can also be slightly bigger than 1, depending on certain features of the phylogenetic tree (Freckleton et al. 2002: 715). λ values that are lower than 1 (or the maximal possible λ based on the phylogenetic tree) generate trees with shorter internal branches, internal branches being the branches that connect the ancestral languages inside the tree structure. If λ is set to 0, the internal structure of the tree is completely eliminated. Figure 4 illustrates the effect of the use of λ as a scaling parameter on a phylogenetic tree.

The maximum likelihood estimation process is used to estimate the λ that best transforms the phylogenetic tree given the amount of shared history present in the comparative dataset. An estimated λ value of 1 (or the maximum value of lambda given the phylogenetic tree) indicates that
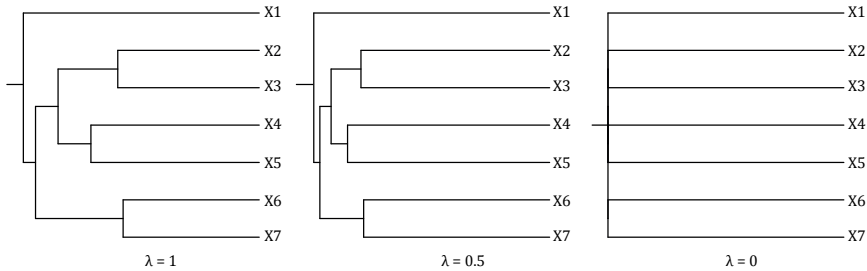
Figure 4. An illustration of the effect of different values of λ on the internal branch lengths of a phylogenetic tree. As λ decreases (from left to right), the internal branches become shorter and disappear completely when λ = 0

the data evolves exactly on the branches of the phylogenetic tree given a random-walk model of evolution (Pagel 1999). An estimated λ value of 0, which completely obliterates any tree structure, implies that the data changes completely independently from the phylogeny and thus that the data is not patterned according to the shared history between the languages. An estimated λ value between 0 and 1 implies that the data is partially dependent on history, but does not evolve exactly along the branches of the given phylogenetic tree. Estimating λ thus gives one an idea to what extent the data is dependent on shared history and indicates whether phylogenetic comparative methods are necessary or not.

It can be tested whether the λ estimate differs significantly from the λ in different models of evolution with the likelihood ratio test (Pagel 1997: 334). Likelihood ratio tests can be used to evaluate whether the λ estimate is significantly different from a model of evolution in which λ is set to 1 and another model of evolution in which λ is set to 0. The likelihood ratio (LR) statistic is calculated as follows:

For testing whether λ is significantly different from 0:
LR = 2[log-likelihood(**estimated λ model**) – log-likelihood(**λ = 0 model**)]

For testing whether λ is significantly different from 1:
LR = 2[log-likelihood(**estimated λ model**) – log-likelihood(**λ = 1 model**)]

The distribution of the LR statistic approximates a χ^2 distribution with 1 degree of freedom. The statistical significance of the difference in likelihood of the two models assessed by the LR statistic can then be determined by looking at the critical significance level of a χ^2 distribution with 1 degree

of freedom. If we choose a significance level of $p < 0.05$, the LR statistic must be 3.841 or higher. If the estimated λ is not significantly different from a model in which λ is set to 1, while it is significantly different from a model in which λ is set to 0, a phylogenetic signal can be said to be present.

The estimation of λ and the likelihood ratio tests were conducted using the function phylosig, part of the R (R Development Core Team, 2011) package *phytools* (Revell 2012) and the function fitContinuous, part of the R package *GEIGER* (Harmon et al. 2008). The function phylosig was used to optimize λ, while the function fitContinuous was used to set λ to 0 and 1 and test for statistical significance as explained above.

3.2.2. *Phylogenetic Generalized Least Squares*

Phylogenetic Generalized Least Squares regression analyses (PGLS) were carried out in order to test whether languages that use the path verb-only and verb-framed constructions more often have larger path verb lexicons (Pagel 1997; Freckleton et al. 2002). PGLS regression is a linear regression technique that takes into account phylogenetic information provided by a sample of phylogenetic trees (Pagel 1997: 337–338). This type of analysis is needed because the data points in the datasets on motion event encoding constructions and path verb lexicon size are not independent: languages that are closely related behave similarly because they have inherited features from a shared ancestor. PGLS analysis takes this shared history into account by using the shared branch lengths of the phylogenetic tree. It accomplishes this by adjusting the β in the following regression formula:

$$Y = a + \beta X + e$$

Here, a variable Y is dependent on a variable X, a is the intercept of the regression line, β is the slope of the regression of the Y variable on the X variable, and e is the measure of error. Specifically, β specifies the amount of change in Y given change in X, that is, how Y changed as X evolved over time (Pagel 1997: 340). In the current analysis, the dependent variable Y is the path verb lexicon size measure, while the independent variable X is one of the two motion event encoding construction measures.

Since the languages are related, the slope of the regression β needs to be adjusted to disregard co-variance caused by shared descent. During PGLS regression analysis, the shared history of each language with each other language is formally assessed by the length of the branches that are shared
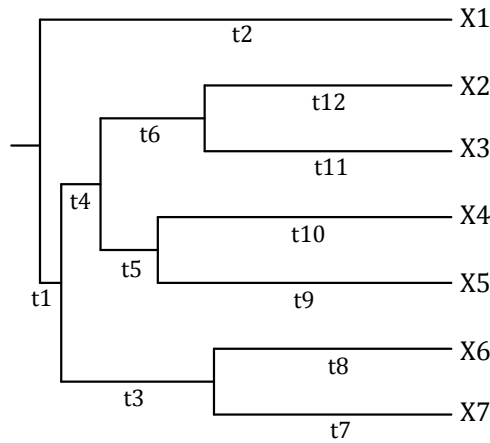
	X1	X2	X3	X4	X5	X6	X7
X1	t2	0	0	0	0	0	0
X2	0	t1+t4+t6+t12	t1+t4+t6	t1+t4	t1+t4	t1	t1
X3	0	t1+t4+t6	t1+t4+t6+t11	t1+t4	t1+t4	t1	t1
X4	0	t1+t4	t1+t4	t1+t4+t5+t10	t1+t4+t5	t1	t1
X5	0	t1+t4	t1+t4	t1+t4+t5	t1+t4+t5+t9	t1	t1
X6	0	t1	t1	t1	t1	t1+t3+t8	t1+t3
X7	0	t1	t1	t1	t1	t1+t3	t1+t3+t7

Figure 5. An example of a tree with a corresponding variance-covariance matrix that describes shared branch lengths between the languages X1 through X7

between them. The idea of shared branch length is illustrated in Figure 5, in which X1 through X7 represent languages and t1 through t12 designate the different branches of the phylogenetic tree. In Figure 5, language X1 and X4 are not closely related and have no (0) shared branch length, while X2 and X3 are closely related sister languages and share the branches that lead to their most recent common ancestor, designated by t1+t4+t6. The information on shared branch length between all languages of the sample is used to remove covariance between the variables that is due to this shared history. Depending on the amount of shared branch length between each language pair, some of the covariance between the score on motion event encoding and the score on path verb lexicon size will be removed for that language pair.

