

Scramble, Scurry and Dash: The Correlation between Motion Event Encoding and Manner Verb Lexicon Size in Indo-European

Annemarie Verkerk

*Evolutionary Processes in Language and Culture Group,
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
a.verkerk@reading.ac.uk*

Abstract

In recent decades, much has been discovered about the different ways in which people can talk about motion (Talmy, 1985, 1991; Slobin, 1996, 1997, 2004). Slobin (1997) has suggested that satellite-framed languages typically have a larger and more diverse lexicon of manner of motion verbs (such as *run*, *fly*, and *scramble*) when compared to verb-framed languages. Slobin (2004) has claimed that larger manner of motion verb lexicons originate over time because codability factors increase the accessibility of manner in satellite-framed languages. In this paper I investigate the dependency between the use of the satellite-framed encoding construction and the size of the manner verb lexicon. The data used come from 20 Indo-European languages. The methodology applied is a range of phylogenetic comparative methods adopted from biology, which allow for an investigation of this dependency while taking into account the shared history between these 20 languages. The results provide evidence that Slobin's hypothesis was correct, and indeed there seems to be a relationship between the use of the satellite-framed construction and the size of the manner verb lexicon.

Keywords

motion events; manner verbs; phylogenetic comparative methods; PGLS; Indo-European

1. Introduction

This paper is concerned with manner of motion verbs (manner verbs) such as *walk*, *swim* and *dash*: how is information on the manner of motion lexicalized in verbs, and is the evolution of these verbs influenced by the syntactic constructions in which they are used? Recent decades have seen an increase in studies in the semantic and lexical typological traditions that address these types of questions (color: Berlin and Kay, 1969; Kay and Maffi, 1999; perception verbs: Viberg, 1983; Evans and Wilkins, 2000; spatial relations: Levinson and Meira, 2003; body parts: Majid et al., 2006; aquamotion verbs: Koptjevskaja-Tamm et al., 2010; see also Koptjevskaja-Tamm et al., 2007 for a review of the field). Several

of these studies focus on implicational hierarchies that become evident from the cross-linguistic study of specific semantic fields. They investigate the universal mechanisms by which lexical items can be added to the lexicon. Other studies investigate the question which words are typically basic and which words are typically derived or peripheral in the languages of the world (Goddard, 2001). Manner verbs have not been investigated in this tradition until now.

The largest amount of information on manner verbs has been gathered in the context of motion event encoding, which has been studied extensively over the last few decades. The study of motion event encoding started with the work of Talmy (1985, 1991), who showed that languages make use of two major construction types: one that encodes the path of motion on the verb (verb-framed construction) and one that encodes the path of motion outside of the verb, on so-called satellites (satellite-framed construction). The use of these two constructions for the encoding of motion events has been described for a range of different languages during the past decades. Examples from my own corpus are provided in (1) and (2). These two examples are translation equivalents: a German and a French translation of the same sentence from a parallel corpus.

- (1) German—satellite-framed construction

Und sie rannte in höchster Eile zurück zu
 and 3SG.F.NOM run.PST.3SG in high.SUP.F.SG.DAT hurry.F back to
der kleinen Tür
 ART.DEF.F.DAT little.F.SG.DAT door.F
 ‘And she ran as fast as she could back to the small door’

- (2) French—verb-framed construction

Elle repart-i-t à tout-e allure vers la petit-e
 3SG.F.SBJ return-PRET-3SG PREP all-F speed.F to ART.DEF.F small-F
porte
 door.F
 ‘She returned as fast as she could to the small door’

The path of movement is the trajectory followed by the person or object as it moves, often in relation to reference points in the environment (Talmy, 1985). The path of movement in (1) and (2) is the protagonist’s trajectory from a small table (this is established in the preceding context) to the small door. The manner of movement is the way in which the action is carried out, which includes the rate of motion (*walk, march*), the rhythm, the motor pattern (*run, swim*), the posture (*crouch, glide*), and any evaluative factors that might be involved with the movement (*sneak, flee*; Slobin, 2004: 255). In (1) and (2), the manner of movement is running or moving very fast.

The German sentence in (1) encodes the path of motion on the adverb *zurück* ‘back’, and the manner of motion on the verb *laufen* ‘run,’ as well as in the

adverbial phrase *in höchster Eile* ‘in an enormous hurry.’ The French sentence in (2) encodes the path of motion on the verb *repartir* ‘return’ and the preposition *vers* ‘to,’ and the manner of motion in the adverbial phrase *à toute allure* ‘with all speed.’ The crucial difference between the two sentences is that the French sentence encodes the most important information about the path of motion (i.e. the movement back towards a place where the protagonist had been before) on the verb, while the German sentence encodes information on path on a locative particle.

However, it is not true that French uses only the verb-framed construction, and German uses only the satellite-framed construction. In certain contexts, French does make use of manner verbs. Aske (1989) was the first to point out that verb-framed languages such as French are restricted in the usage of the satellite-framed construction for telic motion events, later called ‘boundary-crossing’ by Slobin (1997), but not for atelic motion events. A description of such an atelic motion event which contains the satellite-framed construction is given in (3). In this atelic motion event, motion does take place, but the motion is described as simply taking place within a single location, as no end state is postulated in which the person finds herself when the event is finished. This means that, in (3), the reading that the protagonists are still swimming at the end of the event is possible as well as the reading that they have reached the shore. German, on the other hand, can make use of path verbs as well as manner verbs, the former being exemplified in (4).

- (3) French
et tous nag-èr-ent jusqu-’au rivage.
 and all.PL swim-PRET-3PL until-to.ART.DEF.M shore.M
 ‘and they all swam to the shore.’

- (4) German
und Alice verfolgte mit großem Interesse wie
 and Alice.F.NOM follow.PST.3SG with great.N.SG.DAT interest.N how
die Königin zum Baum zurückkehrte
 ART.DEF.F.NOM queen.F to.ART.DEF.M.ACC tree.M return.PST.3SG
 ‘and Alice watched with great interest how the Queen returned to the tree.’

It is clear that most languages have a set of manner verbs, because they may use them in event descriptions like (3). However, there are differences in the structure and the size of the manner verb lexicon between satellite-framed and verb-framed languages (Slobin, 1997, 2005). Slobin (1997) proposes that there are two different types of manner verbs in the manner verb lexicon: “Languages seem to have a ‘two-tiered’ lexicon of manner verbs: the neutral, everyday verbs—like *walk* and *fly* and *climb*, and the more expressive or exceptional verbs—like *dash* and *swoop* and *scramble*. In S-languages [satellite-framed languages, AV],

the second tier is extensive and elaborated, making distinctions that do not play a role in the considerably smaller second tiers in V-languages [verb-framed languages, AV]” (Slobin, 1997: 459). This difference in the number of second-tier manner verbs has consequences for the overall size of the manner verb lexicon: “S-languages will have a larger and more diverse lexicon of manner verbs, in comparison with V-languages” (Slobin, 1997: 458). Manner verb lexicons are thus predicted to have different sizes, depending on the motion-encoding construction that is typically used in the languages being compared. The largest amount of variety in manner verb lexicon size is found for the second-tier manner verbs, with satellite-framed languages having a more extensive class of this type of verb.

In later work, Slobin (2003: 165 ff., 2004, 2006) is more explicit about the evolutionary status of the dependency between being satellite-framed and having a large manner verb class: “If manner is easily accessible, it will be encoded more frequently and, over time, speakers will tend to elaborate the domain in terms of semantic specificity. Consequently, learners will construct a more elaborate conceptual space for manner, allowing each new generation to continue the cycle of attention to manner” (Slobin, 2004: 252 f.). Manner is easily accessible in a language when there exists “an accessible slot for manner in the language” (Slobin, 2004: 250). Accessible slots for manner include the manner verb in satellite-framed languages, the manner verb slot in serializing languages, and the manner ideophone (amongst others). From Slobin’s (2004) discussion it is therefore possible to deduce the hypothesis that, if a language is satellite-framed, it will have a larger manner verb lexicon. Wienold (1995: 322) proposes a similar relationship between lexicalization patterns and the lexicon.

What evidence is there so far that satellite-framed languages have larger manner verb lexicons? Previous comparison of individual languages or language groups seems to yield inconclusive results. Koptjevskaja-Tamm et al. (2010: 339 ff.) point out that there are differences in the size of separate subdomains within the semantic domain of manner of motion: Slavic languages are far less elaborate in their encoding of the subdomain of aquamotion than Germanic and Romance languages, the latter being verb-framed—and therefore expected to make fewer manner distinctions rather than more. Iacobini (2009) makes a similar point when discussing manner verbs in Italian (a verb-framed language), which has far more manner verbs than would be expected due to very elaborate subdomains of manner of motion that encode rapid movement away from a source or towards a goal.

In spite of this issue, Slobin’s (2004) hypothesis seems to hold for several individual languages. Slobin (2005) compares the original and translations of *The Hobbit* (by J.R.R. Tolkien) for satellite-framed English, Dutch, German, Russian and Serbo-Croatian with verb-framed French, Italian, Portuguese, Spanish,

Hebrew and Turkish. He finds that in a single chapter, the satellite-framed languages on average use 25.6 types of manner verbs, while the verb-framed languages use 17.2 types of manner verbs. Özçalışkan and Slobin (2003) show that, in a comparison of 9 English and 9 Turkish novels, English has a larger manner verb lexicon (64 types) than Turkish (26 types). Narasimhan (2003) finds 25 types of Hindi manner verbs as translations of 40 types of English manner verbs, indicating that Hindi (a verb-framed language) has a smaller (although sizeable) manner verb lexicon. Fanego (2012) demonstrates that there has been a massive growth of the English manner verb lexicon in the transition from Old English to Late Modern English, lending support for Slobin's (2004) hypothesis on a diachronic scale.

However, a study that investigates Slobin's (2004) hypothesis on a language family-wide scale has not been carried out. Aside from Ember (1978), who looked at the impact of cultural and biological factors on the growth of basic color term systems, the factors that influence the size of semantic subfields have not been directly studied. The current paper investigates the hypothesis that languages that use the satellite-framed construction more often have larger manner verb lexicons. Using a data set on motion event encoding from 20 Indo-European languages, I show that there is some evidence for this claim.

Since the languages in this data set are all genetically related, it is necessary to make use of phylogenetic comparative methods. Mace and Pagel (1994), among others, have demonstrated that cultures and languages are not independent from one another, as they share features due to common descent and proximity. An example of such a shared feature is word order in the closely related continental Germanic languages Dutch, West Frisian and German: all three languages have SVO word order in main clauses, but SOV in embedded clauses, a pattern not found in any of the other Germanic or surrounding non-Germanic languages. In all likelihood, this pattern was not invented by each of these languages independently, but 1) was inherited from a common ancestor, 2) was borrowed from one language to the others due to their close proximity, or 3) originated in a mixture of these two processes (see also Askedal, 2006). Phylogenetic comparative methods allow one to investigate a correlation between two features, in this case the use of the satellite-framed construction and the size of the manner verb lexicon, while taking into account the phylogenetic relationships that cause these types of interdependencies between languages. These methods are presented in Section 3, after the current data set is introduced in Section 2.

2. The Data Set Used in the Current Study

The data used in this study come from a parallel corpus. In this section, I first introduce the parallel corpus in 2.1. Then, I present the coding used for the motion event constructions in 2.2.1 and discuss the use of the motion event constructions in 2.2.2. The way in which the use of the motion event constructions is aggregated is discussed in 2.2.3. The coding used for distinguishing manner verbs is the topic of 2.3.1, and a new measure for manner verb lexicon size is presented in 2.3.2.

2.1. *The Parallel Corpus*

The source of data in this study is a parallel corpus of motion event descriptions in 20 Indo-European languages: English, Dutch, German, Swedish [Germanic], Irish [Celtic], Portuguese, French, Italian, Romanian [Romance], Russian, Polish, Serbo-Croatian, Latvian, Lithuanian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Iranian], Albanian, Armenian, and Modern Greek. A parallel corpus consists of an original text and its translations into different languages (Cysouw and Wälchli, 2007). The current parallel corpus contains three novels: *Alice's Adventures in Wonderland*; *Through the Looking-Glass, and What Alice Found There* (both by Lewis Carroll); and *O alquimista* (by Paulo Coelho). These three novels were chosen because they were written in two different source languages, English and Portuguese, which are reported to represent the two different types of motion event encoding. In addition, they have been translated into a wide range of languages, which allowed for a comprehensive sample.

