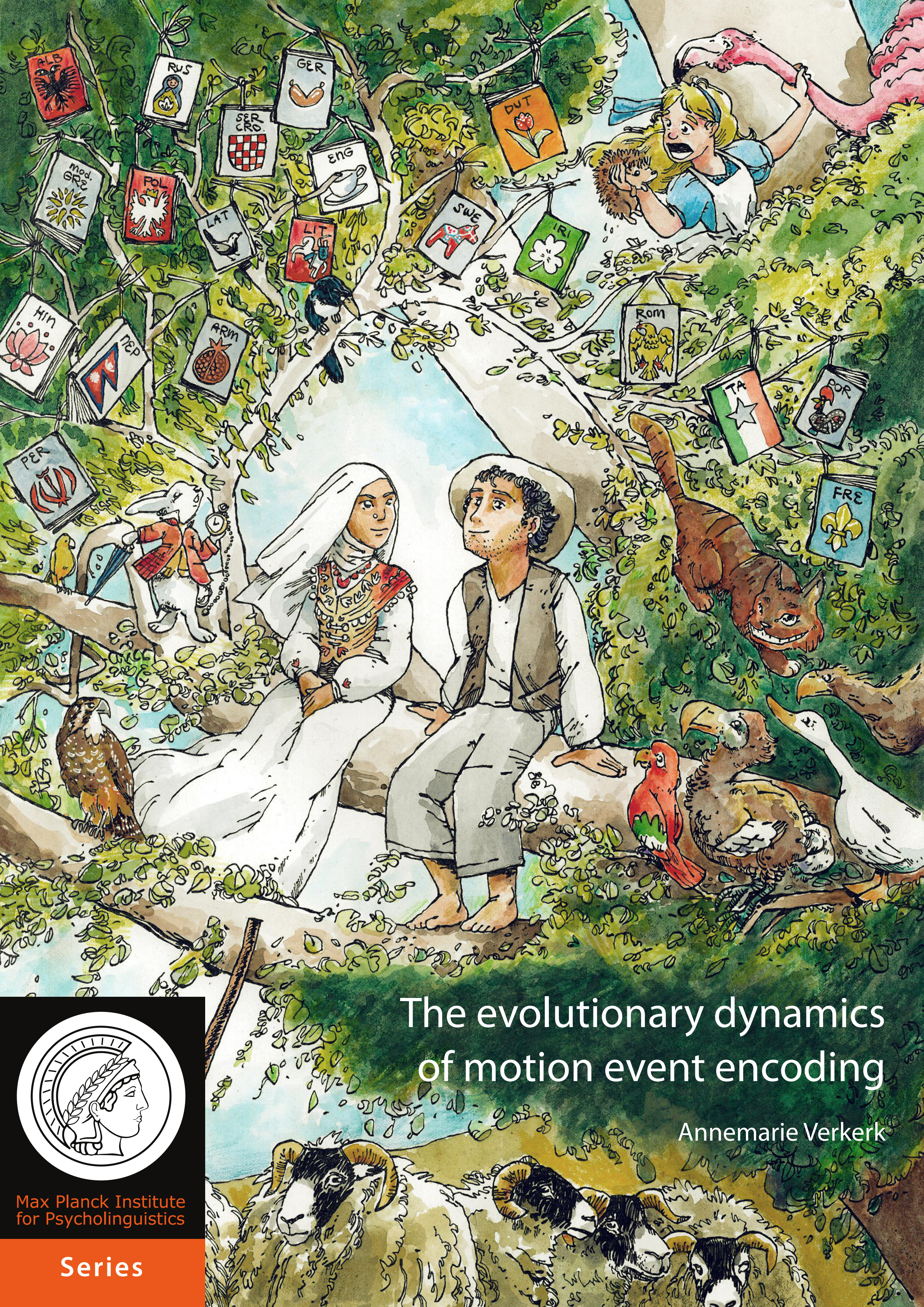


The evolutionary dynamics of motion event encoding

Annemarie Verkerk

Max Planck Institute for Psycholinguistics

Series

The evolutionary dynamics of motion event encoding

ISBN: 978-90-76203-58-4
Cover design: © Jan-Willem Middag
Printed and bound by Ipskamp Drukkers, Enschede, The Netherlands