Similar to the estimation of phylogenetic signal described in Section 3.2.1, the scaling parameter λ (lambda) is used to modify the branches of the tree such that the tree best represents the amount of shared history

present in the dataset. The effect that the scaling parameter λ has on the phylogenetic tree was illustrated in Figure 4. The most likely value of λ is estimated during the PGLS analysis so that the phylogenetic tree represents the best reflection of the amount of covariance due to shared descent between the data points.

The PGLS analysis was conducted on a set of 1,000 phylogenetic trees in order to take into account the uncertainty present in the phylogenetic estimation. PGLS analyses were conducted using the software *BayesTraits* (Pagel 1997, 1999), which uses a Bayesian reversible-jump Markov chain Monte Carlo framework to model and test hypotheses regarding the evolution of biological and linguistic traits. Due to space constraints a full explanation of how PGLS is conducted in a Bayesian framework cannot be given here, but please see Verkerk (to appear) and Pagel & Meade (n.d.) for more information. As a methodological note: the MCMC chains were run for 2×10^9 iterations for all four analyses. The PGLS estimates were sampled every 10^6 iteration. After removing burn-in, a posterior of 1,500 samples was taken from the stationary part of the chain.

4. Results

4.1. Non-phylogenetically corrected correlations

In order to make a comparison between results of non-phylogenetically corrected correlations and the PGLS analyses reported in Section 4.3, Pearson correlation analyses and a Linear Model regression were conducted. This was done both on the 118-sentence sample, which includes data on all 20 languages, as well as on the 192-sentence sample, which includes data on a subset of 16 languages. The measures used for the motion event encoding system as well as the size of the manner verb lexicon and the path verb lexicon are highly correlated when non-phylogenetic correlation analyses are carried out. This is true for both measures across samples (upper part of Table 5) and between motion event encoding and path verb lexicon size measures (lower part of Table 5).

The results from the regular Linear Model regression analysis in Table 6 indicate that without correcting for phylogenetic relatedness, the motion event encoding system as measured by the scores on the first PC1 and the proportion of use of the two main path encoding constructions is signifi-

Table 5. Pearson correlation coefficients for the motion event encoding and path verb lexicon size measures across and within samples

Measure 1	Measure 2	Pearson correlation coefficient
192-sentence PV count	118-sentence PV count	0.95
192-sentence % PVO+VF	118-sentence % PVO+VF	0.99
192-sentence PC1	118-sentence PC1	0.98
192-sentence PC1	192-sentence PV count	−0.67
118-sentence PC1	118-sentence PV count	−0.58
192-sentence % PVO+VF	192-sentence PV count	0.60
118-sentence % PVO+VF	118-sentence PV count	0.65

cantly correlated with the size of the path verb lexicon in each of the four different combinations of measures. The correlations are indicated by the slope of the dependent variable, the number of path verbs (slope PV count), as it is regressed on the independent variable, the scores on the first PC1 (PC1) and the proportion of use of the two main path encoding constructions (% PVO+VF). There is a negative correlation (slopes of −0.57 and

Table 6. Non-phylogenetically corrected LM regression analyses for path verb lexicon size on the motion event encoding system

Sample	Independent variable	Dependent variable	R^2	Coefficients	Standard error	t	p	
192-s.	PC1	intercept	0.42	1.33	0.03	47.54	$< 2 \times 10^{-16}$	***
		slope PV count		−0.57	0.18	−3.18	0.0067	**
118-s.	PC1	intercept	0.32	1.26	0.03	42.21	$< 2 \times 10^{-16}$	***
		slope PV count		−0.56	0.19	−2.94	0.0088	**
192-s.	% PVO+VF	intercept	0.28	1.00	0.13	7.49	2.94×10^{-6}	***
		slope PV count		0.97	0.42	2.32	0.036	*
118-s.	% PVO+VF	intercept	0.31	0.79	0.15	5.11	7.27×10^{-5}	***
		slope PV count		1.18	0.40	2.96	0.0084	**

Note: The independent variables used are the scores on the first principal component (PC1) and the proportion of use of the path verb-only and verb-framed constructions (% PVO+VF); the dependent variable used is the number of unique path verb types attested in the dataset (PV count). Codes for significance levels *p*: *** < 0.001; ** < 0.01; * < 0.05.

−0.56) between the score on the PC1 and the size of the path verb lexicon: a higher score on the PC1 (i.e. a less verb-framed nature) implies a smaller path verb lexicon. There exists a positive relationship (slopes of 0.97 and 1.18) between the proportion of use of the path verb-only and verb-framed constructions and the size of path verb lexicon size: a higher usage rate of the path verb-only and verb-framed constructions implies a larger path verb lexicon. However, since we know that the languages in the current dataset are all genealogically related, it is necessary, as will be evident from the next section, to use an analysis that can estimate phylogenetically corrected correlations. These will be discussed in Section 4.3.

4.2. Phylogenetic signal

The results of the tests for the presence of phylogenetic signal are presented in Table 7. As explained briefly in Section 3.2.1, the estimated λ (lambda) can be slightly higher than 1 due to characteristics of the phylogenetic tree. For the current two tree samples, the maximum λ is 1.28. As the analyses are conducted for all phylogenetic trees in a tree sample of 1,000 trees, the median as well as the range of the estimated λ values is provided. Table 7 also includes the results of the likelihood ratio tests. The number of trees for which λ was estimated *not to be significantly different* from a model of evolution in which λ was set to 1 and the number of trees for which λ was estimated *to be significantly different* from a model in which λ was set to 0 are listed in the last two columns of Table 7. In these two columns, a higher number of trees points to evidence for the presence of phylogenetic signal.

The estimated λ values for the PC1 scores, the proportion of path verb-

Table 7. Phylogenetic signal tests on the three measures in both samples

Sample	Data measure	Median λ	Range λ	λ = 1	λ = 0
192-sentence	PC1	1.16	1.09–1.28	1	1000
118-sentence	PC1	1.16	0.95–1.28	86	1000
192-sentence	% PVO+VF	1.01	1.07–1.22	1000	1000
118-sentence	% PVO+VF	1.01	0.87–1.14	1000	1000
192-sentence	PV count	0.95	0.83–1.08	1000	1000
118-sentence	PV count	0.93	0.82–1.03	1000	1000

Note: λ = 1 lists the number of trees for which λ was not estimated to be significantly different from 1 ($p < 0.05$); λ = 0 lists the number of trees for which λ was estimated to be significantly different from 0 ($p < 0.05$).

only and verb-framed constructions, and the path verb count are all quite high. This was the case for both sentence samples. For all six measures, the estimated λ was significantly different from 0 for all phylogenetic trees. For the proportion of path verb-only and verb-framed constructions and the simple path verb count, the estimated λ was found not to be significantly different from 1 for all phylogenetic trees. This was not the case for the estimated λ values for the PC1 scores. The reason for this is that the phylogenetic signal analyses should have used the maximum λ value for each given phylogenetic tree, and not simply λ = 1, to test the statistical significance of the estimated λ values. The maximum possible λ values ranged from 1.090 to 1.279 for the tree sample used for the analysis with the 192-sentence sample, and from 1.090 to 1.277 for the tree sample used for the analysis with the 118-sentence sample. However, these values could not be used as maximum possible λ values, as they are not accepted by the corPagel function from the R package *ape* (Paradis et al. 2004). This function only accepts fixed values for λ between 0 and 1. Despite this small issue with conducting the likelihood ratio tests, the high estimated lambda values as well as the fact that all estimated λ values are significantly different from 0 suggests that there is a clear phylogenetic signal for these three different data measures in both sentence samples.