From these three books, a selection of descriptions of motion events was extracted. As including all descriptions of motion events found in the three novels would have resulted in a far too large data set, a balanced sample of motion event descriptions was taken. This sample was constructed taking into account the type of motion event encoding construction (see 2.2.1) and the main motion verb, providing for a large and balanced set of different motion event encoding construction types. In addition, at least one instance of each type of verb was included in order to maximize the lexical diversity in the sample.

Motion events were defined as “situations in which an animate being moves from one place to another,” following Özçaliskan and Slobin (2003: 259), although I also included inanimate entities. Each motion extract was a single sentence describing a single situation (event or activity; Berman and Slobin, 1994: 657). The majority of the motion extracts have a single main verb that describes the motion between the starting point and the endpoint of motion (as in examples (1)–(4)). Those motion event descriptions that feature more than one finite verb (of which there were very few in the original motion event

descriptions) are classified as instances of the subordinate construction or the coordinate construction (see 2.2.1).

The sample amounted to 215 sentences that describe voluntary (non-causative) motion. 24 of these sentences only encode manner and not path, and are therefore included in the data set used for the investigation of the manner verb lexicon, but not in the data set of motion event encoding constructions. For the analyses presented in this paper, I distinguish between two sets of sample sentences: the full sample of 215 sentences (of which 191 encode path) called the ‘215-sentence sample’ and a smaller sample of 132 sentences (of which 118 encode path) called the ‘132-sentence sample.’ The 215-sentence sample includes sentences from all three books, while the data in the 132-sentence sample is a subset of the 215-sentence sample and includes only sentences from *Alice’s Adventures in Wonderland* and *O alquimista*. The 132-sentence sample is available for all 20 languages discussed in this paper. The 215-sentence sample is available for 16 languages, excluding Hindi, Nepali, Persian and Albanian, because there is no translation of *Through the Looking-Glass, and What Alice Found There* available for these four languages.

All original and translated sentences were found and glossed with the help of either native speakers or language specialists. Native speakers helped to categorize each motion verb as one of the verb types (including manner verbs) that were distinguished.

2.2. Motion Event Encoding Constructions

2.2.1. Coding of Motion Event Encoding Components and Constructions

Different elements or components of motion event encoding are distinguished. These components include verbal and non-verbal elements. Coding of the different motion encoding constructions is based on the different motion event encoding components that are present in each original or translated sample sentence.

The coding of the components of motion events is as follows. The verbs are coded as either manner verbs, path verbs, deictic verbs, manner plus path verbs, non-motion verbs used as manner verbs (see Section 2.3.1), or neutral verbs. The non-verbal motion event components are path satellites and manner expressions. Path satellites include all elements that signify path, i.e., adpositions, adverbs, case markers, verbal prefixes, et cetera. I broaden Talmy’s (1985: 102) original definition of path satellite to include this wide range of elements, following the arguments of Filipović (2007: 35), Beavers et al. (2010: 337) and Croft et al. (2010: 205–206). See Table 1 for descriptions and examples of each motion event component.

Table 1. Motion event encoding components distinguished in the current paper

Nature	Component	Semantics	Examples
verbal	manner verb	indicates the way in which a person or an object moves	<i>walk, run</i>
	path verb	indicates the path or trajectory of motion	<i>enter, descend</i>
	deictic verb	indicates the path of motion as seen from a deictic center	<i>go, come</i>
	manner plus path verb	indicates both manner and path of motion	Greek <i>skarfalono</i> ‘climb up’; Persian <i>goriktan</i> ‘run away’
	non-motion manner verb	indicates manner of motion in satellite-framed construction, but is not a manner of motion verb (see 2.3.1)	<i>wriggle, flutter, rattle, crash</i>
	neutral verb	indicates none of the above	<i>move, travel, find oneself, continue</i>
non-verbal	path satellite	indicates the path or trajectory of motion	<i>back, from, into</i>
	manner expression	is an adverb or manner verb participle that signifies manner, i.e., the way in which a person or an object moves	<i>slowly, quickly, running, swimming</i>

The combination of the motion event components presented in Table 1 results in a set of motion event encoding constructions, listed in Table 2. Most of these constructions are familiar from the literature (Talmy, 1985, 1991; Croft et al., 2010). For more information on these constructions see Verkerk (2014).

2.2.2. Use of Motion Event Encoding Constructions in the Sample

A short discussion of how often each motion event encoding construction is used in the languages of the sample follows. Figure 1 provides an overview of the frequency of these various motion encoding constructions as attested in the 132-sentence sample.

Several recent studies (Slobin, 2004; Beavers et al., 2010; Croft et al., 2010) point out that languages often use a set of different constructions to encode motion events. Figure 1 supports these claims: most languages make use of

Table 2. Motion event encoding constructions distinguished in the current paper¹

Name	Components	Example
1. satellite-framed construction	manner verb + path satellite	<i>Alice ran into the forest</i>
2. verb-framed construction	path verb + manner expression	<i>Alice entered the forest running</i>
3. path only construction	path verb, no indication of manner	<i>Alice entered the forest</i>
4. manner only construction	manner verb, no indication of path	<i>Alice ran in the forest</i> (locative)
5. manner plus path verb construction	only a manner plus path verb	<i>Alice [ran+into] the forest</i> (analogous to <i>Alice fled the forest</i>)
6. deictic verb construction	deictic verb, no indication of manner	<i>Alice went into the forest</i>
7. deictic verb-framed construction	deictic verb + manner expression	<i>Alice went into the forest running</i>
8. subordinate construction	any two motion verbs, one is subordinate	<i>Alice entered to run in the forest</i>
9. coordination construction	any two motion verbs, coordinated	<i>Alice entered and ran in the forest</i>

¹) Note that a less detailed, but more conventional and purely structural definition of construction types strictly in line with Talmy's (1991) classification would be the following. The verb-framed construction would be defined as a sentence with a single main verb denoting path (the use of path satellites is optional). The satellite-framed construction would be defined as a sentence with a single main motion verb (manner verb, non-motion manner verb, deictic verb, or neutral verb, but not path verb or manner plus path verb) plus at least one path satellite. The verb-framed construction, defined thus, would collapse construction types 2, 3, and 5 as listed in Table 2. The satellite-framed construction, defined thus, would include construction type 1, as well as a part of what falls under types 6 and 7, as the use of path satellites is optional for types 6 and 7. Construction types 4, 8 and 9 in Table 2 would fall outside of this classification. It is clear that definitions of construction types vary considerably between different papers in the literature on motion event encoding. To make the current results more easily comparable to other studies, which potentially use different definitions of construction types, I will briefly report on the results of a correlation analysis between the use of motion event encoding constructions defined in this purely structural way and the size of the manner verb lexicon as well as using the more detailed list of constructions in Table 2. The results of the analyses using the purely structural definition of the motion event encoding constructions can be found in Footnote 3.

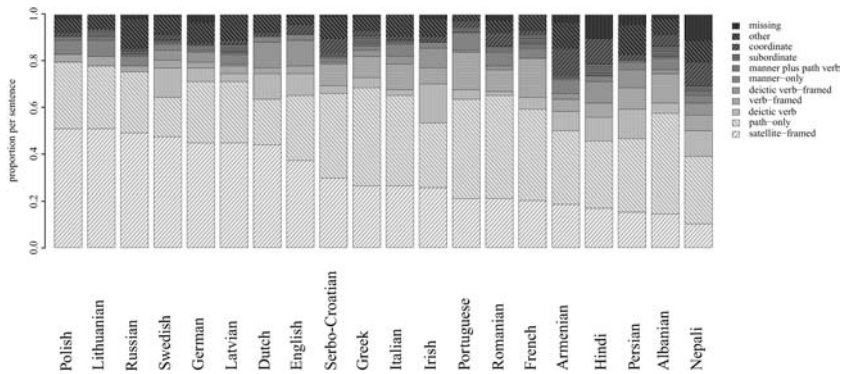


Figure 1. Proportions of the use of different motion event encoding constructions in the 132-sentence sample

several different constructions, and some languages use certain constructions more than others (see also Verkerk, 2014). However, Talmy's (1991) typology is not invalidated by this diversity, as it is possible to construct it as a distributional rather than absolute typology—which also was Talmy's (1991) aim.

The languages have been ordered in Fig. 1 in such a way that three clusters of languages emerge. The first group consists of those languages that use the satellite-framed construction often: Russian, Polish, Lithuanian, Latvian, Swedish, German, Dutch, and English. I call these languages 'satellite-framed' in the remainder of this paper, in line with convention. However, note that this does not imply that these languages do not use other strategies, or that there is no internal differentiation in this group. The second group comprises those languages that do not use the satellite-framed construction often, while they do use the path-only and the verb-framed constructions often: Italian, Portuguese, Romanian, French, Albanian, Greek, and Serbo-Croatian. In the remainder of this paper I call these languages 'verb-framed,' in line with convention but under the same caveat as above. The third 'group' consists of those languages that use neither the satellite-framed nor the path-only construction often, while they do use the deictic verb construction relatively often, as well as the coordinate and subordinate constructions in certain cases: Armenian, Persian, Hindi, Nepali, and Irish. This is not a unified group of languages, but rather one that emerges because these five languages are neither satellite-framed nor verb-framed.

The motion event encoding constructions that were not discussed above are all quite uncommon throughout the sample. The patterns found in Fig. 1 generally correspond to those that have been described in the literature for these languages (for a comparison between these results and the classifications made in the literature, see Verkerk, 2014).

2.2.3. Aggregation of the Use of Motion Event Encoding Constructions

The current data sets consist of 132 motion event encoding sentences in 20 languages and 215 sentences in 16 languages. In order to reduce the dimensionality of these data sets, two different measures were used. The first measure is simply the proportion of use of the satellite-framed construction. This measure was used because it allows for a direct test of Slobin's (2004) hypothesis. However, note that it is a continuous measure, rather than the discrete measure originally proposed by Slobin (2004), that is, the dichotomy between satellite-framed and verb-framed languages. This reflects the more recent consensus that the typology is clinal rather than absolute.

The second measure is a more holistic characterization of the use of all the motion event encoding constructions. To this end, a principal components analysis was conducted. Principal components analysis is a ubiquitous data reduction technique used in comparative analyses in biology, anthropology, and linguistics. The use of the different motion event encoding constructions appeared to be highly correlated, which allowed for the use of a principal components analysis for dimensionality reduction.

The principal components analysis was performed on the proportion of usage of each of the nine motion event encoding constructions (see Fig. 1 for a visual representation of these proportions). Instead of a regular principal components analysis, I used the phylogenetic principal components analysis developed by Revell (2009). This was done in order to remove the variance that can be attributed to the genetic relationships between the languages in the samples. The results are graphically depicted in Figs 2 and 3.

In both principal components analyses, the first and second principal components are the most important and together account for a large proportion of the variance. For the principal components analysis conducted on the 132-sentence sample (depicted in Fig. 2), the first principal component (PC1), given on the x-axis, accounts for 79.1% of the variance and can be interpreted to relate to the Talmyan scale: languages situated in the far right of Fig. 2 are the most satellite-framed, while languages situated in the far left of Fig. 2 are the most verb-framed. The second principal component (PC2), given on the y-axis, accounts for 9.8% of the variance and relates to the proportion of use of the deictic construction and the deictic verb-framed construction, with languages that use these constructions relatively often situated in the bottom of Fig. 2. For the principal components analyses conducted on the 215-sentence sample depicted in Fig. 3, PC1 explains 85.0% of the variance, while PC2 explains 6.8%. The same interpretations of the principal components apply.