The evolutionary dynamics of motion event encoding

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op het gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 13 juni 2014
om 16.30 uur precies

door

Annemarie Verkerk
geboren op 24 september 1985
te Oss

Promotor:
Prof. dr. Stephen C. Levinson

Copromotor:
Dr. Michael Dunn (Max Planck Institute for Psycholinguistics)

Manuscriptcommissie:
Prof. dr. Ans M.C. van Kemenade
Prof. dr. Bernhard Wälchli (Stockholm University)
Dr. Laura Fortunato (University of Oxford)

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany, as part of the Max Planck Research Group "Evolutionary Processes in Language and Culture".

Contents

Acknowledgements

This thesis could not have been completed without the help of the many people that provided inspiration, instruction, encouragement and other forms of assistance along the road. I would like to thank everybody that was a part of my PhD-life for your time, help, and kind interest. There are some people that I would like to mention in particular.

First of all, I would like to mention my supervisors, Michael Dunn and Fiona Jordan. My sincerest thanks to both of you for your belief in my abilities, for your support in all academic matters, and most importantly for being a daily inspiration. Michael, thank you for giving me the opportunity to start on this journey and for supporting me in every way possible. You let me set my own course but steered me in the right direction at the same time. Our conversations usually left me hoping that one day I might approach your level of understanding and perspective. Fiona, thank you for showing me that doing a PhD is not just about becoming the very best researcher that one can be, it is also about becoming the best organizer, time-manager, writer, speaker, reviewer and colleague that one can be. Your though questions and valuable suggestions have helped me so very much.

I am grateful to Steve Levinson for his continued interest in this project, for agreeing to be my promotor, and for his important feedback on my manuscript. I wish to thank my reading committee for their time and their constructive and invaluable comments on my manuscript: Laura Fortunato, Ans van Kemenade and Bernhard Wälchli. I feel honored that Asifa Majid, Marian Klamer, Seán Roberts and Elisabeth Norcliffe have agreed to be part of the corona. I am also immensely grateful for having Saskia van Putten and Ruti Vardi as my paranymphs. Your presence at my side fills me with joy and confidence.

Many people have been indispensable for the collection of the data on which this thesis is based. Many thanks go to Henri Ruizenaar for very generously allowing me to borrow books from his Alice collection; Gemma Capdevila from Sant Jordi Asociados Agencia Literaria for providing me with free translations of *O Alquimista*; Zeynel Abidin, Ankur Deep Bordoloi, Bijay K. Danta, and Vidyagauri Tilak for providing me with various translations, several for free; Maarten van de Heuvel for finding the Persian translation online; and Jon Lindseth and Joel Birenbaum for help with finding translations. A very special thanks goes to the MPI library for buying twenty-five translations just for this project; and specifically to Meggie Uijen for searching for translations with so much enthusiasm and good will.

A special thanks also goes to all the informants who glossed translations and helped with the analysis of the data. Their enthusiasm and hard work has been of immense importance for this thesis: Sahar Bahrami-Kh., Róise Ní Bhaoill, Eni Bilbili, Claudia Boshouwers, Dana van Breukelen, Petra Brucic, Ishan Chakrabarti, Marco Checchi, Johnny Cheung, Loig Cheveau, Alain Corbeau, Rubina Dawud, Kyriaki Dimopoulou, Marianna Donnart, Linda Eriksson, Sandra Grinberga, Paul Hulsenboom, Irina Ivanchenko, Madri Kakoti, Aruni Kashyap, Ashwini Kulkarni, Daan van Loon, Els Lucassen, Borana Lushaj, Lilla Luu, Hrach Martirosyan, Parisa Mausavi, Cyrielle Messabis, Friederike Möller, Ergin Opengin, Netra P. Paudyal, Yashoda Pokharel, Catalina Ratala, Serieke de Ruyter, Algirdas Sabaliauskas, Rita Schrama, Áine Seoighe, Guilherme Silva, Lusine Simonyan, Elżbieta Sroczyńska, and Stefanie Voigt. Many of you have become friends, for which I am grateful as well. My two student assistants, Agata Blichewicz and Maddalena Tacchetti, helped immensely with data entry, for which my heartfelt thanks.