4.3. Phylogenetic Generalized Least Squares

The results of the PGLS analyses are presented in Table 8. Since the analyses were conducted over a sample of 1,000 trees, only medians are shown. The medians of the λ values that were estimated during the PGLS analyses ranged between 0.67 and 0.84. This suggests that common descent explains at least some of the covariance between the measures for the motion event encoding system and the size of the path verb lexicon. The use of a technique that takes into account phylogenetic relatedness is therefore validated. The p values for the coefficients presented in the last column indicate that a statistically significant relationship (on the $p < 0.05$ level) between the motion event encoding system and the size of the path verb lexicon were found in all four PGLS analyses (PC1 + unique path verb count for both sentence samples and proportion of use of path verb-only and verb-framed constructions + unique path verb count for both sentence samples). Two of these relationships are statistically significant on a stricter level, that is, $p < 0.01$.

Table 8. PGLS regression analyses for the path verb lexicon size and the motion event encoding system

Sample	Independent variable	Dependent variable	R^2	λ	Coefficients	Standard error	p	
192-s.	PC1	intercept	0.33	0.67	1.31	0.05	$< 2 \times 10^{-16}$	***
		slope PV count			−0.61	0.23	0.011	*
118-s.	PC1	intercept	0.42	0.82	1.24	0.04	$< 2 \times 10^{-16}$	***
		slope PV count			−0.75	0.21	0.002	**
192-s.	% PVO+VF	intercept	0.23	0.72	0.97	0.16	$< 2 \times 10^{-16}$	***
		slope PV count			0.98	0.45	0.028	*
118-s.	% PVO+VF	intercept	0.41	0.84	0.66	0.17	$< 2 \times 10^{-16}$	***
		PV count			1.45	0.40	0.0013	**

Note: The independent variables used are the scores on the first principal component (PC1) and the proportion of use of the path verb-only and verb-framed constructions (% PVO+VF); the dependent variable used is the unique path verbs attested in the dataset (PV count). Codes for significance levels p: *** < 0.001; ** < 0.01; * < 0.05.

It seems useful to compare the results in Table 8 with those for the non-phylogenetically corrected LM regression analyses presented in Table 6. As was the case in Table 6, the results presented in Table 8 indicate that there exists a negative relationship between PC1 score and path verb lexicon size: a higher score on the PC1 (i.e. a more satellite-framed nature) implies a smaller path verb lexicon. There exists a positive relationship between the proportion of use of the path verb-only and verb-framed constructions and path verb lexicon size: a higher usage rate of the path verb-only and verb-framed constructions implies a larger path verb lexicon. R^2, which denotes the proportion of the variation in the dependent variable (size of the path verb lexicon) that is predicted by the independent variable (type of motion event encoding system), is between 0.23 and 0.42 in both Table 6 and 8. The significant PGLS results in Table 8 indicate that the size of the path verb lexicon is dependent on the amount of verb-framed behavior, but as R^2 is not very high (a high score would be .8 to 1.0), there clearly are other factors that act on path verb lexicon size, as also suggested in the literature (see Section 1). The values of the intercept and slope of the regression analyses, which describe the regression line, are quite similar in Table 6

and Table 8. Note that this is not necessarily the case: there are famous examples in biology for which a supposed correlation was no longer found after correcting for phylogenetic relationships. An example is the relationship between body size of different species of birds and the number of parasites found on them (Poulin 1995). Even though the non-phylogenetically corrected LM regression analyses and the PGLS analyses have quite similar results, the phylogenetic comparative method was needed to determine
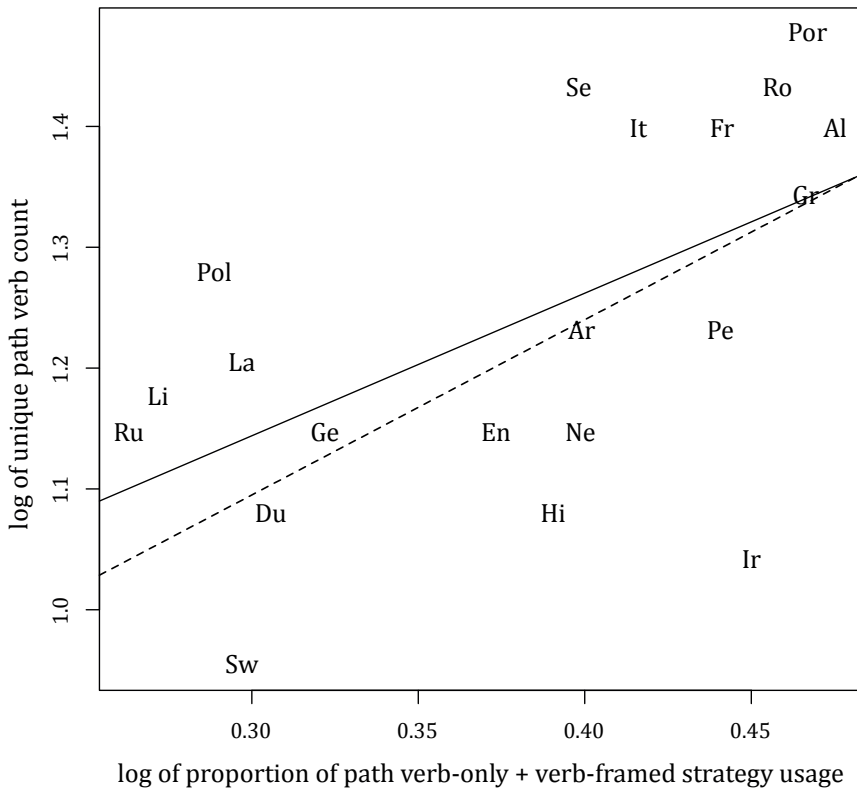
Figure 6. Comparison of the Linear Model regression (solid line) to the Phylogenetic Generalized Least Squares regression (dotted line) as found for the 118-sentence sample; with the independent variable (proportion of path verb-only + verb-framed strategy usage) on the x-axis and the dependent variable (unique path verb count) on the y-axis. Abbreviations used for language names: Al: Albanian; Ar: Armenian; Du: Dutch; En: English; Fr: French; Ge: German; Gr: Modern Greek; Hi: Hindi; Ir: Irish; It: Italian; La: Latvian; Li: Lithuanian; Ne: Nepali; Pe: Persian; Pol: Polish; Por: Portuguese; Ro: Romanian; Ru: Russian; Se: Serbo-Croatian; Sw: Swedish.

whether the correlation found in Table 6 holds after controlling for shared history so as to avoid finding misleading correlations of the type found by Poulin (1995).

In Figure 6, the results of one of the non-phylogenetically controlled LM regression analyses can be compared against the results of one of the PGLS analyses. These are the results for the analyses that correlate the proportion of use of path verb-only and verb-framed constructions with the unique path verb count for the 118-sentence sample. The x-axis represents the log of the proportion of use of path verb-only and verb-framed constructions, while the y-axis represents the log of the unique path verb count. The language abbreviations indicate the position of each language in this feature space. The solid line represents the regression function that was estimated by the non-phylogenetic LM analysis, whereas the dotted line represents the median regression function that was estimated by the PGLS analyses. The fit of both regression functions to the data suggests both methods describe the data adequately. But the most important result, as indicated by Table 8 and the dotted regression line in Figure 6, is that the relationship between motion event encoding and path verb lexicon size holds even when phylogenetic dependencies are taken into account.