The PC1 score of each language for both principal components analyses can be used to position that language on a Talmyan scale: from maximally verb-framed

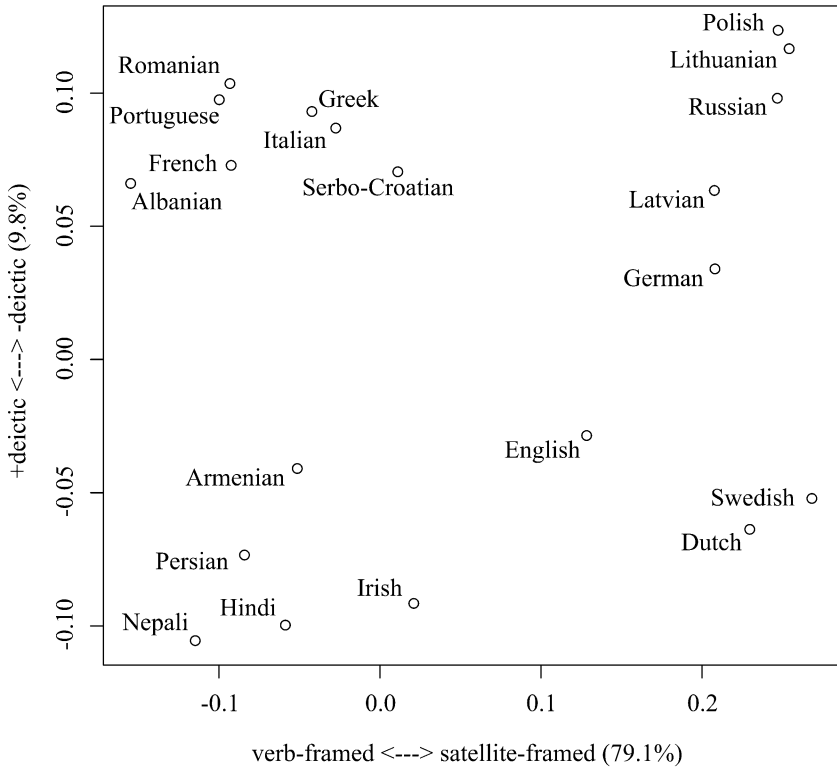


Figure 2. A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 132-sentence sample for 20 Indo-European languages

on the left side of Figs 2 and 3 to maximally satellite-framed on the right side of Figs 2 and 3. Since they provide a holistic characterization of the motion event encoding system, the PC1 scores for the 132-sentence sample and the 215-sentence sample have been used for further phylogenetic comparative analysis, reported on in Sections 3 and 4.

2.3. *Manner Verbs*

2.3.1. Coding of Manner Verbs

Although the terms ‘manner’ and ‘path’ have been in use for a long time, a cross-linguistic classification of verbs with respect to these features is not without difficulties. Many verbs seem to specify information about both manner

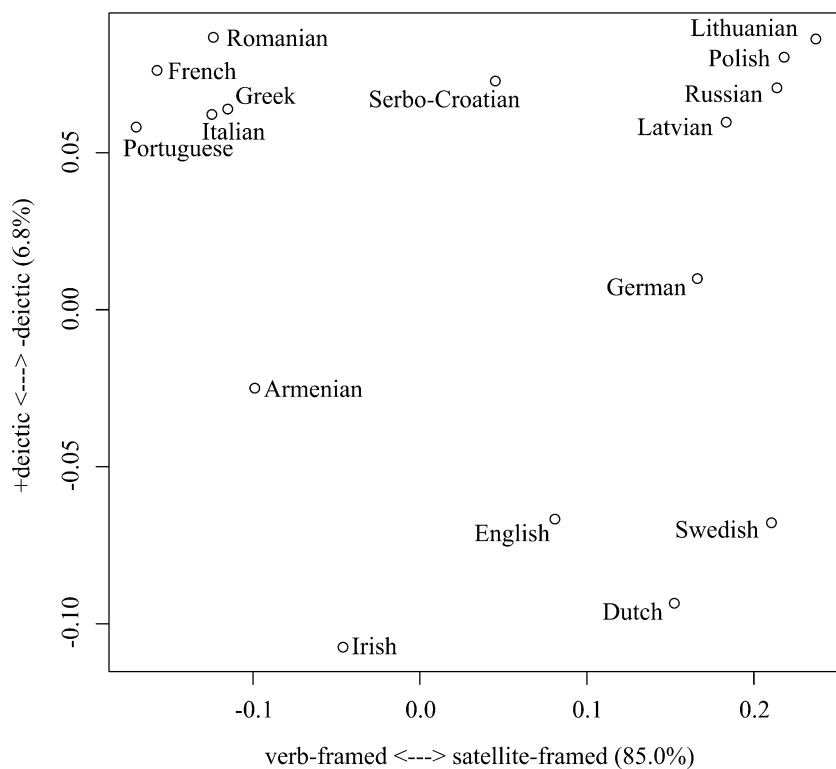


Figure 3. A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 215-sentence sample for 16 Indo-European languages

and path. English *climb* and Dutch *klimmen* are good examples: they can be used for all kinds of paths, including up, down, into and out of a reference object, but without further specification of direction, they indicate movement upwards. Another example is the Hindi verb *bbāgnā*, which may signify both ‘run’ and ‘run away.’ In other words, many manner verbs seem to have a path preference. In my coding of manner verbs, the definition of a manner verb is that it signifies the way in which a person or object moves, while it can be used with different types of path. English *climb* and Dutch *klimmen* are therefore classified as manner verbs. If a verb indicates both a clear manner and path and is incompatible with different path specifications, it is not classified as a manner verb but as a manner plus path verb.

Former studies of the size of the manner verb lexicon in different languages, such as Slobin (1997) and Narasimhan (2003), were typically based on

dictionaries, although Slobin (2005) and Hsiao (2009), among others, are exceptions. For the current study, I am using the manner verbs that appear in the restricted set of sample sentences selected from the parallel corpus. Since this set of sentences does not feature all possible contexts in which we could find manner verbs—there are, for instance, no events of driving, cycling, ice-skating, sailing, etc. in the parallel corpus—the resulting list is not a complete list of manner verbs in these languages. However, the most important ones relating to various common types of manner are included. The benefit of this approach is that it accounts for the most frequent manner verbs, while obsolete or less commonly used verbs (as one might find in dictionaries) are not included. Note that although the original *Alice's Adventures in Wonderland* and *Through the Looking-Glass, and What Alice Found There* were published in the nineteenth century, they can be considered contemporary apart from certain lexical choices; the translations have been published between 1978 and 2010. The selection of manner verbs is therefore controlled by the sample itself, and not determined by the size and the quality of dictionaries available for the languages, as was the case for some of the earlier studies.

There are two categories of verbs that are not treated as manner verbs in this study, but nevertheless deserve some attention before the actual manner verbs are discussed. These are phrasal manner verbs and non-motion verbs used as manner verbs.

Phrasal manner verbs are not included in the quite narrow definition of manner verb employed in this study. The reason for this coding decision is that phrasal verbs are often related to full verbs (*jump* and *make a jump*, for instance) and that a certain sense of arbitrariness sometimes appears to be involved in their creation. Should *make a jump* and *give a jump*, for instance, be counted as two different verbs? The phrasal verbs that were attested in the samples are listed in Appendix 2 in order to provide an overview. From these tables, it seems clear that phrasal verbs are most common in Germanic and Romance, while they are not common in any of the other languages.

With regard to the second category, it is well known that in English, verbs for various types of non-motion events can be used as motion verbs that indicate manner (Levin and Rappaport Hovav, 1991; Goldberg and Jackendoff, 2004). These include so-called 'self-contained motion verbs' such as *wriggle* (Talmy, 1991: 489), verbs of sound emission such as *crash* (Levin and Rappaport Hovav, 1991), and metaphors such as *skim* or *shoot*. This is not only true for English, but also for other satellite-framed languages. Examples from the current data sets in which sound emission verbs are used as manner verbs are provided in (5) and (6).

(5) Serbo-Croatian

Jao eno od-e njegov zlat-n-i
 Oh there go.PFV-PRS.3SG 3SG.M.POSS.M.NOM.SG gold-ADJ-M.NOM.SG
nos-ić!— Golem tiganj
 nose-DIM.M.NOM.SG huge.M.NOM.SG saucepan.M.NOM.SG
bi-o je odista
 be-PST.ACT.PTCP.M.SG be.PRS.3SG really
pro-zvižda-o mimo dijete i
 PREFIX-whistle.IPFV-PST.ACT.PTCP.M.SG next.to child.N.ACC.SG and
gotovo mu odvali-o
 almost 3SG.M.DAT bear.off.PFV-PST.ACT.PTCP.M.SG nose.M.ACC.SG
 ‘‘Oh, there goes his precious nose!’ [Alice shouted] as a huge saucepan whistled past the
 child and almost carried the nose off.’

(6) Swedish

Under tid-en roa-de sig mygga-n med att
 under time-SG.DEFUT amuse-PST REFL.3SG gnat-SG.DEFUT with to
surra runt huvud-et på henne.
 hum.INF round head-SG.DEF.N on 3SG.F.OBJ
 ‘While the gnat amused itself by humming around her head.’

Narasimhan (2003: 130 ff.) points out that, while the constructions presented in (5) and (6) are common in satellite-framed languages, they are not possible in Hindi, which is traditionally viewed as verb-framed. The current data support this observation, since translating this type of sentence directly seems troublesome in other verb-framed languages as well. The type of sentence in (5) is typically translated with a verb meaning ‘pass,’ ‘fly,’ or ‘brush’ by the verb-framed languages in the sample.

In the current analysis, only verbs that signify both manner and motion from one location to another in simple declarative sentences without path satellites, such as ‘John swam,’ are coded as manner verbs. The so-called ‘self-contained motion verbs’ such as *wriggle*, sound emission verbs such as *whistle*, and metaphors such as *skim* do not signify motion when used without path satellites: ‘Mary wriggled’ means that Mary wriggled on the spot, maybe in her seat, ‘Jack whistled’ means that Jack produced a whistling sound, and ‘Lisa skimmed the milk’ means that Lisa removed something from the surface of the milk. When they are used in constructions with path satellites, such as ‘Mary wriggled into the pipe,’ these verbs certainly denote manner of motion, but the sense of motion from one location to the other derives from the combination of the verb with a path satellite. This sense of motion is therefore not part of the semantics of these verbs themselves. Consequently, these verbs are not strictly part of the manner verb lexicon, and they are not coded as manner verbs in the current analysis.

However, since there might be interesting cross-linguistic differences in the use of these verbs, the ones that were encountered in the sample are listed in Appendix 3. These verbs are of the following types: ‘self-contained motion verbs’ (*flutter, wriggle*), sound emission verbs (*rattle, hum, splash*), verbs that are used in the subordinate construction (*struggle, have trouble to*), verbs that are used in a metaphorical sense (*skim*), and verbs that signify increasing or diminishing speed. In line with Narasimhan’s (2003) observations on the use of this type of verb in Hindi, the use of non-motion verbs appears to be more restricted in verb-framed languages than it is in satellite-framed languages. This seems to be especially the case for sound emission verbs.

A full overview of the manner verbs encountered in the sample is presented in Appendix 1. Table 3 lists the type counts for the manner verbs attested in the 132-sentence and the 215-sentence sample for each language. Appendix 4 lists the type counts for each book individually.

Table 3. Type counts for manner verbs in 20 Indo-European languages

Language	132-sentence sample	215-sentence sample
English	21	31
Dutch	23	34
German	24	31
Swedish	21	26
French	16	18
Portuguese	17	23
Italian	15	16
Romanian	16	18
Russian	19	23
Polish	25	28
Serbo-Croatian	20	27
Lithuanian	22	36
Latvian	26	33
Hindi	17	–
Nepali	14	–
Persian	13	–
Albanian	14	–
Armenian	14	14
Greek	17	20
Irish	14	21

From Table 3, Appendix 1, and Appendix 4, it becomes clear that there are differences in the size of the manner verb lexicon from language to language. In the 132-sentence sample, Persian has the smallest number of manner verbs (13 types), while Latvian has the biggest number (26 types). In the 215-sentence

sample, the variation is even larger: Armenian has 14 types of manner verbs, while Lithuanian has 36 types. Between languages that are more closely related, the numbers resemble each other more than for less closely related languages.

Appendix 1 suggests that every language in the 132-sentence sample has manner verbs for the most prototypical types of manner, i.e. words for the concepts RUN, FLY, SWIM, and WALK. RUN and WALK are also the concepts for which both satellite-framed and non-satellite-framed languages typically have more than one verb. Most languages have at least one verb that denotes STROLL alongside the general verb ‘to walk.’ Other manner types that are lexicalized in all languages include the concepts of ROLL (except Albanian), JUMP (some satellite-framed as well as some non-satellite-framed languages have more than one verb for jumping), DASH, or, more generally, very speedy movement (lexicalized by three to six different verbs in some satellite-framed and some non-satellite-framed languages), and WANDER/ROAM (except Persian). The concept of STEP/TREAD and the concept of SLIDE/SLIP are lexicalized by most satellite-framed and most non-satellite-framed languages.