I am also thankful to many colleagues. During this project, I have benefited greatly from discussions with John Beavers, Melissa Bowerman, Bill Croft, Michael Cysouw, Russell Gray, Simon Greenhill, Silvia Luraghi, Elisabeth Norcliffe, Loretta O'Connor, and Dan Slobin, as well as with members of audiences at conferences and participants of courses and workshops around the world. I have especially benefitted from discussing (and complaining about) motion data with Inge Alferink, Rebecca Defina, Reyhan Furman, Beyza Sümer and other members of the 2010 and 2011 motion reading groups. For meticulous proofreading I would like to thank Olga Krasnoukhova, Dan Slobin, Angela Terrill, Ruprecht von Waldenfels, as well as several anonymous referees. A special thanks goes to my writing support group buddies Jiyoun Choi, Rosanne Hebing, Vanja de Lint and Tanja Poellmann. Thanks go to Joy de Jong for valued feedback on my writing. Thanks for help with statistics go to Dan Dediu. Thanks for help with phylogenetic matters goes to Freek Bakker, Lars Chatrou, Natalie Cooper, Andrew Meade, Charlie Nunn, Chris Venditti, as well as various members of the R phylogenetics special interest group.

Thanks for valuable support from the following people is due here as well: Paul Lommen and Edith Sjoerdsma for taking care of all kinds of administrative matters, the MPI librarians for all the interlibrary loans, to Tobias van Valkenhoef for much needed Mac support, and to Dirkje van der Aa, Els den Os, and Rachel Sheer for organizing many worthwhile IMPRS courses and evenings. Thanks for help with Filemaker goes to Maarten Witberg and various people at filemakertoday.com.

After two and a half years of working on this thesis at the MPI, I went abroad to spend four months at the University of Texas at Austin. Thanks to the financial support of the University of Texas at Austin and the organizational skills of David Birdsong and Steve Levinson, I could spend the 2012 Spring semester in Tandy Warnow's lab at the Department of Computer Sciences. David, thank you for your wonderful help to make our stay in Austin as pleasant as it was. I will never forget the trips we made in your red Chevy. Tandy, thank you for teaching me about phylogenetics and for the chance to see the inner workings of another research group. Thank you to Shamsuzzoha Bayzid, Andrei Margea, Siavash Mirarab, Nam-Phuon Nguyen, and Yashoda Pokharel for making my stay in Texas such a great experience.

I would like to offer my gratitude to Mark Pagel for giving me the opportunity to continue my journey at his lab in Reading. Thank you for making me feel welcome and for allowing me work on this thesis to a far larger extent than just crossing t's and dotting i's. Thanks also to my Reading lab mates for your concern and support, and your interest in and help with this project.