5. Discussion

Tests for phylogenetic signal in Section 4 demonstrated that in the current data set, the language scores on the measures that relate the motion event encoding system and path verb lexicon size are dependent on the phylogenetic relationships between the languages (see again Table 7). In other words, languages that are closely related display similar behavior, both with respect to the encoding of motion events as well as the size of their path verb classes. The presence of this phylogenetic signal necessitated the use of statistical analyses such as PGLS that can detect and eliminate these dependencies in order to find whether there truly exists a correlation between the type of motion event encoding system and path verb lexicon size. The results of the PGLS analyses suggest that indeed there exists a correlation between the motion event encoding system, particularly the use of the path verb-only and verb-framed constructions, and the size of the path verb lexicon. They support the findings reported in the literature

on motion event encoding as discussed in Section 1, which also generally seem to suggest that verb-framed languages typically have larger path verb lexicons as compared with satellite-framed languages.

One issue that needs to be noted is that by employing two different samples and two different motion event encoding measures, which necessitates the execution of four different PGLS analyses, the chances of finding a significant correlation increase. Doing multiple statistical analyses inevitably increases the chances of finding results that confirm the researcher's hypothesis. Since the p-values of the PGLS analyses conducted on the 192-sentence sample are higher than those of the 118-sentence sample, ideally a larger sample of languages should be used. However, even if only a single PGLS using one sample (the biggest, 118-sentence sample) would have been conducted, using just one measure of motion event encoding (the traditional measure that relates the proportion of use of the path verb-only and verb-framed constructions), a statistically significant correlation ($p < 0.01$) between motion event encoding and path verb lexicon size would have been found.

The current results leave us asking if this correlation truly exists, through what mechanisms has it come about? There seem to be two processes involved in the creation and maintenance of the correlation between motion event encoding and path verb lexicon size in the Indo-European languages. The first process, which relates to the creation of path verbs and the emergence of path verb-only and verb-framed constructions in Romance and Indo-Iranian, is a diachronic process. As was explained in Section 2.2.2, Proto-Indo-European and the ancient Indo-European languages had a system of path denoting particles that could be positioned anywhere in the sentence (Delbrück 1893; Kuryłowic 1964; Bloch 1965; Lehmann 1974; Watkins 1964). The path particles that were closely associated with verbs formed the preverb systems that later became spatial path prefixes. During the transition from Latin to the modern Romance languages (see Kopecka 2013 on French), as well as from Sanskrit to the modern Indo-Aryan languages, these path prefixes merged with the verb completely. In some Indo-European languages, including many of the Balto-Slavic languages, the system of path prefixes is still in place. Almost all Indo-European languages have path verbs that are demonstrable mergers of path prefixes with verbs.

Verkerk (to appear) has studied the etymologies of the path verbs that were encountered in the 118-sentence sample in detail, and has demon-

strated that most of the path prefix + verb mergers are contemporary: they originated after the languages have split up from their closest sister languages. An example of this is the parallel evolution of verbs denoting RETURN in many Indo-European languages, including French *retourner*, Portuguese *retornar*, Dutch *terugkeren*, German *zurückkehren*, Swedish *återvända*, Greek *epistrefo*, Persian *bāzgaštan*, and Russian *vozvrašat'sja*. These verbs are all based on the combination of a prefixal element meaning something like 'back' with a verb meaning 'turn'. Other contemporary path prefix + verb mergers are Albanian *përparoj* 'advance', French *contourner* 'go around', Irish *éirigh* 'rise', Latvian *nonākt* 'reach', Lithaunian *atsidurti* 'get somewhere', Polish *przeprawiać* 'cross', and Serbo-Croatian *unići ~ ući* 'enter'. The merging of path prefix + verbs therefore has played an important role and is still involved in adding new verbs to the path verb lexicons of most Indo-European languages.

However, an important set of path verbs can be demonstrated to have a much longer history: these are path prefix + verb mergers that have merged at a much earlier stage and they have subsequently been inherited by the Romance languages and the Indo-Iranian languages. Examples of these are French *descendre*, Portuguese *descer*, and Italian *scendere* 'to descend', from Latin *descendere*, a prefix-verb combination of *de* 'from' and *scandere* 'to ascent, to mount' (Walde 1930–1956). Another example is Hindi *pahuṃcnā* and Nepali *pahūcnu* 'to arrive', from Sanskrit *prábhurati*, a prefix-verb combination of *pra* 'forward' and *bhuráti* 'to move' (Turner 1962–1985). Sixteen percent of the Romance and 19% of the Indo-Iranian path verb lexicons consists of these inherited path prefix + verb mergers, whereas none of the other big Indo-European subgroups (Germanic and Balto-Slavic) show evidence for such ancient path prefix + verb mergers (Verkerk to appear).

The merging of path prefixes with verb roots has not only affected the path verb lexicons of Indo-European languages, but has also had a pronounced effect on the motion event encoding system – especially in Romance and Indo-Iranian, where the most frequent use of path verb-only and verb-framed constructions is attested. As the path prefix system became unproductive, the satellite-framed construction was used less and less, as only a limited set of prepositions, adverbs, and for some languages cases remained to encode path. At the same time, new monomorphemic path verbs emerged as path prefixes merged with verb roots. In effect, then, the emergence of larger path verb lexicons and a more frequent use of path verb-only and verb-framed construction in Romance and Indo-Iranian are

two sides of the same coin, the coin being the merging of path prefixes with verb roots. The correlation between the motion event encoding system and the size of the path verb lexicon can therefore be attributed to the earlier onset and far more extensive emergence of path prefix + verb mergers in Romance and Indo-Iranian as opposed to Germanic and Balto-Slavic.

Sketching this scenario leads to the question why this diachronic merging process affected Romance and Indo-Iranian earlier and to a larger extent than Germanic and Balto-Slavic, as well as why the differences between Romance and Balto-Slavic and Germanic are still maintained. The most plausible second process that seems to be acting on motion event encoding in Indo-European is that of areal contact. Several researchers, including Wälchli (2009: 214) and Slobin (2005), have pointed out the existence of a north-south divide in motion event encoding: north and central Europe are satellite-framed (including the Germanic languages and the Balto-Slavic languages, several Finno-Ugric languages such as Estonian, Finnish, and Hungarian, several Daghestanian languages such as Lezgian, as well as Georgian) whereas south Europe is verb-framed (including the Romance languages, Basque, Albanian, and Turkish), with some languages in between (including various Italian dialects (Kramer 1981; Talmy 2000: 145; Slobin 2005; Masini 2006; Iacobini & Masini 2007), Serbo-Croatian (Filipović 2007; Vidaković 2012), and Modern Greek (Hickmann et al. to appear)). In fact, Wälchli (2009) has shown that most languages around the globe are verb-framed, making satellite-framed northern and central Europe one of a limited set of exceptions. It seems likely that the systems that north and central European languages employ to encode path on satellites are maintained because of the close proximity of and contact with similar languages.