Domains in which satellite-framed languages have more verbs than non-satellite-framed languages include the concept of CRAWL (only some satellite-framed languages and Greek feature more than one word for CRAWL), the concept of SCRAMBLE (only some satellite-framed languages feature a word for SCRAMBLE), the concepts of GALLOP and TROT (more satellite-framed languages have verbs denoting GALLOP and TROT than non-satellite-framed languages), the concept of FLOAT (although this is particularly so in Germanic, and not so much so in Balto-Slavic; see also Koptjevskaja-Tamm et al., 2010 for similar results), the concepts of MARCH and SNEAK, and the concept of RIDE. The concept of CLIMB is lexicalized only sporadically and mostly by satellite-framed languages; in the other languages climbing manner is typically lexicalized as a verb meaning ‘to climb up,’ having both path and manner, which would be classified as a manner plus path verb in the current coding.

Slobin (1997) claims that manner verbs can be divided into a first tier of general or neutral verbs and a second tier of more specific and expressive verbs. Satellite-framed languages have extensive second-tier manner verb lexicons, while the first-tier verbs should be more or less the same set in both non-satellite-framed and satellite-framed languages. This distinction seems to explain the patterns found in Table 3: as apparent from Appendix 1, there exists a clear set of concepts that are lexicalized into verbs in each of these languages, including WALK, STROLL, RUN, FLY, SWIM, ROLL, JUMP, RUSH/HURRY, and WANDER. The other semantic subdomains are lexicalized into verbs in some languages but not in others, and, generally, to a greater degree by satellite-framed languages than by non-satellite-framed languages. In addition, satellite-framed languages typically have more verbs in a given semantic subdomain than

non-satellite-framed languages. See, for instance, the use of six different verbs for DASH/SPEEDY MOTION in Lithuanian and Serbo-Croatian (Appendix 1c) and the use of four different verbs for JUMP in English (Appendix 1a).

In this section, I have shown that there is considerable variation in the size of the manner verb lexicon in the sample of 20 Indo-European languages. The data suggest that a core set of manner verbs is shared by all languages, and extension of the manner verb lexicon is possible in the periphery surrounding this core set. Satellite-framed languages seem to have larger manner verb lexicons than non-satellite-framed languages. In addition, closely related languages often have a similar number of manner verbs. Although a multitude of factors are likely to have played a role in the evolution of these 20 manner verb lexicons, I focus on the influence of the use of the satellite-framed construction in Sections 3 and 4.

2.3.2. Measures of Manner Verb Lexicon Size

Two different measures of the size of the manner verb lexicon are used in this study. The first is simply the type count of unique manner verbs attested in both sentence samples, as presented in Table 3. The second is what I call the Chao index, a measure of species diversity used by ecologists (Chao and Lee, 1992; Bunge and Fitzpatrick, 1993). The Chao index gives an estimate of how many species are present in a habitat, given that the sample will in all likelihood not contain a specimen of all the species actually present in the habitat. The Chao index is given in (7), with $u = (1 - \text{number of verbs occurring only once or twice} / \text{number of total attested verbs})$ and $y =$ a bias correction term that estimates the coefficient of variation of the proportion of occurrence for each verb type.

(7) *Chao index*

$$\text{true M verb lexicon size} = \frac{\text{observed M verb lexicon size}}{u} + \frac{n(1-u)}{u} y^2$$

Given that not all manner verbs are attested in the current sample, the Chao index is used as a measure to approximate the true manner verb lexicon size more closely. The Chao index was calculated using the function `ChaoLee1992` available in the R (R Development Core Team, 2011) package `SPECIES` (Wang, 2011).

2.4. Summary

Table 4 provides a short overview of the measures used for the analyses described in Section 3 below. The abbreviated designations of each measure are given in brackets.

Table 4. Samples and measures used in the current paper

Sample	Motion event encoding construction measure	Manner verb lexicon size measure
132-sentence	PC1 score (PC1)	unique verb type count (uni. count)
132-sentence	PC1 score (PC1)	Chao index
132-sentence	proportion satellite-framed (% sat)	unique verb type count (uni. count)
132-sentence	proportion satellite-framed (% sat)	Chao index
215-sentence	PC1 score (PC1)	unique verb type count (uni. count)
215-sentence	PC1 score (PC1)	Chao index
215-sentence	proportion satellite-framed (% sat)	unique verb type count (uni. count)
215-sentence	proportion satellite-framed (% sat)	Chao index

The measures used to assess motion event encoding constructions as well as manner verb lexicon size are all highly correlated. This is illustrated by a set of non-phylogenetic correlation analyses presented in Appendix 5.

3. Method

This study employs phylogenetic comparative analyses in order to investigate the potential correlation between the motion event encoding system and the size of the manner verb lexicon. This is necessary because all languages included in the current data set are genetically related to some extent. Phylogenetic comparative analyses allow investigating the relationship between these two linguistic features while taking into account the genetic relationships between the languages at the same time. In order to conduct these analyses, a sample of phylogenetic trees is needed as a representation of these genetic relationships. This tree sample is discussed in 3.1. The phylogenetic comparative analyses themselves are the topic of 3.2.

3.1. *Phylogenetic Trees*

I use a set of phylogenetic trees taken from Bouckaert et al. (2012) as a representation of the history of the Indo-European language family. Using a sample of trees rather than a single individual tree makes it possible to account for the uncertainty that is part of every phylogenetic estimation due to the different histories represented by different linguistic features. Bouckaert et al. (2012) constructed their sample of phylogenetic trees on the basis of a data set consisting of cognate-coded lexical data for 103 languages. This data set was then recoded in a binary fashion, in which each language was characterized to have a cognate (1) or not (0) for each of the 5047 cognate sets. Posterior phylogenetic tree distributions were estimated from this data set using a Bayesian Markov Chain Monte Carlo

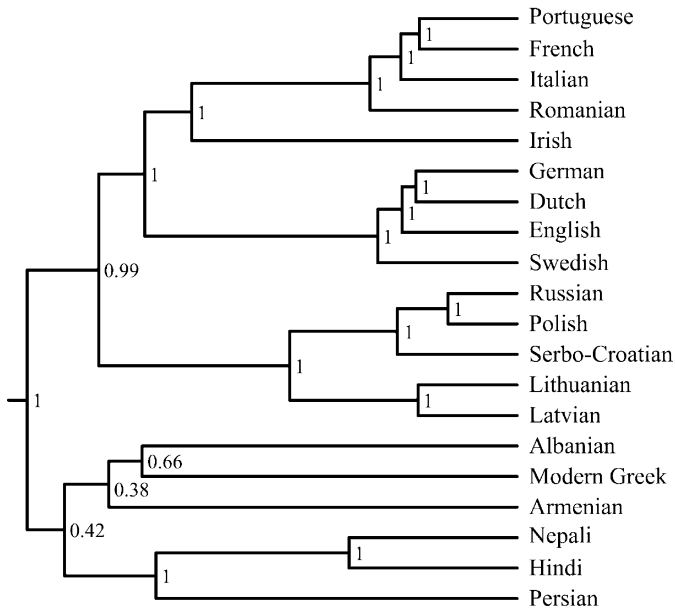


Figure 4. A maximum clade credibility tree for 20 Indo-European languages from the tree sample presented in Bouckaert et al. (2012)

approach (Huelsenbeck et al., 2001) available in the software BEAST (Drummond et al., 2012). Sampling from the posterior distribution of estimated trees generated a sample of 12,500 trees. From this sample, I randomly selected a subset of 1000 trees to use in the current study.

Retaining from this sample of trees only those languages on which the current study presents data, a sample of 1000 trees for 20 languages (for use with the 132-sentence sample) and a sample of 1000 trees for 16 languages (for use with 215-sentence sample) were created. To illustrate what these tree samples look like, a maximum clade credibility tree of the 1000-tree sample for 20 languages was calculated using TreeAnnotator v.1.6.1 (Drummond et al., 2012). This tree is presented in Fig. 4. Note, however, that the phylogenetic comparative analyses presented below were conducted over all trees in both samples of 1000 phylogenetic trees.

The numbers for the nodes of the tree presented in Fig. 4 are support-values that indicate how commonly each subgroup is attested in the tree sample. For instance, the number for the node of the subgroup consisting of Albanian, Armenian, Modern Greek and the Indo-Iranian languages (Nepali, Hindi, and Persian), 0.42, indicates that this subgroup is attested in 42% of the trees in the

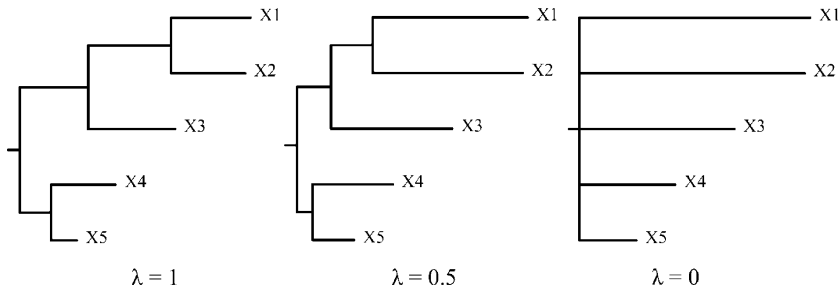


Figure 5. The effect of different values of λ on the internal branch lengths of a phylogenetic tree

1000-tree sample. The lengths of the branches of the phylogenetic trees are drawn in proportion to time. Since the current sample only includes contemporary languages, the trees are ultrametric, meaning that the lengths of the branches leading from the root node to the different languages on the tips of the tree are equal.

3.2. *Phylogenetic Comparative Analyses*

3.2.1. Phylogenetic Signal

The presence of phylogenetic signal was tested to see whether the current data on the motion event encoding system and the manner verb lexicon are dependent on genetic descent. Phylogenetic signal is the tendency of closely related species (or languages, in this case) to behave similarly, while languages that are not closely related may show differentiated behavior. For a data set of this type, it is important to know whether phylogenetic signal is present, because if it is, the use of phylogenetic comparative methods is necessary.

Pagel's lambda was used as a test of phylogenetic signal (Pagel, 1999; Freckleton et al., 2002): the estimation of a scaling parameter λ that, assuming a random walk model of evolution, adjusts tree branch lengths depending on the amount of shared history between the languages. Its most likely value was estimated using a maximum likelihood approach. In this optimization process, the tree is adjusted by giving λ different values (see Fig. 5). These values generally range between 0 and 1, but can also be slightly higher than 1, depending on features of the phylogenetic tree (Freckleton et al., 2002: 715). Values below 1 generate trees with shorter internal branches. If λ is set to zero, the internal structure of the tree is completely deleted. Figure 5 illustrates the effect of the use of λ as a scaling parameter on a phylogenetic tree.

The maximum likelihood estimation process determines the λ value that provides the best scaling transformation to the tree, given the amount of shared history evident in the data set. An estimated λ value of 1 (in other words, the maximum possible value of λ given the phylogenetic tree) signifies that the feature being studied is evolving exactly along the branches of the phylogenetic tree under a random walk model of evolution (Pagel, 1999). A λ value of 0 signifies that the feature is evolving completely independent from the phylogeny, i.e., that the feature is not dependent on history. A range of different processes aside from genetic descent could of course influence the feature, including borrowing. A λ value that is somewhere in between 0 and 1 indicates that this feature is partially dependent on history, but does not evolve exactly along the branches of the phylogenetic tree. By estimating the parameter λ for the measures used for the motion encoding system and manner verb lexicon size, it is possible to evaluate whether these measures are dependent on phylogenetic history or not.

Statistical significance of the optimized estimation of λ can be assessed with log-likelihood ratio tests (Pagel, 1997: 334). These tests determine whether λ is significantly different from a model of evolution in which λ is set to 1 and another model in which λ is set to 0. If the estimated λ is not significantly different from a model in which λ is fixed to be 1, while it is significantly different from a model in which λ is fixed to be 0, phylogenetic signal is present.

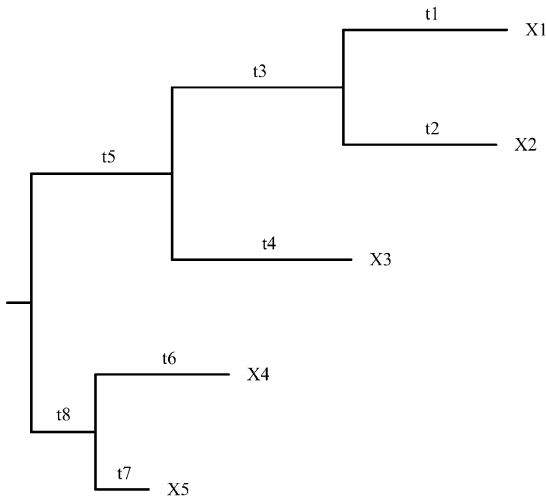
These tests were conducted using the function `physig`, part of the R package `phytools` (Revell, 2012) and the function `fitContinuous`, part of the R package `GEIGER` (Harmon et al., 2008).