I would like to thank my MPI colleagues for encouragement and comments on my work but especially for lots of fun over many, many, many cups of tea, colorful lunches, Refter dinners, and other events. First, the best roommate I could have wished for: Vishnupriya Kolipakam. Vishnu, I miss our times in 715. You have always been supporting, caring, cheerful, and most importantly, making me laugh! Thank you for the good times. Thanks go to my other Midi Planck neighbors for making the Midi Planck the best building on campus: Kate Bellamy, Dan Dediu, Anouk Diederen, Jeremy Hammond, Gertie Hoymann, Lieke Hoenselaar, Edwin van Leeuwen, Carolin Lorenz, Julia Ludewig, Dejan Matic, Saskia van Putten, Disa Sauter, Robert Van Valin, and Ruti Vardi. Thanks also go to other MPI colleagues for their support and for making the MPI such an amazing work place: Helen Buckler, Mark Dingemanse, Tyko Dirksmeyer, Sarah Dolscheid, Nick Enfield, Christine Fawcett, Claire Hill, Asifa Majid, Ger Reesink, Shakila Shayan, Gunter Senft, Marijt Witteman, and Ewelina Wnuk. And I would also like to thank other Nijmegen colleagues for making Nijmegen such a great place to study and work for the past 11 years: Margot van den Berg, Joshua Birchall, Benedicte Haraldstad Frostad, Rik van Gijn, Harald Hammarstrom, Olga Krasnoukhova, Sander Lestrade, Linda van Meel, Pieter Muysken, Neele Müller, Job Schepens, and Leon Stassen.

To my friends and family, who did not let the incomprehensibility of this project deter them from encouraging and supporting me, thank you so much. Lieve familie en vrienden, bedankt dat jullie er altijd voor me zijn, ook al begrepen jullie niet helemaal waar dit project over gaat (en dat lag natuurlijk altijd aan mij). Ik hoop dat jullie het na het lekenpraatje helemaal snappen! Erik, bedankt dat je me al deze tijd met twee benen op de grond hebt gehouden, voor al je pogingen om deze dissertatie te lezen, en voor al je praktische hulp. Jouw steun en liefde maken me erg gelukkig.

Glosses

In linguistic examples I use the Leipzig Glossing Rules (Comrie et al. 2004) as well as a few additional glosses. All are listed below.

1	first person
2	second person
3	third person
ACC	accusative
ACT	active
ADJ	adjective
ADV	adverb
ANTIC	anticausative
AOR	aorist
ART	article
AUX	auxiliary
COMPA	comparative
COND	conditional
COP	copula
DAT	dative
DEF	definite
DEM	demonstrative
DEP	dependent
DIM	diminuitive
DIR	directive
DIST	distal
ERG	ergative
EZ	ezāfe
F	feminine
FUT	future
GEN	genitive
IND	indicative
INDF	indefinite
INDR	indirect
INF	infinitive
INS	instrumental
IPFV	imperfective
LOC	locative
M	masculine
N	neuter
NEG	negation
NOM	nominative
OBJ	object (pronoun)
OBL	oblique
PART	particle
PASS	passive
PFV	perfective
PL	plural
POSS	possessive
PRET	preterite
PRFX	prefix
PROG	progressive
PROX	proximal / proximate
PRS	present
PST	past
PTCP	participle
REFL	reflexive
RES	resultative
SBJ	subject (pronoun)
SBJV	subjunctive
SG	singular
SUP	superlative
UT	uter (Swedish)
VF	verb formative

Chapter 1: Introduction

1.1 Motion events

People that speak different languages talk about motion in different ways. An example of this are these two headlines reporting the crossing of the Niagara Falls by tightrope walker Nik Wallenda, one in English and one in French:

1) Daredevil Wallenda becomes first person to walk on tightrope across Niagara Falls[1]

2) Le funambule Nik Wallenda traverse les chutes du Niagara sur un fil[2]

The English headline refers to Nik Wallenda *walking across* the Niagara Falls on a tightrope, while the French headline indicates that he *crossed* the Niagara Falls on a tightrope. The same act is linguistically encoded in different ways in these two headlines. The English headline features a verb that signifies the manner of motion, *walk*, and a preposition, *across*. The French headline only features a (transitive) verb that signifies the path of motion, *traverser* 'to cross'. The fact that Nik Wallenda walked on the tightrope is not indicated by the French headline. To a non-specialist, this difference might be odd: being able to perceive and perform movement is central to all animals including humans, and people conceptualize and talk about motion on a daily basis. Why would there exist differences in the syntax, semantics and lexicon that humans use to linguistically encode motion?