The path prefix + verb merging process seems to have generated a special prevalence to encode path on verbs in the Romance languages. Many path verbs have cognates throughout the Romance subfamily, such as French *entrer*, Italian *entrare*, Romanian *intra* and Portuguese *entrar* 'to enter', from Latin *intrare* 'to enter', which is derived from *inter* 'among, between' (De Vaan 2008). Strikingly, the rest of the Romance path verbs are actually not contemporary path prefix + verb mergers, but often they are semantic shifts or derivations from non-verbal elements like adverbs and spatial nouns (Verkerk to appear). This suggests that once the Romance languages had these merged path verbs, they employed other mechanisms to generate even more path verbs. There are also many near-synonyms,

such as Portuguese *regressar*, *voltar*, and *retornar* 'to return'. Clearly, the existence of one verb that means 'to return' does not necessarily inhibit the formation of verbs with similar meanings. In verb-framed languages, path verbs can diversify to have a higher degree of semantic granularity as path verbs from satellite-framed languages (just as manner verbs in satellite-framed languages, see Slobin 2004; Verkerk 2013).

Although Latin and Sanskrit underwent very similar changes, resulting in merged path verbs in both the modern Romance and Indo-Aryan languages, the Indo-Aryan languages do not have a large path verb lexicon like the Romance languages (see Table 3). As was also evident from Figure 1 and 2, the Indo-Iranian languages Hindi, Nepali, and Persian are not verb-framed like the Romance languages, although they use the path verb-only and verb-framed construction relatively frequently as well. The difference between the Romance and the Indo-Aryan languages seems to be that while the Romance languages continued to add path verbs to their lexicons throughout their history, making it bigger all the time, the Indo-Aryan languages did not (or if they did, some path verbs were lost). My impression is that this might be due to frequent usage of the deictic verb-only strategy, which is slightly more common in Indo-Iranian than in Romance. However, the difference in deictic verb-only strategy usage is not very big in the current sample (see Figure 1), and clearly more investigation of the Indo-Iranian languages is needed. In addition, it is clear that many more factors influence the size and structure of classes of the lexicon such as the path verb class (see again Section 1 and the R^2 values in Table 8). This is especially evident for languages such as the Indo-Iranian languages that do not have large classes of manner verbs (Verkerk 2013) nor large classes of path verbs. It could be that certain factors are acting on their motion verb lexicons to keep them small. However, it could also be the case that Romance has an exceptional number of path verbs, and Germanic and Balto-Slavic have an exceptional number of manner verbs. Indo-European languages are not the norm for what is common or unmarked in the languages of the world, and very often are quite exotic (Dahl 1990).

6. Summary and conclusion

This article investigated the relationship between the motion event encoding system, specifically the use of the path verb-only and verb-framed

motion event encoding constructions, and the size of the path verb lexicon, measured by the number of unique verb types. In order to do this, comparable data on motion encoding and path verbs were collected from a parallel corpus of novels in 20 Indo-European languages. Testing for phylogenetic signal revealed that languages behave similar because of phylogenetic relationships. In order to account for these dependencies, Phylogenetic Generalized Least Squares (PGLS) regression analyses were conducted. These analyses uncovered a relationship between motion event encoding and the size of the path verb lexicon: more frequent use of the path verb-only and verb-framed motion event encoding constructions is correlated with a larger path verb lexicon. This finding can be explained by the merging of path prefixes with verb roots, a diachronic process that has affected most Indo-European languages but especially the Romance and Indo-Aryan subfamilies. Areal contact between satellite-framed languages in north and central Europe as well as between verb-framed languages in south Europe has possibly played a role in maintaining and amplifying the difference between verb-framed and path verb rich Romance in the south and satellite-framed and path verb poor Germanic and Balto-Slavic in the north.

Understanding the dynamics of the relationship between the constructions we use and the words we put in them will continue to be an important aspect in motion event encoding research. This article has made a contribution to this topic by showing that these dynamics are real and that it is possible for these dynamics to act on both syntax and the lexicon.

Appendix 1. Motion encoding classifications

Table A1. Motion encoding classifications made in the literature (this is not intended to be a full overview for individual languages)

Language	Classification	Source
Dutch	satellite-framed	Slobin (2005, 2006); Croft et al. (2010)
English	satellite-framed	Talmy (1985)
Swedish	satellite-framed	Viberg (2006)
German	satellite-framed	Berthele (2006)
Russian	satellite-framed	Slobin (2005)
Polish	satellite-framed/mixed	Kopecka (2009)

(cont.)

Table A1. (Cont.)

Language	Classification	Source
Serbo-Croatian	satellite-framed/mixed	Filipović (2007); Slobin (2005)
Lithuanian		no previous Talmian literature
Latvian		no previous Talmian literature
Portuguese	verb-framed	Slobin (2005)
French	verb-framed	Jones (1983); Kopecka (2006); Pourcel and Kopecka (2005)
Italian	verb-framed	Folli (2008); Iacobini & Masini (2006)
Romanian		no previous Talmian literature
Hindi	verb-framed	Narasimhan (2003)
Nepali		no previous Talmian literature
Persian	mixed	Feiz (2011)
Irish		no previous Talmian literature
Albanian		no previous Talmian literature
Armenian		no previous Talmian literature
Greek	verb-framed/mixed	Papafragou et al. (2006); Talmy (2007: 105); Hickmann et al. (to appear)

Appendix 2. Lists of path verb types encountered in the sample

The following tables present an overview of the path verbs used by the 20 languages in the sample. Verbs that occurred in the 118-sentence sample are presented in plain text. Verbs that occurred only in the 192-sentence sample are presented in italics. This distinction is made in order to give an overview of path verbs in these 20 languages that is as extensive as possible, while also making clear which verbs appear only in the 192-sentence sample. The number of path verbs in each language has have been provided in Table 3 in the main text. The path verbs are categorized as follows (partly taken from Narasimhan 2003):

1. complex endpoint (encodes features of the endpoint, e.g. whether it is an enclosure or a moving object): *enter, follow, land*
2. complex source-point (encodes features of the source point, e.g. whether it is an enclosure): *exit, escape*
3. simple endpoint: *arrive*

4. simple potential endpoint: *head for*, *approach*
5. simple source-point: *leave*
6. mid-point: *pass*
7. direction up: *go up*, *rise*
8. direction down: *go down*, *fall*, *dive*
9. direction forward: *advance*
10. direction around: *go around*, *turn*
11. direction back: *return*
12. direction behind: *go behind*
13. spatial features of the path: *cross*, *penetrate*

Table A2a. Path verbs in Romance

French	Portuguese	Italian	Romanian
Complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object			
entrer 'enter'	entrar 'enter' enfiar-se 'enter'	entrare 'enter'	intra 'enter'
suivre 'follow' *poursuivre 'pursue'* *traquer 'pursue'* atterrir 'land'	seguir 'follow' *perseguir 'pursue'* aterrar 'land'	seguire 'follow' accodarsi 'follow' *inseguire 'pursue'* atterrare 'land'	urma 'follow' *urmări 'pursue'*
Complex source-point; encodes features of the source point, e.g. whether it is an enclosure			
fuir 'flee' s'échapper 'escape'	sair 'go out' fugir 'flee' escapulir-se 'escape' escaper 'escape' deixar 'abandon' abandonar 'leave'	uscire 'go out' lasciare 'leave' scappare 'escape' (s)fuggire 'flee'	ieşi 'go out' scăpa 'escape' părăsi 'abandon'
Simple endpoint			
arriver 'arrive' *atteindre 'reach'*	chegar 'arrive'	(rag)giungere 'arrive' arrivare 'arrive' *ficcarsi 'get to'*	ajunge 'arrive'

(cont.)