3.2.2. Phylogenetic Generalized Least Squares

In order to test for a correlation between the motion event encoding system and the size of the manner verb lexicon, Phylogenetic Generalized Least Squares (PGLS) analyses were carried out (Pagel, 1997; Freckleton et al., 2002). PGLS is a linear regression analysis that takes into account phylogenetic information (Pagel, 1997: 337 ff.). In a regular linear regression analysis, the dependent variable X is the product of the estimated relation coefficient β and the independent variable Y plus residual variation e :

$$(8) \quad X = a + \beta Y + e$$

If the measures of motion event encoding and manner verb lexicon size were independent, the value β could be estimated using conventional regression techniques. However, the language data are not independent due to the shared history of the languages. Therefore the values in X and Y are interdependent, with closely related languages having more similar values as compared to less closely related languages.



	X1	X2	X3	X4	X5
X1	$t_5+t_3+t_1$	t_5+t_3	t_5	0	0
X2	t_5+t_3	$t_5+t_3+t_2$	t_5	0	0
X3	t_5	t_5	t_5+t_4	0	0
X4	0	0	0	t_8+t_6	t_8
X5	0	0	0	t_8	t_8+t_7

Figure 6. An example of a tree with a corresponding variance-covariance matrix

This interdependency can be formalized by taking into account the shared branch lengths in the phylogenetic tree. See Fig. 6 for illustration. The expected variance (variability in a character) of a feature value is proportional to the time it has been evolving, which is proportional to the length of the branch leading from the root to the tip. The expected covariance (lack of independence) of a feature value is proportional to the branch length shared between each pair of languages. In Fig. 6, X1 and X2 are closely related, which can be stated formally in terms of the shared branch length between X1 and X2, t_5+t_3 . In contrast, X1 and X3 are less closely related, which can be stated formally in terms of the shared branch length between X1 and X3, t_5 (a subset of the shared branch length between X1 and X2). A PGLS analysis first constructs a variance-covariance matrix that embodies all the information on unique and shared branch length (see Fig. 6). It then uses this variance-covariance matrix to estimate a phylogenetically adjusted value of β . This phylogenetically adjusted β describes the relationship between the motion event encoding system and the size of

the manner verb lexicon while taking into account the shared history between the languages.

The variance-covariance matrix taken from a phylogenetic tree assumes that the feature is evolving exactly along the branches of the tree under a random walk model of evolution. However, this model of evolution might not provide the best fit for the data. Therefore, the parameter λ , which measures the degree to which the phylogeny predicts the covariance between the languages, is estimated by the PGLS analysis in parallel with the construction of the covariance matrix. The most likely value of λ is estimated to modify the phylogenetic tree in such a way that it best reflects the amount of phylogenetic dependence between the variables (see 3.2.1 and Fig. 5). If $\lambda = 1$, the tree remains as it is, if $\lambda = 0$, the tree topology and branch lengths are reduced to a star phylogeny that has no phylogenetic information (the analysis then becomes a regular regression analysis, without correction for phylogeny). If λ is between 1 and 0, the tree topology and branch lengths are modified accordingly as illustrated in Fig. 5.

In short: a PGLS analysis is a maximum likelihood estimation, and therefore provides the most likely answer to the question whether two features are related while taking into account the most likely estimate of the amount of covariance that is explained by the phylogenetic tree. PGLS analyses for the present study were conducted using the function `corPagel` from the R package `ape` (Paradis et al., 2004) and the function `gls` from the R package `nlme` (Pinheiro et al., 2012).

4. Results

4.1. *Phylogenetic Signal*

First, I discuss the presence of phylogenetic signal. Table 5 gives an overview of the estimated λ values for the data measures. The estimation of λ was conducted over the tree sample of 1000 phylogenetic trees. Therefore, the median of the estimated λ values is presented in the third column of Table 5, while the range of the estimated λ values is presented in the fourth column. The last two columns give the number of trees for which λ was estimated to be significantly different from $\lambda = 1$ and $\lambda = 0$, respectively, on the $p < 0.05$ significance level.²

² The estimated λ should be tested to be significantly different from a model in which λ is set to have the maximum possible value of λ given the phylogenetic tree, not simply to be significantly different from a model in which $\lambda = 1$. The maximum possible λ values ranged from 1.090 to 1.279 for the tree sample used for the analysis with the 215-sentence sample, and 1.090 to 1.277 for the tree sample used for the analysis with the 132-sentence sample. However, these values could not serve as maximum possible λ values in the calculations, as they are not accepted by the `corPagel` function from the R package `ape` (Paradis et al., 2004). This function only accepts fixed values between 0 and 1 for λ .

Table 5. Tests for phylogenetic signal

Sample	Data measure	Median λ	Range λ	$\lambda = 1$	$\lambda = 0$
215-sentence	PC1	1.16	1.09–1.28	999	1000
132-sentence	PC1	1.16	0.95–1.28	914	1000
215-sentence	% sat	1.16	1.08–1.27	180	1000
132-sentence	% sat	1.15	0.97–1.27	4	1000
215-sentence	uni. count	1.02	0.95–1.11	0	1000
132-sentence	uni. count	0.84	0.77–0.93	9	1000
215-sentence	Chao index	1.14	1.03–1.27	0	1000
132-sentence	Chao index	0.40	0.36–0.44	1000	0

$\lambda = 1$ lists the number of trees for which λ was estimated to be significantly different from 1; $\lambda = 0$ lists the number of trees for which λ was estimated to be significantly different from 0.

As presented in Table 5, the estimated λ values for the measures used in the current analyses are generally quite high. They are always significantly different from an evolutionary model in which λ was set to 0 (except for the Chao index on the 132-sentence sample, see below), and sometimes not significantly different from a model in which λ was set to 1. This indicates that there exists a clear phylogenetic signal, and the behavior of the languages with regard to these data measures is dependent on phylogenetic history. The outlier, the Chao index for the 132-sentence sample, is by accident a rather arbitrary measure. For instance, for Armenian it is 34, while for Swedish it is 33. This is due to the high number of manner verbs occurring only once or twice in proportion to the total number of verbs for Armenian. This phenomenon occurred for several different languages in the 132-sentence sample, resulting in a relatively random Chao index for the 132-sentence sample and low estimated λ values.

4.2. *Phylogenetic Generalized Least Squares*

Let us now turn to the results from the PGLS (Phylogenetic Generalized Least Squares) analysis. A known quirk of the function `corPagel` in the R package `ape` (Paradis et al., 2004), which was used to estimate λ for the PGLS analyses, is that it can estimate λ to be smaller than zero (negative) or higher than one. Negative λ values as well as λ values much larger than 1 imply negative branch lengths, which have a debatable meaning and value in evolutionary models. I have therefore removed from the results those trees that generated negative λ values or λ values larger than 1.28. This latter value was chosen because it corresponds to the highest maximum λ value that was possible for certain trees from the tree samples (1.279 for the tree sample used for the analysis of the 215-sentence sample, and 1.277 for the tree sample used for the analysis of the 132-sentence sample). This is discussed further in Section 5.

The results of the PGLS analyses are presented in Tables 6 and 7. Table 6 shows the estimated λ values for the PGLS analyses.

Table 6. Estimated lambda values for the PGLS analyses

Sample	Indep. variable	Dep. variable	Included trees	Median λ	Range λ	$\lambda = 1$	$\lambda = 0$
215-sentence	PC1	uni. count	836	0.24	0.001–0.72	0	0
132-sentence	PC1	uni. count	0	–	–	–	–
215-sentence	% sat	uni. count	289	0.018	5.23×10^{-5} –0.13	0	0
132-sentence	% sat	uni. count	468	0.038	0.0003–0.16	1000	0
215-sentence	PC1	Chao index	1000	1.12	0.90–1.25	0	0
132-sentence	PC1	Chao index	0	–	–	–	–
215-sentence	% sat	Chao index	539	0.74	0.33–1.19	0	0
132-sentence	% sat	Chao index	994	0.26	0.19–0.33	1000	0

The column ‘Included trees’ lists the number of trees for which λ was estimated to be between 0 and 1.28, and which were therefore included in the analysis; $\lambda = 1$ lists the number of trees for which λ was estimated to be significantly different from 1; $\lambda = 0$ lists the number of trees for which λ was estimated to be significantly different from 0.

For most measures, the PGLS analysis produced λ values within the bounds of 0–1.28 for at least more than 200 out of the 1000 trees; however, this is not the case for the PC1 + unique manner verb count analysis for the 132-sentence sample and the PC1 + Chao index for the same sample. For these analyses, the estimation procedure did not arrive at viable λ values for any of the trees. These two analyses are therefore disregarded in the further discussion of the results.

The analyses on the other measures produced estimates for λ that ranged, on average, between 0.018 and 1.12, depending on the measures and the sample. This suggests that shared history explains at least some part of the covariance between the measures of motion event encoding and the measures of size of the manner verb lexicon, and the use of a linear regression that adjusts for phylogenetic relatedness is validated. The λ values for the two analyses using the proportion of use of the satellite-framed construction + unique manner verb count are low, suggesting that shared history does not play a large role for this combination of measures.

Table 7 shows the results of the PGLS analyses proper. Since these were conducted over the two 1000-tree samples, only the medians of the relevant values are given in Table 7.

Table 7. PGLS results for the 215-sentence and the 132-sentence sample

Sample	Incl. trees	Indep. variable	Dep. variable	R ²	Coefficients	Stan. error	t	P
215-sentence	836	PC1	constant uni. count	– 0.79	3.64 0.73	0.14 0.80	26.55 4.04	2.25×10^{-13} 0.0012
132-sentence	0	PC1	constant uni. count	– –	– –	– –	– –	– –
215-sentence	289	% sat	constant uni. count	– 0.88	3.82 0.64	0.12 0.12	30.62 5.52	3.15×10^{-14} 0.00007
132-sentence	468	% sat	constant uni. count	– 0.88	3.39 0.39	0.08 0.05	44.89 7.15	6.22×10^{-20} 1.17×10^{-6}
215-sentence	1000	PC1	constant Chao index	– 0.71	3.89 0.59	0.38 0.43	10.30 1.37	6.43×10^{-8} 0.19
132-sentence	0	PC1	constant Chao index	– –	– –	– –	– –	– –
215-sentence	539	% sat	constant Chao index	– 0.83	4.31 0.75	0.31 0.25	14.02 2.99	1.24×10^{-9} 0.0096
132-sentence	994	% sat	constant Chao index	– 0.86	3.94 0.38	0.23 0.16	16.84 2.43	1.83×10^{-12} 0.026

The column 'Incl. trees' lists the number of trees for which λ was estimated to be between 0 and 1.28, and which were included in the analysis; the independent variables used are the scores on the first principal component (PC1) and the proportion of use of the satellite-framed construction (% sat); the dependent variables used are the number of unique manner verbs attested in the dataset (uni. count) and the score on the Chao index (Chao index).

Table 7 can be read like the results from any regular regression analysis. The most important information in Table 7 is the value of coefficient β (given under 'coefficients') and whether it is statistically significant on a $p < 0.05$ level (given under 'P'). For 5 out of the 6 usable PGLS analyses using the different measures, there exists a statistically significant positive relationship between the motion event encoding system and the size of the manner verb lexicon:

- PC1 + unique manner verb count with the 215-sentence sample,
- proportion of use of the satellite-framed construction + unique manner verb count for both the 215-sentence and the 132-sentence sample,
- and proportion of use of the satellite-framed construction + Chao index for both the 215-sentence and the 132-sentence sample.

For the remaining one of the 6 usable PGLS analyses (PC1 + the Chao index with the 215-sentence sample), this relationship was also found, but it was not significant on a $p < 0.05$ level. This means that this data set provides some evidence for the hypothesis that, if a language has a position more to the right on the first principal component as depicted in Figs 2 and 3, that is, when it uses the satellite-framed construction more often, it will also have a more sizable manner verb lexicon. Because PGLS was used, we can say that this correlation was not just a chance result of shared history.³

5. Discussion and Conclusion

The results from the PGLS analyses indicate the existence of a positive relationship between the motion encoding system, especially the use of the satellite-framed construction, and the size of the manner verb lexicon. This correlation stands even though the genetic relationships between the languages in the data set explain part of the variance. Further inquiry will hopefully allow a more precise analysis of the evolutionary processes that underlie this correlation.