For linguists, the difference between the English and French headlines is not odd, as questions regarding the linguistic encoding of motion have been asked by a growing number of linguists over the last three decades. One of the most important findings has been that the difference between the English and French headlines given above is not due to an accidental word choice, but reflects the prevalent linguistic encoding of motion in these two languages. Two semantic aspects of motion encoding are essential to understand the difference: the manner of motion and the path of motion (Talmy 1985). The manner of motion is the way in which the person or object moves. In the example given above, the manner of motion is walking on a tight rope. The path of motion is the trajectory of the movement of the person or object. In the example given above,

[1] news article published on 15-06-2012: http://www.foxnews.com/us/2012/06/15/wallenda-begins-walks-across-niagara-falls-wire/

[2] news article published on 16-06-2012: http://www.huffingtonpost.fr/2012/06/16/nik-wallenda-funambule-traversee-chutes-niagara_n_1602489.html

the path of motion is from one side of the Niagara Falls to the other side of the Niagara Falls. Languages such as English encode the manner of motion on the verb, (e.g. *walk* in (1)), and the path of motion on a so-called 'satellite' (e.g. *across* in (1)). This type of language is therefore called 'satellite-framed' (Talmy 1991). Languages such as French encode the path of motion on the verb (e.g. *traverser* 'to cross' in (2)), and the manner of motion on an adverbial or gerund (e.g. *à petit pas* 'with small steps', which could have been added to (2)). This type of language is therefore called 'verb-framed' (Talmy 1991).

These two motion event encoding construction types, the satellite-framed construction and the verb-framed construction, as well as several others (Croft et al. 2010; Zlatev and Yangklang 2004), have been investigated in a multitude of languages. Particularly the work of Dan Slobin (Slobin 1991, 1996a, 1996b, 1997, 2000, 2003, 2004, 2005a, 2005b, 2006; Berman & Slobin 1994; Slobin & Hoiting 1994; Özçalışkan & Slobin 2003) has extended and refined Talmy's typology. Various other studies have been listed in Tables 1.1 and 1.2, which are by no means a comprehensive overview.

Table 1.1*: Motion event encoding research in Indo-European languages*

Language	**Reference**
Dutch	Slobin (2004, 2005a, 2005b); Croft et al. (2010)
English	Slobin (2004); Talmy (1985)
French	Fong and Poulin (1998); Jones (1983); Kopecka (2006, 2009a); Pourcel (2004); Pourcel and Kopecka (2005)
German	Berthele (2004, 2006); Slobin (2004)
Modern Greek	Hickmann et al. (to appear); Papafragou et al. (2006); Talmy (2007)
Hindi	Narasimhan (2003)
Icelandic	Ragnarsdóttir and Strömqvist (2004)
Italian	Folli (2008); Folli and Ramchand (2001, 2005); Iacobini and Masini (2006, 2007); Masini (2005)
Persian	Feiz (2011)
Polish	Kopecka (2009b)
Portuguese	Slobin (2005b)
Russian	Slobin (2004, 2005b)
Serbo-Croatian	Filipović (2007)
Spanish	Aske (1989); Naigles et al. (1998); Slobin (1996b)
Swedish	Ragnarsdóttir and Strömqvist (2004)

Table 1.2*: Motion event encoding research in non-Indo-European languages*

Language	**Reference**
Akan	Ameka & Essegbey (2004)
Arrernte	Wilkins (2004)
Basque	Ibarrexte-Antuñano (2003, 2004)
Mandarin Chinese	Chen and Guo (2009, 2010); Guo and Chen (2009)
Ewe	Ameka & Essegbey (2004)
Japanese	Wienold (1995)
Korean	Choi and Bowerman (1992); Oh (2009)
Thai	Zlatev and Yangklang (2004); Zlatev and David (2006)
Turkish	Özçalışkan (2009); Özçalışkan and Slobin (2003)
Tzeltal	Brown (2004)
West-Greenlandic	Engberg-Pedersen and Blytmann Trondhjem (2004)