Table A2a. (*Cont.*)

French	Portuguese	Italian	Romanian
Simple potential endpoint			
se diriger 'head for'	dirigir-se 'go to'		(se) lua 'take oneself' se îndrepta 'go toward'
s'approcher 'approach'	aproximar-se 'approach' *abeirar-se 'draw near'*	avvicinarsi 'approach'	se apropia 'come near' se alătura 'approach'
Simple source-point			
partir 'leave' sortir 'leave' quitter 'leave' s'éloigner 'depart'	partir 'leave' afastar-se 'distance oneself' distanciar-se 'distance oneself'	partire 'leave' andarsene 'go away' allontanarsi 'go away' *avviarsi 'set out'*	pleca 'leave' (în)depărta 'leave' porni 'depart' se duce 'leave'
Midpoint			
passer 'pass' *franchir 'pass'*	passer 'pass'	passare 'pass'	trece 'pass'
Direction up			
monter 'go up' *s'élever 'rise'*	subir 'go up' erguer-se 'rise'	montare 'go up' sollevarsi 'rise' (ri)salire 'go up'	sui 'go up' se ridica 'rise' urca 'go up'
direction down (re)tomber 'fall' *basculer 'fall'*	cair 'fall'	cadere 'fall'	cădea 'fall'
descendre 'come down'	descer 'descend'	scendere 'descend'	coborî 'descend'
	mergulhar 'dive'		cufunda 'dive'
Direction forward			
(s')avancer 'go forward'		avanzare 'move forward'	

French	Portuguese	Italian	Romanian
Direction around			
contourner 'go around'	contornar 'go around'		ocoli 'go around'
	virar 'turn'	(ag)girare 'turn' *svoltare 'turn'*	coti 'turn'
Direction back			
repartir 'return' revenir 'return' retourner 'return' rentrer 'return' rejoindre 'return' *regagner 'return'*	regressar 'return' voltar 'return' retornar 'return'	(ri)tornare(se) 'return'	se întoarce 'return'
Spatial features of the path			
traverser 'cross' *croiser 'cross'* pénétrer 'penetrate'	atravessar 'cross' cruzar 'cross' penetrar 'penetrate'	(at)traversare 'cross' penetrare 'penetrate' *inoltrarsi 'penetrate'*	traversa 'cross' străbate 'cross' pătrunde 'penetrate'

Table A2b. Path verbs in Germanic

English	German	Dutch	Swedish
Complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object			
pursue follow	folgen 'follow' *verfolgen 'pursue'*	volgen 'follow' *achtervolgen 'pursue'*	följa 'follow' hamna 'land'
Complex source-point; encodes features of the source point, e.g. whether it is an enclosure			
flee	verlassen 'leave'	ontsnappen 'escape' verlaten 'leave' vluchten 'flee'	fly 'flee'

(cont.)

Table A2b. (*Cont.*)

English	German	Dutch	Swedish
Simple endpoint			
reach get arrive	sich begeben 'set off' gelangen 'reach' eintreffen 'arrive' *erreichen 'reach'*	arriveren 'arrive' *bereiken 'reach'*	*anlända 'arrive'*
Simple potential endpoint			
	sich nähern 'approach'	*aanzetten 'approach'*	*närma sig 'approach'*
Simple source-point			
leave	sich entfernen 'depart' *weichen 'leave'*	zich begeven 'set off'	avlägsna sig 'remove oneself' lämna 'leave' ta sig 'take oneself'
Midpoint			
pass	überholen 'pass'	*passeren 'pass'*	*passera 'pass'*
Direction up			
rise	sich erheben 'rise'	stijgen 'rise'	*lyfta 'take off'*
Direction down			
descend fall	fallen 'fall' eintauchen 'dive'	vallen 'fall'	falla 'fall'
Direction around			
turn		omslaan 'turn' *draaien 'turn'*	vika 'turn'
Direction back			
return	zurückkehren 'return'	terugkeren 'return'	återvända 'return'

English	German	Dutch	Swedish
Spatial features of the path			
cross	durchqueren 'cross'	oversteken 'cross'	
	überqueren 'cross'		
	eindringen	doordringen	
	'penetrate'	'penetrate'	

Table A2c. Path verbs in Balto-Slavic

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
Complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object				
	następować 'follow'	*progoniti 'pursue'*	sekti 'follow'	sekot 'follow'
	ganiać się 'chase'		*vaikyti 'chase'*	
		unići ~ ući 'enter'		iekļūt 'get in'
		se uvaliti 'enter'		
Complex source-point; encodes features of the source point, e.g. whether it is an enclosure				
	porzucić 'abandon'	napustiti 'abandon'	palikti 'abandon'	pamest 'abandon'
				izkļūt 'get out'
				bēgt 'flee'
Simple endpoint				
	przybyć 'arrive'	stići 'arrive'	patekti 'get'	ierasties 'arrive'
	dostać 'get'	*dospjeti 'arrive'*	atsidurti 'get'	tikt 'get'
	dotrzeć 'get'		*siekti 'reach'*	*nokļūt 'arrive'*
				sasniegt 'reach'
				nonākt 'reach'
				aizkļūt 'get'
Simple potential endpoint				
pravit'sja 'direct oneself to'	skierować się 'go to'		traukti 'go to'	

(*cont.*)

Table A2c. (*Cont.*)

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
blizit'sja 'approach'	zbliżać się 'approach'	pridolaziti 'approach' se približiti 'approach'		

Simple source-point

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
dvinut'sja 'set out' brat'sja 'take oneself' *skryt'sja* *'go away'*	opuścić 'leave' wybrać się 'set out' oddalić 'go away' ruszyć 'set out'	otići ~ odlaziti 'leave' izaći 'leave' poći 'leave' se ukloniti 'go away' se uputiti 'set out' ostaviti 'leave' se udaljiti 'go away' krenuti 'set out'	tolti 'go away' sileisti 'let oneself go'	doties 'set out' attālināties 'go away' atstāt 'leave'

Midpoint

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
minovat' 'pass'	*minąć 'pass'*	prolaziti ~ proći 'pass'		

Direction up

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
podnjat'sja ~ podnimat'sja 'go up'		dignuti ~ *dizati* 'rise' se penjati 'rise' se popeti 'go up'	kilti 'rise'	

Direction down

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
padat' ~ past' 'fall' valit'sja 'fall' spuskat'sja ~ spustit'sja 'go down'	padać 'fall'	padati ~ pasti 'fall' se rušiti 'fall' sići ~ silaziti 'go down'	kristi 'fall' virsti 'fall' byrėti 'fall'	krist 'fall' birt 'fall'

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
	nurkować 'dive'	zaroniti 'dive'		nirt 'dive'

Direction around

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
kružiť 'go round' ogibať 'go round'	krążać 'go around' zbaczać 'go around'	zaobići 'go around'	lenkti 'go around'	*riņķot 'circle'*
	kręcić 'turn'		sukti 'turn' vingiuoti 'turn'	griezties 'turn'

Direction back

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
vozvrašať sja 'return' vernuť sja 'return'	wrócić 'return'	vratiti ~ vraćati 'return'	grįžti 'return'	