However, the results leave some space for further speculation. This is for the most part due to a lack of statistical power in the current data set, which became evident when the PGLS analyses were conducted. As explained in Section 3, the function `corPagel` in R package `ape` (Paradis et al., 2004) that was used to estimate λ for the PGLS analyses presented in Section 4 does not restrict the estimated lambda to be between 0 and the maximum possible lambda for the tree. The results of the PGLS analyses presented in Section 4 indicate that this is not ideal. Because the PGLS analyses sometimes produced debatable λ values that were lower than 0 or much larger than 1, a small part of the PGLS analyses became uninterpretable and thus unusable. For this reason, PGLS is preferably conducted by the function `pgls` in R package `caper` (Orme, 2011) or in `BayesTraits` (Pagel and Meade, 2005). However, using these methods to

³) As discussed in Footnote 1, PGLS analyses (as described in Section 3) using a purely structural definition of the verb-framed and the satellite-framed construction were also conducted, in order to maximize future comparability with studies employing potentially different coding systems. The results of these additional PGLS analyses were highly comparable to the results presented in this section. Out of the eight analyses conducted with the different samples and measures, one was uninterpretable because λ was estimated to be outside the bounds of 0–1.28 for all the trees. The remaining seven analyses indicated a relationship between the motion event encoding system and the size of the manner verb lexicon, three of which were statistically significant on the $p < 0.05$ level. This suggests that the correlation found is quite robust, as it can be detected using a less detailed coding system.

conduct PGLS resulted in λ values of 0 for all samples and measures (except for a single combination of measures). This estimate of $\lambda = 0$ had extremely large confidence intervals, indicating that in the PGLS analyses conducted in *capre* and *BayesTraits*, λ could have been anything between 0 and the maximum possible lambda given the tree. The present data set is thus too small to allow these methods to arbitrate between a model in which λ is intermediate to high (as found in Section 4) or a model in which $\lambda = 0$. This leads me to conclude that the results presented in Section 4 provide some evidence for a correlation between the motion encoding system and the size of the manner verb lexicon, as the majority of the PGLS analyses reported on in Section 4 were usable; ideally, however, a larger data set is needed to provide irrefutable evidence of the relationship between these two features.

Although the results presented in Section 4 do suggest a correlation between having a satellite-framed type of motion encoding system and having a large manner verb lexicon, they do not provide any evidence for the direction of evolutionary change. The results are simply evidence for the existence of a correlation between these two features in consideration of the phylogenetic dependencies among the languages, but they do not have any implications for the hypothesis that, diachronically, languages become satellite-framed first and then develop a larger manner verb lexicon. To test such a hypothesis using phylogenetic comparative methods, a much larger data set would be needed. Diachronic data on motion event encoding patterns and manner verb lexicons may be used to guide these analyses as well as to compare the results against the historical record. However, diachronic material on motion event encoding and manner verbs will not be available for the large majority of languages spoken today.

Even though further research is needed, this study is the first language family-wide positive confirmation of the earlier suggestions (Wienold, 1995: 322; Slobin, 2004: 252), thus far only demonstrated to hold true for English (Fanego, 2012), that there exists a relationship between motion event encoding patterns and the size of the manner verb lexicon. One of the most interesting aspects of this relationship is the evolutionary mechanisms on a population-wide scale that are involved in the emergence and continuation of manner salience. From Slobin (2006) we know that speakers of a satellite-framed language such as English demonstrate detailed conceptual understanding of manner of motion and its linguistic expression from an early age onwards. Is the common use of the satellite-framed construction one of the causes of this attention to manner, as suggested by the current results? If so, when and through which mechanisms has the satellite-framed construction become common? Is this emphasis on manner matched in other semantic domains, and does such prevalence of manner information throughout the linguistic system support the saliency of manner in the motion

domain? These are all questions of interest if the emergence and continuation of manner salience in motion event encoding are to be further understood.

These findings also open the door for the investigation of other correlations between motion event encoding constructions and aspects of the lexical items that feature in these constructions. For instance, languages with other motion event encoding construction types with high manner salience, such as serial verb constructions and manner ideophones, have not been considered here and would provide further insight into these correlations. Wienold (1995) has already demonstrated that verb-framed languages such as Japanese, Korean, and Thai may have a small manner verb lexicon, but a large lexicon of manner ideophones and adverbials. Within the larger motion event encoding system, it would be interesting to see whether similar correlations can be found between motion event encoding constructions and other semantic domains, such as the path verb lexicon, adposition and adverb lexicons, and case systems. If such correlations could be found, this would support the argument that strong interdependencies between syntactic constructions and the lexicon exist.

The relationship between syntactic constructions and the lexical classes that appear in these constructions definitely seems to be one of the mechanisms involved in the expansion or contraction of lexical classes, but it is certainly not the only one. A language that stands out in this data set is Irish, which has a rather small manner verb lexicon (14 verbs in the 132-sentence sample, 21 in the 215-sentence sample) for a language that uses the satellite-framed construction more often than any of the languages classified here as verb-framed, according to Fig. 1 above. However, Irish seems to be characterized by a rather small verb lexicon not only in the motion domain but throughout, with high degrees of polysemy. Clearly, there are other factors at play that have shaped the Irish verb lexicon, and common use of the satellite-framed construction does not necessarily guarantee the existence of a large manner verb lexicon. Another aspect that might play a role is borrowing. Wienold (1995) has described the resistance of German to borrowing path verbs; but in different conditions, satellite-framed languages may very well borrow path verbs, as is illustrated by the use of path verbs of Romance origins in English. Manner verbs seem to be borrowed very easily if a language is already satellite-framed (Fanego, 2012), which could cause an inflated picture of the observed relationship between the use of the satellite-framed construction and the size of the manner verb lexicon. In addition, we have to take into account that a lexical class such as that of manner verbs consists of quite a few different subclasses, all of which may behave differently for a variety of reasons, as was demonstrated by Koptjevskaja-Tamm et al. (2010) and Iacobini (2009). These various aspects of the increase or reduction in size of lexical classes are important for future studies in semantic and lexical typology.

A first step to carry this investigation further is to look at different language families and see whether similar correlations can be found. It is as of yet unknown how much variation language families typically exhibit with regard to motion event encoding and the size of the manner verb lexicon—in other words, how unusual the Indo-European language family is in comprising languages as dissimilar as Latvian and Nepali. For large language families, extensive data sets may be constructed that would also allow for the investigation of directional diachronic hypotheses. A further step would be to investigate implicational hierarchies of manner concepts, i.e., whether there is a universal order in which manner concepts are added to manner verb, ideophone, and adverb lexicons.

The study of semantic and lexical typology, specifically in relation to syntactic or typological domains of language, is still quite young. However, the range of unanswered questions demonstrates that it has great potential in uncovering cross-linguistic patterns that will allow us to gain a better understanding of the structure of meaning in human language.

Acknowledgments

I would like to thank the AnthroTree Workshop (supported by NSF BCS-0923791) for training in phylogenetic comparative methods; Michael Dunn for organizing the Beyond Phylogenies workshop hosted by the 45th Annual Meeting of the Societas Linguistica Europaea, which led to the publication of this special issue; the audience at the Beyond Phylogenies workshop for their attention and questions; as well as Dan Slobin, Michael Dunn, Jiyoun Choi, Katja Poellmann and two anonymous referees for valuable feedback on an earlier version of this paper.

Glosses

3	third person
ACC	accusative
ACT	active
ART	article
DAT	dative
DEF	definite
DIM	diminutive
F	feminine
INF	infinite
IPFV	imperfective
M	masculine
N	neuter

NOM	nominative
OBJ	object (pronoun)
PFV	perfective
PL	plural
POSS	possessive
PRET	preterite
PRFX	prefix
PRS	present
PST	past
PTCP	participle
REFL	reflexive
SBJ	subject (pronoun)
SG	singular
UT	uter (noun gender)

References

- Aske, Jon. 1989. Path predicates in English and Spanish: A closer look. In *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*, 1–14. Berkeley: Berkeley Linguistic Society.
- Askedal, John Ole. 2006. Some typological differences between the Modern Germanic languages in a historical and geographical perspective. *Tidsskrift for Sprogforskning* 4: 187–207.
- Beavers, John, Beth Levin, and Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46: 331–377.
- Berlin, Brent and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- Berman, Ruth A. and Dan I. Slobin (eds.). 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337: 957–960.
- Bunge, J. and M. Fitzpatrick. 1993. Estimating the number of species: A review. *Journal of the American Statistical Association* 88: 364–373.
- Chao, Anne and Shen-Ming Lee. 1992. Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* 87: 210–217.
- Croft, William, Johanna Barðdal, Willem Hollman, Violeta Sotirova, and Chiaki Taoka. 2010. Revising Talmy's typological classification of complex event constructions. In Hans C. Boas (ed.), *Contrastive Studies in Construction Grammar*, 201–235. Amsterdam: John Benjamins.
- Cysouw, Michael and Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *STUF—Sprachtypologie und Universalienforschung* 60: 95–99.

- Drummond, Alexei J., Marc A. Suchard, Dong Xie, and Andrew Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.
- Ember, Melvin. 1978. Size of color lexicon: Interaction of cultural and biological factors. *American Anthropologist* 80(2): 364–367.
- Evans, Nicholas and David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76: 546–592.
- Fanego, Teresa. 2012. Motion events in English: The emergence and diachrony of manner salience from Old English to Late Modern English. *Folia Linguistica Historica* 33: 29–85.
- Filipović, Luna. 2007. *Talking about Motion: A Crosslinguistic Investigation of Lexicalization Patterns*. Amsterdam: John Benjamins.
- Freckleton, Robert P., Paul H. Harvey, and Mark Pagel. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. *The American Naturalist* 160(6): 712–726.
- Goddard, Cliff. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5(1): 1–65.
- Goldberg, Adele and Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80: 532–568.
- Harmon, Luke J., Jason T. Weir, Chad D. Brock, Richard E. Glor, and Wendell Chalmers. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24(1): 129–131.
- Hsiao, Hui-Chen Sabrina. 2009. *Motion Event Descriptions and Manner-of-motion Verbs in Mandarin*. PhD dissertation, University at Buffalo, State University of New York.
- Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen, and Jonathan P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310–2314.
- Iacobini, Claudio. 2009. The number and use of manner verbs as a cue for typological change in the strategies of motion events encoding. In Giovanna Marotta, Alessandro Lenci, Linda Meini, and Francesco Rovai (eds.), *Space in Language: Proceedings of the Pisa International Conference*, 495–514. Pisa: Edizioni ETS.
- Kay, Paul and Luisa Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* 101(4): 743–760.
- Koptjevskaja-Tamm, Maria, Dagmar Divjak, and Ekaterina Rakhilina. 2010. Aquamotion verbs in Slavic and Germanic: A case study in lexical typology. In Victoria Hasko and Renee Perlmutter (eds.), *New Approaches to Slavic Verbs of Motion*, 315–341. Amsterdam: John Benjamins.
- Koptjevskaja-Tamm, Maria, Martine Vanhove, and Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11: 159–185.
- Levin, Beth and Malka Rappaport Hovav. 1991. Wiping the slate clean: A lexical semantic exploration. *Cognition* 41: 123–151.
- Levinson, Stephen C. and Sérgio Meira. 2003. 'Natural concepts' in the spatial topological domain—adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79: 485–516.
- Mace, Ruth and Mark Pagel. 1994. The comparative method in anthropology. *Current Anthropology* 35: 549–564.
- Majid, Asifa, Nicholas J. Enfield, and Mirjam Van Staden (eds.). 2006. *Parts of the Body: Cross-linguistic Categorisation*. *Language Sciences* 28 (special issue).

- Narasimhan, Bhuvana. 2003. Motion events and the lexicon: A case study of Hindi. *Lingua* 113: 123–160.
- Orme, David. 2011. The caper package: Comparative analysis of phylogenetics and evolution in R. Downloadable at <http://cran.r-project.org/web/packages/caper/> (accessed November 5, 2013).
- Özçalışkan, Şeyda and Dan I. Slobin. 2003. Codability effects on the expression of manner of motion in English and Turkish. In A. Sumru Özsoy, Didar Akar, Mine Nakipoğlu-Demiralp, Eser Erguvanlı-Taylan, and Ayhan Aksu-Koç (eds.), *Studies in Turkish Linguistics: Proceedings of the Tenth International Conference on Turkish Linguistics*, 259–270. Istanbul: Bogazici University Press.
- Pagel, Mark. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26: 331–348.
- . 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.
- Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Development Core Team. 2012. nlme: Linear and nonlinear mixed effects models. Downloadable at <http://cran.r-project.org/web/packages/nlme/index.html> (updated version; accessed November 5, 2013).
- Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.
- R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Downloadable at <http://www.r-project.org/> (accessed November 5, 2013).
- Revell, Liam J. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution* 63: 3258–3268.
- . 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.
- Slobin, Dan I. 1996. Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibatani and Sandra A. Thompson (eds.), *Grammatical Constructions: Their Form and Meaning*, 195–219. Oxford: Clarendon Press.
- . 1997. Mind, code, and text. In Joan Bybee, John Haiman, and Sandra A. Thompson (eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón*, 437–467. Amsterdam: John Benjamins.
- . 2003. Language and thought online: Cognitive consequences of linguistic relativity. In Deirdre Gentner and Susan Goldin-Meadow (eds.), *Advances in the Investigation of Language and Thought*, 157–191. Cambridge: MIT Press.
- . 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömquist and Ludo Verhoyen (eds.), *Relating Events in Narrative: Typological and Contextual Perspectives*, 219–257. Mahwah, NJ: Lawrence Erlbaum Associates.
- . 2005. Relating narrative events in translation. In Dorit D. Ravid and Hava B. Shyldkrot (eds.), *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman*, 115–129. Dordrecht: Kluwer.
- . 2006. What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In Maya Hickmann and Stéphane Robert (eds.), *Space in Languages: Linguistic Systems and Cognitive Categories*, 59–82. Amsterdam: John Benjamins.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timo-

- thy Shopen (ed.), *Language typology and syntactic description: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- . 1991. Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 480–519. Berkeley: Berkeley Linguistic Society.
- Verkerk, Annemarie. 2014. Where Alice fell into: Motion events in a parallel corpus. In Benedikt Szmrecsanyi and Bernhard Wälchli (eds.), *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech*, 324–354. Berlin: Walter de Gruyter.
- Viberg, Åke. 1983. The verbs of perception: A typological study. *Linguistics* 21: 123–162.
- Wang, Ji-Ping. 2011. SPECIES: An R package for species richness estimation. *Journal of Statistical Software* 40: 1–15.
- 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, and Götz Wienold (eds.), *Lexical Knowledge in the Organization of Language*, 301–340. Amsterdam: John Benjamins.

Appendices

Appendices 1 and 2 present an overview of the types of manner verbs used by the 20 languages in the sample. Appendix 3 presents an overview of the non-motion verbs used as manner verbs. Appendix 4 presents an overview of the manner verb types encountered in the three novels separately. Appendix 5 presents non-phylogenetic correlations between the different motion event encoding and manner verb lexicon size measures.

The following applies to Appendices 1–3: types of verbs that occur in the 132-sentence sample are presented in plain text. Types of verbs that occur only in the 215-sentence sample are presented in italics. This distinction is made in order for the overview of manner verbs in these 20 languages to be as extensive as possible, while also showing the size of the manner verb lexicon on the basis of the 132-sentence sample and on the basis of the 215-sentence sample. The type counts for manner verbs in each language are provided in Table 3. In the appendices, I distinguish the following categories of manner verbs (taken from Narasimhan, 2003): mode (*walk*), medium (*fly*), velocity (*rush*), attitude and display (*sneak*), contact (*slide*), course (*wander*), and vehicle (*ride*).

1. Manner Verbs

1a. Manner Verbs in Germanic

English	Dutch	German	Swedish
Manner type: mode			
crawl creep	kruipen 'crawl'	kriechen 'crawl'	krypa 'crawl'
<i>scramble</i>	<i>klauteren 'scramble'</i>	<i>klettern 'scramble'</i>	
tumble		purzeln 'tumble'	
roll	rollen 'roll'	rollen 'roll'	rulla 'roll'
climb	klimmen 'climb'		klättra 'climb'
run	rennen 'run' hollen 'run'	rennen 'run'	springa 'run' kila 'scurry'
walk	lopen 'walk, run'	gehen 'walk, go' laufen 'walk, go, run'	gå 'walk, go' promenera 'walk'
<i>saunter</i>	kuieren 'stroll' wandelen 'stroll' trekken 'hike' <i>slenteren 'saunter'</i>	schlendern 'stroll' spazieren 'stroll'	vanka 'saunter'
	huppelen 'hop' trippelen 'patter'		
<i>bobble</i>	<i>hobbelen 'bobble'</i>	hoppeln 'lollop'	<i>linka 'limp'</i>
step	<i>stappen 'step'</i>	<i>schreiten 'stride'</i> treten 'step'	
jump leap <i>spring</i> <i>bound</i>	springen 'jump' <i>hoppen 'hop'</i>	springen 'jump' <i>hüpfen 'hop'</i> <i>setzen 'jump'</i>	hoppa 'jump' <i>skutta 'leap'</i>
Manner type: animal mode			
<i>gallop</i>	<i>galloperen 'gallop'</i>	<i>sprengen 'gallop'</i>	<i>gallopera 'gallop'</i>
trot	draven 'trot'	<i>traben 'trot'</i> trotten 'trot'	trava 'trot' lunka 'trot'

English	Dutch	German	Swedish
Manner type: medium			
<i>float</i> <i>drift</i>	<i>drijven</i> 'float' <i>zweven</i> 'bover' dwarrelen 'twirl'	<i>treiben</i> 'float'	<i>sväva</i> 'bover' <i>driva</i> 'drift'
fly	vliegen 'fly'	fliegen 'fly'	flyga 'fly'
swim	zwemmen 'swim' baantjes trekken 'swim'	schwimmen 'swim' segeln 'sail'	simma 'swim'
Manner type: velocity			
hurry rush scurry <i>dart</i>	<i>zich haasten</i> 'hurry' spurten 'scurry'	eilen 'hurry' schnellen 'hurry' stürzen 'dash'	skynda 'hurry' rusa 'rush'
Manner type: attitude and display			
march	marcheren 'march' sluipen 'sneak'	marschieren 'march'	marschera 'march' smyga 'sneak, creep'
Manner type: contact			
slither slip <i>glide</i> <i>drift</i>	<i>glijden</i> 'glide' <i>schuiven</i> 'glide'	gleiten 'glide'	<i>glida</i> 'glide' ringla 'slither'
Manner type: course			
wander	<i>dwalen</i> 'wander' <i>zwerfen</i> 'roam'	wandern 'wander' umherstreifen 'wander' ziehen 'wander'	vandra 'wander'
Manner type: vehicle			
ride	rijden 'ride'	reiten 'ride'	rida 'ride'

1b. Manner Verbs in Romance

French	Portuguese	Italian	Romanian
Manner type: mode			
ramper 'crawl'	rastejar 'crawl'	strisciare 'crawl'	
			rostogoli 'tumble'
rouler 'roll'	rolar 'roll' <i>rebolar</i> 'roll'	rotolare 'roll'	
courir 'run'	correr 'run'	correre 'run'	alerga 'run'
marcher 'walk' cheminer 'walk'	andar 'walk' caminhar 'walk'	camminare ~ incamminarsi 'walk'	merge 'walk'
se promener 'stroll'	passear(-se) 'stroll'	passaggiare 'stroll'	plimba 'stroll'
			tropăi 'tramp'
<i>boitiller</i> 'limp'	<i>coxear</i> 'limp'		<i>șchiopăta</i> 'limp'
fouler 'tread'	pisar 'tread'		păși 'step'
bondir 'jump, rush' sauter 'jump'	pular 'jump' saltar 'leap'	balzare 'jump' saltare ~ saltellare ~ salterellare 'jump'	sări 'jump' topăi 'hop'
Manner type: animal mode			
	<i>galopar</i> 'gallop'	<i>galoppare</i> 'gallop'	
trotter 'trot'	trotar 'trot'	trottare ~ trotterellare 'trot'	
Manner type: medium			
<i>flotter</i> 'float'	<i>flutuar</i> 'float' <i>vogar</i> 'sail, float'		<i>pluti</i> 'float'
voler 'fly'	voar 'fly'	volare 'fly'	zbura 'fly'
nager 'swim'	nadar 'swim'	nuotare 'swim'	înota 'swim'
Manner type: velocity			
foncer 'dash' filer 'dash'	precipitar-se 'rush'	precipitarsi 'rush' filare 'dash'	zori 'rush' fugi 'hurry'

French	Portuguese	Italian	Romanian
Manner type: contact			
glisser 'slide'	esgueirar-se 'slip' escorregar 'slip' deslizár 'slide'	scivolare 'slip'	(a)luneca 'slide'
Manner type: course			
errer 'roam'	vagar 'wander'	vagare 'wander'	umbla 'wander' rătăci 'wander'
Manner type: vehicle			
	cavalgar 'ride'	cavalcare 'ride'	călări 'ride'

1c. Manner Verbs in Balto-Slavic

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
Manner type: mode				
lezt' 'crawl'	pełzać 'crawl' czołgać się 'crawl'	puzati 'crawl'	rėplioti 'crawl' <i>ropoti 'crawl'</i>	list 'crawl'
	leźć 'scramble'		ropšti 'scramble' <i>repečkenti 'scramble, crawl'</i>	<i>rausties 'scramble'</i>
	koziółkować się 'tumble'		<i>versti 'roll, tumble'</i>	kūleņot 'somersault'
katit'sja 'roll'	toczyć 'roll' <i>turlać się 'roll'</i>	kotrljati 'roll'	riedėti 'roll' <i>risti 'roll'</i>	ripot 'roll' <i>velties 'roll'</i>
			lipti 'climb' sliuogti 'climb'	kāpt 'climb' rāpties 'scramble, climb'
bežat' 'run'	biec 'run'	trčati 'run'	bėgti 'run'	skriet 'run'
iditi ~ hodit' 'walk'	iść ~ chodzić 'walk'	hodati 'walk, stroll'	eiti 'walk' vaikščioti 'walk'	iet 'walk' <i>kumurot 'walk'</i>
guljat' 'stroll'	spacerować 'stroll'	šetati 'stroll'	<i>pėdinti 'pace'</i>	pastaiģāties 'stroll' cilpot 'pace'

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
plestis' 'trudge'	<i>człapać</i> 'waddle' <i>wlec się</i> 'trudge'		<i>kėblinti</i> 'drag <i>oneself along</i> '	<i>slāt</i> 'strut'
<i>kovyljat</i> 'hobble'	tuptać 'toddle'	gegati 'waddle' <i>šepesati</i> 'hobble'	<i>slubuoti</i> 'limp'	tipināt 'toddle' <i>steberēt</i> 'hobble'
stupat' ~ stupit' 'tread'		koračati 'tread' <i>stupiti</i> 'step'	ženhti 'step' mindyti 'tread'	mīt 'tread'
prygnut' ~ prygivat' 'jump' <i>skakat</i> 'hop'	skoczyć 'jump'	skakati ~ skočiti ~ skakutati 'jump'	šokti 'jump'	lēkt ~ lēkāt ~ linkāt 'jump'
Manner type: animal mode				
	<i>cwałować</i> 'gallop'	<i>galopirati</i> 'gallop'	<i>šuoliuoti</i> 'gallop'	<i>auļot</i> 'gallop'
trusit' 'trot' rysit' 'trot'		kaskati 'trot'	risnoti 'trot' tursenti 'trot'	
Manner type: medium				
		<i>plutati</i> 'float' <i>lebdjeti</i> 'float'	<i>sklęsti</i> 'hover, <i>float</i> '	lidināties 'flutter, hover'
letet' ~ letat' 'fly'	lecieć 'fly' frunąc 'fly'	letjeti 'fly'	skristi 'fly' <i>skrieti</i> 'fly'	lidot 'fly' laisties 'fly'
plavat' ~ pl'it' 'swim'	plywać 'swim'	plivati 'swim'	plaukti 'swim'	peldēt 'swim'
				spraukties 'force'
Manner type: velocity				
mčat' ~ mčat'sja 'rush' rvanut'sja 'dash' <i>nestis</i> ' ~ <i>neslas</i> ' 'dash'	spieszyć 'hurry' pędzić 'bolt' gnać 'rush' mknąć 'zoom' smyrgnąć 'dash'	hitati 'hasten' juriti ~ jurnuti 'rush' navaliti 'rush' zuriti 'hurry' strugnuti 'dash' <i>sunuti</i> 'dart'	skubėti 'hurry' lēkti 'dash' šmurkštelėti 'dart' nerti 'rush' <i>mauti</i> 'dash' <i>lakstyti</i> 'scurry'	steigties 'hurry' traukties 'rush' mesties 'dash' drāzties 'dash' šauties 'dart'
Manner type: attitude and display				
šagat' 'march' <i>šestvovat</i> 'walk in <i>procession</i> '	maszerować 'march'			soļot 'march'