Direction behind

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
		zalaziti 'go behind'		

Spatial features of the path

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
pereseč ~ peresekat' 'cross'	przeprawiać 'cross' *przekraczać 'cross'*	prelaziti ~ prijeći 'cross'		šķērsot 'cross'
		prodrijeti 'penetrate'		

Table A2d. Path verbs in Indo-Aryan

Hindi	Nepali	Persian
Complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object		
samānā 'go in'	pasnu 'go in'	dāḵel šodan 'go in' vāred šodan 'go in' darāmadan 'come in'
Complex source-point; encodes features of the source point, e.g. whether it is an enclosure		
nikalnā 'go out' tyāgnā 'abandon'	niskanu 'go out' chodnu 'abandon'	ḵārej šodan 'go out' tark kardan 'abandon'
Simple endpoint		
pahuṃcnā 'arrive'	pugnu 'reach'	(az rāh) residan 'arrive'
Simple potential endpoint		
		nazdik šodan 'approach'
Simple source-point		
choṛnā 'leave'		dur sakhtan 'go away' vāgozāštan 'leave' (be) rāh oftādan 'set out'
Midpoint		
guzarnā 'pass'		gozaštan 'pass'
Direction up		
	uklinu 'go up'	barḵāstan 'rise'
Direction down		
girnā 'fall'	khasnu 'fall' jharnu 'fall' barsinu 'fall'	oftādan 'fall'
utarnā 'go down'	orlinu 'go down' dubnu 'go down'	
Direction forward		
calnā 'go forward, advance' baṛhnā 'advance'	aghi badhnu 'advance'	

Hindi	Nepali	Persian
Direction around		
ghūmnā 'turn'		dowr zadan 'go around'
Direction back		
	pharkinu 'return'	bāzgaštan 'return'
		bāzāmadan 'return'
Spatial features of the path		
pār karnā 'cross'	tarnu 'cross'	
	par garnu 'cross'	
		nofuz kardan 'penetrate'

Table A2e. Path verbs in Albanian, Armenian, Modern Greek and Irish

Albanian	Armenian	Greek	Irish
Complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object			
futem 'go in'	mtnel 'go in'	mpaino 'go in'	
hyj 'go in'			
ndjek 'follow'	hetevel 'follow'	akoloytho 'follow'	lean 'follow'
	hetapndel 'pursue'	kynigo 'chase'	*tóraigh 'pursue'*
		egkataleipo	
		'abandon'	
Complex source-point; encodes features of the source point, e.g. whether it is an enclosure			
dal 'go out'		vgaino 'go out'	
iki 'escape'			éalaigh 'escape'
braktis 'abandon'	t'oghnel 'abandon'		
Simple endpoint			
arrij 'arrive'	hasnel 'arrive'	ftano 'reach'	sroich 'reach'
mbërrij 'arrive'			
Simple potential endpoint			
afrohem 'approach'	motenal 'approach'	plisiazo 'approach'	
qasem 'approach'			

(cont.)

Table A2e. (*Cont.*)

Albanian	Armenian	Greek	Irish
Simple source-point			
drejtohem 'head for'		travo 'head for'	
largohem 'depart' u nisem 'set out' lë 'leave'	herranal 'go away'	(xe)feygo 'go away' afino 'leave' apomakryno 'remove oneself'	imigh 'leave' fág 'leave' *scoith 'leave'*
Midpoint			
	ants'nel 'pass'	perno 'pass'	cuir do 'pass'
Direction up			
ngrihem 'rise' ngjitem 'go up' hipi 'go up'	bardzranal 'rise'	anevaino 'go up' sikono 'rise'	éirigh 'rise'
Direction down			
bie 'fall'	ynkel 'fall' t'ap'vel 'fall' teghal 'fall'	pefto 'fall'	tit 'fall'
zbres 'go down'	ijnel 'go down' suzvel 'dive'	katevaino 'go down'	
Direction forward			
përparoj 'advance'	arrajanal 'go forward'		
Direction around			
rrotullohem 'go round'	shrjants'el 'go round'	parakampto 'go round' strivo 'turn'	*cuir 'turn'*
Direction back			
kthej ~ kthehem 'return'	(vera)darrnal 'return'	gyrizo 'return, turn' epistrefo 'return'	fill 'return'

Albanian	Armenian	Greek	Irish
Spatial features of the path			
kaloj 'cross' përshkoj 'cross' kapërcej 'cross'		diaschizo 'cross'	trasnaigh 'cross' cuir de 'cross'
	t'ap'ants'el 'penetrate'	dieisdyo 'penetrate'	

References

Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46: 331–377.

Berman, Ruth A. & Dan. I. Slobin, eds. 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berthele, Raphael. 2006. *Ort und Weg. Die sprachliche Raumreferenz in Varietäten des Deutschen, Rätoromanischen und Französischen*. Berlin: De Gruyter.

Bloch, Jules. 1965. *Indo-Aryan: From the Vedas to modern times*. Paris: Librarire Adrien Maisonneuve.

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.

Cifuentes Férez, Paula. 2010. The semantics of the English and the Spanish motion verb lexicons. *Review of Cognitive Linguistics* 8: 233–271.

Chen, Liang & Jiansheng Guo. 2009. Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics* 41: 1749–1766.

Croft, William, Johanna Barðdal, Willem Hollman, Violeta Sotirova & Chiaki Taoka. 2010. Revising Talmy's typological classification of complex event constructions. In Hans C. Boas, ed. *Contrastive studies in construction grammar*. Amsterdam: John Benjamins, 201–235.

Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *STUF – Sprachtypologie und Universalienforschung* 60: 95–99.

Dahl, Östen. 1990. Standard Average European as an exotic language. In Johannes Bechert, Giuliano Bernini & Claude Buridant, eds. *Towards a typology of European languages*. Berlin: Mouton de Gruyter, 3–8.

Delbrück, Berthold. 1893. *Vergleichende Syntax der Indogermanischen Sprachen*. Strassburg: K. J. Trübner.

Drummond, Alexei J., Marc A. Suchard, Dong Xie & Andrew Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Feiz, Parastou. 2011. Travelling through space in Persian and English: A comparative analysis of motion events in elicited narratives. *Language Sciences* 33: 401–416.

Filipović, Luna. 2007. *Talking about motion: A crosslinguistic investigation of lexicalization patterns*. Amsterdam: John Benjamins.

Folli, Raffaella. 2008. Complex PPs in Italian. In Anna Asbury, Jakub Dotlačil, Berit Gehrke and Rick Nouwen, eds. *Syntax and semantics of spatial P*. Amsterdam: John Benjamins, 197–220.

Freckleton, R. P., Paul H. Harvey & Mark Pagel. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. *The American Naturalist* 160: 712–726.

Harmon, Luke J., Jason T. Weir, Chad D. Brock, Richard E. Glor & Wendell Challenger. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24: 129–131.

Hickmann, Maya, Henriëtte Hendriks, Efstathia Soroli, Tatiana Iakovleva & Yinglin Ji. to appear. *Space and language typology: Encoding motion across languages*.

Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian inference of phylogeny and its impact of evolutionary biology. *Science* 294: 2310–2314.

Iacobini, Claudio & Francesca Masini. 2006. The emergence of verb-particle constructions in Italian: Locative and actional meanings. *Morphology* 16: 155–188.