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
				ložņāt 'sneak'
Manner type: contact				
skol'zit' ~ skol'znut' 'glide'	sunąć 'glide' ślizgać się 'slip'	kliznuti ~ klizati 'slip'	slysti 'slide' čiaužti 'slide' smukti 'slip'	slīdēt 'glide'
Manner type: course				
bresti ~ brodit' 'wander, trudge'	błąkać się 'wander' wędrować 'wander'	obilaziti 'wander' lutati 'wander' urludati 'wander'	klaidžioti 'wander'	klist 'wander'
Manner type: vehicle				
chat' 'ride'	jechać 'ride'	jahati 'ride'	joti 'ride'	jāt 'ride'

1d. Manner Verbs in Indo-Iranian

Hindi	Nepali	Persian
Manner type: mode		
rengnā 'crawl'		ḵazidan 'creep'
luḥhknā 'roll'	gudnu 'roll' ladnu padnu 'roll'	ḡaltāndan ~ ḡaltidan 'roll'
Manner type: climb		
daḥṇā 'run' bhāḡnā 'run'	daudinu 'run' dagurnu 'run' badbadaudai cha 'run'	davidan 'run'
	hidnu 'walk'	rāh raftan 'walk'
ṭahalnā 'stroll' cahlakdamī karnā 'walk slowly'	lamkinu 'stride'	
		qadam gozāštan ~ zadan 'step'
kūdṇā 'jump' chalāmg lagnā 'jump'	uphranu 'jump'	jastan ~ jast zadan 'jump' jahidan 'leap' kizan 'jump'

Hindi	Nepali	Persian
Manner type: medium		
uṛnā ‘fly’	udnu ‘fly’	parridan ‘fly’ parvāz kardan ~ nemudan ‘fly’
tairnā ‘swim’	paudi khelnu ~ paudinu ‘swim’	šenā kardan ‘swim’
Manner type: velocity		
ḍubkī laganā ‘swoop’ lapaknā ‘dart’ jhapaṭnā ‘dash’	hatāra cha ‘hurry’	shetaftan ‘hurry’
Manner type: attitude and display		
mārc karnā ‘march’		
Manner type: contact		
laǰzidan ‘slide’		
Manner type: course		
bhaṭaknā ‘wander’ ghūmnā ‘roam’	bhautārinu ‘wander’ dulnu ‘wander’	
Manner type: vehicle		
haknu ‘ride’		

I.e. Manner Verbs in Albanian, Armenian, Modern Greek and Irish

Albanian	Armenian	Greek	Irish
Manner type: mode			
	soghal ‘crawl’	<i>mpoysoylo</i> ‘crawl’ sernomai ‘creep’	<i>lámbacán</i> ‘crawl’
	glorvel ‘roll’	kylo ‘roll’ strifogyrizo ‘roll, whirl’	<i>rothlaigh</i> ‘roll’ <i>rollail</i> ‘roll’ <i>sil</i> ‘trickle’
			dreap ‘climb’

Albanian	Armenian	Greek	Irish
vrapoj 'run'	vazel 'run'	trecho 'run'	rith 'run'
eci 'walk'	k'ayle 'walk'	perpato 'walk'	siúil 'walk'
shëtis 'stroll'		<i>peridiavaino 'stroll'</i>	
çapitet 'stride'		<i>alafropato 'trudge'</i>	<i>bacadail 'limp'</i>
shkel 'step'		pato 'step'	
kërcej 'jump'	ts'atkel 'jump'	pidao 'jump'	léim 'jump' <i>preab 'jump'</i>
Manner type: animal mode			
		kalpazo 'gallop'	
	vargel 'trot'		sodar 'trot'
Manner type: medium			
fluturoj 'fly'	t'rrch'el 'fly'	peto 'fly'	eitil 'fly'
notoj 'swim'	loghal 'swim'	kolympo 'swim'	snámh 'swim'
Manner type: velocity			
nxitroj 'hurry'	shtapel 'hurry'	speydo 'hurry'	deifir 'hurry'
u sulem 'rush'	slanal 'rush'	viazomai 'hurry'	brostaigh 'hurry'
	sural 'dash'	ormo 'rush'	
	khoyanal 'rush'		
Manner type: attitude and display			
u zvarrit 'drag oneself'		vadizo 'walk, march'	máirseáil 'march'
Manner type: contact			
shkas 'slide'	sahel 'glide'	glistro 'slide'	sleamhnaigh 'slide' sciorr 'slide'
Manner type: course			
sillet 'wander'	t'ap'arrel 'wander'	trigyrizo 'wander'	bheith ag fánaíocht 'wander'
baret 'wander'		planiemai 'wander'	<i>fálróid 'wander'</i>
Manner type: vehicle			
			marcaigh 'ride'

2. Phrasal Manner Verbs

2a. Phrasal Manner Verbs in Germanic

English	Dutch	German	Swedish
make one's way find one's way			
take a walk	gaan wandelen 'go for a walk'	spazieren gehen 'go for a walk'	
give a leap <i>give a bound</i>		einen Satz machen 'make a jump'	
make a rush			göra en rusning

2b. Phrasal Manner Verbs in Romance

French	Portuguese	Italian	Romanian
			da dea dura 'tumble'
		fare qualche passo 'take some steps'	
faire un bond 'jump'	dar um salto 'jump'	fare un balzo 'jump' <i>dare un balzo 'jump'</i> <i>fare un salto 'jump'</i> <i>spiccare il salto 'leap'</i>	face un salt 'jump'
			da buzna 'rush'
	deixar à deriva 'leave to drift'		
		mettersi in marcia 'march'	

2c. Phrasal Manner Verbs in Balto-Slavic

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
			galvotrūkčiais leisti 'run fast'	
idti guljat' 'go for a walk'			eiti vaikščioti 'go for a walk'	
				mest kūleņus 'summersault'

2d. Phrasal Manner Verbs in Indo-Iranian

Hindi	Nepali	Persian
	ghodā daudinu 'horse ride'	

2e. Phrasal Manner Verbs in Albanian, Armenian, Modern Greek and Irish

Albanian	Armenian	Greek	Irish
	vaz tal 'run'		
běj një kërcim 'jump'	ts'atk anel 'jump' vostyun anel 'jump'	kano ena pidima 'jump'	<i>tabhair léim</i> 'jump'
			<i>tabhair fogha</i> 'rush'

3. Non-motion Verbs Used as Manner Verbs

Manner types:

1. stationary movement verb
2. sound/light emission verb
3. verbs used as part of a subordinate construction
4. verbs used in a metaphorical sense
5. transitive motion verb

3a. Non-motion Verbs Used as Manner Verbs in Germanic

Manner type	English	Dutch	German	Swedish
1.	flutter wriggle <i>wiggle</i>	fladderen 'flutter' kronkelen 'wriggle' <i>wringen</i> 'wring'	flattern 'flutter'	fladdra 'flutter' <i>flaxa</i> 'flutter' <i>slingra</i> 'wriggle'
2.	pop rattle <i>hum</i> <i>crash</i>	kletteren 'rattle' <i>zoemen</i> 'hum'	platschen 'splash' knattern 'rattle' prasseln 'crackle' <i>summen</i> 'hum' <i>rauschen</i> 'rustle'	plaska 'splash' <i>surra</i> 'hum'
3.	help <i>struggle</i>	<i>werken</i> 'work'		<i>hjälp</i> a 'help'

Manner type	English	Dutch	German	Swedish
4.	skim	stuiven 'blow' schieten 'shoot' scheren 'shave'	stürmen 'storm'	

3b. Non-motion Verbs Used as Manner Verbs in Romance

Manner type	French	Portuguese	Italian	Romanian
1.	se tortiller 'wriggle'	contorcer-se 'wriggle' esvoaçar 'flutter'	svolazzare 'flutter'	flutura 'flutter'
2.	<i>bourdonner</i> 'hum'		<i>ronzare</i> 'hum'	pleoscăi 'splash' zbărmai 'whiz'
3.	avoir un peu de mal 'have a bit of trouble' aider 'help' <i>démener</i> 'struggle'			<i>lupta</i> 'fight'
4.	<i>se faufiler</i> 'slip through'			
5.				iuți 'go faster'

3c. Non-motion Verbs Used as Manner Verbs in Balto-Slavic

Manner type	Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
1.	kružít'sja 'flutter'	zatrzepotać 'flutter' <i>wić się</i> 'wriggle'	lepršati 'flutter' <i>đipati</i> 'frisk' <i>kobeljati</i> 'wriggle'	trenkti 'strike' <i>plasnoti</i> 'flutter' <i>pleventi</i> 'flutter' <i>klibinti</i> 'wobble'	ķepuroties 'flounder' laidelēties 'fly about' locīties 'wriggle' griezties 'wriggle' <i>ietrausties</i> 'snuggle'

Manner type	Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
2.		prysnąć 'sprinkle' rąbnać 'bang' gwizdnąć 'whistle' <i>bzykać</i> 'hum'	zviždati 'whistle' <i>zujati</i> 'hum'	švilpti 'whistle' švystelėti 'flash' <i>zvimbti</i> 'hum' <i>dardėti</i> 'crash'	
4.				šauti 'shoot'	<i>šūpot</i> 'swing'
5.	<i>medlit</i> 'slow down'				

3d. Non-motion Verbs Used as Manner Verbs in Indo-Iranian

Manner type	Hindi	Nepali	Persian
1.	pharphaṛānā 'flutter' bal khānā 'wriggle'	phyatphyat garnu 'flutter'	
2.		chapnu 'splash' chapchap awaj garnu 'make splashing sound'	
3.	pareśānī honā 'be difficult' madad karnā 'help'	samhālno 'help'	

3e. Non-motion Verbs Used as Manner Verbs in Albanian, Armenian, Modern Greek and Irish

Manner type	Albanian	Armenian	Greek	Irish
1.		p'rrvel 'loll'	fteroygizo 'flap'	<i>foluain</i> 'flutter'
2.	përplas 'bang' troket 'clatter'	t'prtal 'splash' <i>bzzal</i> 'hum'		cleatráil 'clatter' <i>tuairteáil</i> 'crash' <i>crónan</i> 'hum'
3.	kam të vështirë 'have difficulties' ndihmoj 'help'	haght 'aharel 'overcome difficulties'	dyskolevo 'have difficulty' <i>paleyo</i> 'struggle'	<i>streachailt</i> 'struggle'

Manner type	Albanian	Armenian	Greek	Irish
5.	nxitoj 'speed'			

4. *Number of Manner Verbs (Type Count) per Novel*

The numbers in bold are the numbers of manner verb types encountered in the original novels; the numbers in regular font are the numbers of manner verb types encountered in the translations.

Language	Alice's Adventures	Through the Looking-Glass	O alquimista
English	16	23	9
Dutch	17	25	11
German	18	22	9
Swedish	15	16	11
French	11	13	8
Portuguese	10	16	10
Italian	11	9	7
Romanian	9	11	11
Russian	13	18	10
Polish	17	19	11
Serbo-Croatian	17	21	9
Lithuanian	16	25	14
Latvian	22	19	10
Hindi	12	–	10
Nepali	10	–	8
Persian	10	–	6
Albanian	8	–	9
Armenian	11	12	7
Greek	12	11	9
Irish	13	17	6

5. *Correlation Analyses for the Different Motion Event Encoding and Manner Verb Lexicon Size Measures*

Measure 1	Measure 2	Pearson correlation coefficient
215-sentence PC1	215-sentence uni. count	0.78
132-sentence PC1	132-sentence uni. count	0.84
215-sentence PC1	215-sentence Chao index	0.73
132-sentence PC1	132-sentence Chao index	0.62
215-sentence % sat	215-sentence uni. count	0.79
132-sentence % sat	132-sentence uni. count	0.82
215-sentence % sat	215-sentence Chao index	0.73
132-sentence % sat	132-sentence Chao index	0.60
215-sentence uni. count	132-sentence uni. count	0.87
215-sentence Chao index	132-sentence Chao index	0.70
215-sentence uni. count	215-sentence Chao index	0.92
132-sentence uni. count	132-sentence uni. count	0.86