Iacobini, Claudio & Francesca Masini. 2007. Verb-particle constructions and prefixed verbs in Italian: typology, diachrony and semantics. In Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli & Sergio Scalise, eds. *Proceedings of the Fifth Mediterranean Morphology Meeting*. Bologna: Università degli Studi di Bologna, 157–184.

Ibarretxe-Antuñano, Iraide. 2004. Language typologies in our language use: The case of Basque motion events in adult oral narratives. *Cognitive Linguistics* 15: 317–349.

Jones, Michael A. 1983. Speculations on the expression of movement in French. In Jacques Durand, ed. *A festschrift for Peter Wexler*. Essex: University of Essex Language Center, 165–194.

Kawachi, Kazuhiro. 2011. Are events that are expressed in verb-framed languages using non-characteristic patterns really macro-events? An examination of motion expressions in Sidaama (Sidamo). *Papers from the 11th National Conference of the Japanese Cognitive Linguistics Association*, 114–124.

Kopecka, Anetta. 2006. The semantic structure of motion verbs in French: Typological perspectives. In Maya Hickmann & Stéphane Robert, eds. *Space in languages: Linguistic systems and cognitive categories*. Amsterdam: John Benjamins, 83–102.

Kopecka, Anetta. 2009. Motion events in Polish: Lexicalization patterns and the description of Manner. In Victoria Hasko & Renee Perelmutter, eds. *New approaches to Slavic verbs of motion*. Amsterdam: John Benjamins, 225–246.

Kopecka, Anetta. 2013. Describing motion events in Old and Modern French. In Juliana Goschler & Anatol Stefanowitsch, eds. *Variation and change in the encoding of motion events*. Amsterdam: John Benjamins, 163–183.

Kramer, Johannes. 1981. Die Übernahme der deutschen und der niederländischen Konstruktion Verb + Verbzusatz durch die Nachbarsprachen. In Wolfgang Meid & Karin Heller, eds. *Sprachkontakt als Ursache von Veränderungen der Sprach- und Bewusstseinsstruktur: Eine Sammlung von Studien zur sprachlichen Interferenz*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 129–40.

Kuryłowicz, Jerzy. 1964. *The inflectional categories of Indo-European*. Heidelberg: Carl Winter Universitätsverlag.

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

Levinson, Stephen C. & Niclas Burenhult. 2009. Semplates: A new concept in lexical semantics? *Language* 85: 153–174.

Masini, Francesca. 2006. Diacronia dei verbi sintagmatici in italiano. *Archivio Glottologico Italiano 2006*, XCI: 67–105.

Matsumoto, Yo. 1997. Kuukan-idoo no gengo-hyoogen to sono kakuchoo [Linguistic expressions of spatial motion and their extensions]. In Shigenori Tanaka & Yo Matsumoto, eds. *Kuukan to idoo no hyoogen*. Tokyo: Kenkyusha.

Matsumoto, Yo. 2003. Typologies of lexicalization patterns and event integration: Clarifications and reformulations. In Shuji Chiba, ed. *Empirical and theoretical investigations into language: A festschrift for Masaru Kajita*. Tokyo: Kaitakusha.

Naigles, Letitia R., Ann R. Eisenberg, Edward T. Tako, Melissa Highter & Nancy McGraw. 1998. Speaking of motion: Verb use in English and Spanish. *Language and Cognitive Processes* 13: 521–549.

Narasimhan, Bhuvana. 2003. Motion events and the lexicon: A case study of Hindi. *Lingua* 113: 123–160.

Özçalışkan, Şeyda. 2004. Typological variation in encoding the manner, path, and ground components of a metaphorical motion event. *Annual Review of Cognitive Linguistics* 2: 73–102.

Özçalışkan, Şeyda. 2009. Learning to talk about spatial motion in language-spe-

cific ways. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan, eds. *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*. New York: Psychology Press, 263–276.

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*. Istanbul: Bogazici University Press, 259–270.

Pagel, Mark. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26: 331–348.

Pagel, Mark. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.

Pagel, Mark & Andrew Meade. n.d. *Manual BayesTraits*, ms. Reading: University of Reading.

Papafragou, Anna, Christine Massey & Lila Gleitman. 2006. When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition* 98: B75–B87.

Paradis, Emmanuel, Julien Claude & Korbinian Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.

Poulin, Robert. 1995. Phylogeny, ecology, and the richness of parasite communities in vertebrates. *Ecological Monographs* 65: 283–302.

Pourcel, Stéphanie & Anetta Kopecka. 2005. Motion expression in French: Typological diversity. *Durham & Newcastle Working Papers in Linguistics* 11: 139–153.

Revell, Liam J. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution* 63: 3258–3268.

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

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*. Amsterdam: John Benjamins, 437–467.

Slobin, Dan I. 2003. Language and thought online: Cognitive consequences of linguistic relativity. In Deidre Gentner & Susan Goldin-Meadow, eds. *Advances in the investigation of language and thought*. Cambridge: MIT Press, 157–191.

Slobin, Dan I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömqvist & Ludo Verhoven, eds. *Relating events in narrative: Typological and contextual perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, 219–257.

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*. Dordrecht: Kluwer, 115–129.

Slobin, Dan I. 2006. What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In Maya Hickmann & Stéphane Robert, eds. *Space in languages: Linguistic systems and cognitive categories*. Amsterdam: John Benjamins, 59–82.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen, ed. *Language typology and syntactic description: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press, 57–149.

Talmy, Leonard. 1991. Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: University of California, 480–519.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Talmy, Leonard. 2007. Lexical typologies. In Timothy Shopen, ed. *Language typology and syntactic description*. Vol 2: *Grammatical categories and the lexicon*. Cambridge: Cambridge University Press, 66–168.

Turner, Ralph Lilley. 1962–1985. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

De Vaan, Michiel. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.

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: 169–217.

Verkerk, Annemarie. 2014. 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*. Berlin: Walter de Gruyter, 324–354.

Verkerk, Annemarie. to appear. Where do all the motion verbs come from? The speed of development of manner verbs and path verbs in Indo–European. *Diachronica*.

Viberg, Åke. 2006. Towards a lexical profile of the Swedish verb lexicon. *STUF – Sprachtypologie und Universalienforschung* 59: 103–129.

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*. Amsterdam: John Benjamins, 233–262.

Wälchli, Bernhard. 2009. *Motion events in parallel texts: A study in primary-data typology*. Habilitationsschrift, Universität Bern.

Walde, Alois. 1930–1956. *Lateinisches etymologisches Wörterbuch*. Heidelberg: Winter.

Watkins, Calvert. 1964. Preliminaries to the reconstruction of Indo-European sentence structure. In Horace G. Lunt, ed. *Proceedings of the IX International Congress of Linguists*. The Hague: Mouton, 1035–1045.

Wienold, Götz. 1995. Lexical and conceptual structures in expressions for movement and space: With reference to Japanese, Korean, Thai, and Indonesian as compared to English and German. In Urs Egli, Peter E. Pause, Christoph Schwarze, Arnim von Stechow & Götz Wienold, eds. *Lexical knowledge in the organization of language*. Amsterdam: John Benjamins, 301–340.

Author's address:
Evolutionary Processes in Language and Culture Group
Max Planck Institute for Psycholinguistics
P. O. Box 310
6500 AH Nijmegen
The Netherlands
a.verkerk@reading.ac.uk