The view that is emerging from these studies is that motion event encoding is characterized by both cross-linguistic and language internal diversity. Languages have a set of motion-independent linguistic devices at their disposal that they can use to encode motion (Beavers et al. 2010; Ibarretxe-Antuñano 2009: 410ff). These include linguistic units such as serial verb constructions, adverbials, and subordinate clauses, and it is these units that determine the encoding patterns that are available to encode motion. Preferences or markedness constraints, in turn, determine which patterns are pervasive in an individual language. Verb-framed languages can make use of the satellite-framed construction, and, vice versa, satellite-framed languages can make use of the verb-framed construction (Huang and Tanangkingsing 2005). In addition, while a motion event in which both manner and path are expressed has initially been considered to be a unified type of complex event, recent work demonstrates that the term 'motion event' can be applied to a set of related, but different types of complex events (Croft et al. 2010). Pulling these different types of complex events apart and looking at their individual encoding patterns allows for more sophisticated answer than saying that every language is typologically 'split' or 'mixed'.

Although the mixed typological nature of motion event encoding is now described for many languages, there do not exist many studies on what drives change in motion event encoding. There are some descriptions of typological change based on the comparison of ancient languages with contemporary languages (Acedo Matellán and Mateu 2008, 2010; Iacobini and Masini 2007; Kopecka 2006, 2009a; Peyraube 2006, Talmy 2007: 154) as well as descriptions of ongoing change in contemporary languages (Croft et al. 2010; Kramer 1981; Slobin 2005b). There is also a limited amount of work done on the processes of change in the motion domain and their consequences: Croft et al. (2010) present two grammaticalization pathways that lead to the emergence of the verb-framed construction, while Slobin (2000: 110, 113, 2003: 11, 2004: 252-253) proposes that satellite-framed languages have larger manner verb lexicons, which emerge over time due to the saliency of manner in these languages. However, we still know rather little about how motion event encoding changes over time and what the drivers behind these changes are.

This dissertation investigates diachronic change in motion event encoding in the Indo-European language family. In order to answer the questions on what drives change in motion event encoding, data from a sample of 20 Indo-European languages is gathered: French, Italian, Portuguese, Romanian [Romance], Irish [Celtic], Dutch, English, German, Swedish [Germanic], Latvian, Lithuanian, Polish, Russian, Serbo-Croatian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Iranian], Modern Greek [Hellenic], Albanian, and Armenian. The study of motion encoding in a single language family is an excellent way to investigate diachronic change, since it enables the investigation of changes that have taken place along the branches of the language family tree that lead from the root of the family to the contemporary languages on the tips of the tree. This investigation is carried out by focusing on four themes: language-internal diversity, diachronic change of the use of motion event encoding constructions, correlations between syntactic and lexical features of motion encoding, and rates of lexical evolution. Phylogenetic comparative methods are used to investigate these four themes, as they can model the diachronic changes that have occurred on the branches of a phylogenetic tree. Since these methods have been mostly applied by evolutionary biologists and their application within linguistics is novel, discussion of these methods and their relevance for the investigation of diachronic change in motion encoding is presented in section 1.3, after a brief introduction of the Indo-European language family in section 1.2.

1.2 The Indo-European language family

The Indo-European language family is one of the world's major language families, and Indo-European languages are among those most widely spoken around the world (Spanish: 406 million speakers; English: 335 million speakers; Hindi: 260 million speakers; Portuguese: 202 million speakers; Bengali: 193 million speakers; Russian: 162 million speakers; Ethnologue 2013). Its pre-colonial territory spreads out from Iceland in the northwest to Bangladesh in the southeast, and covers most of Europe, the Iranian plateau, and the Indian subcontinent, as depicted in Figure 1.1. The Ethnologue (2013) lists 443 Indo-European languages, the large majority of which are Indo-Iranian (312 languages). The Indo-European languages included in this study are mapped onto the expansion of the Indo-European languages in Figure 1.1.

The Indo-European language family is considered to be discovered in 1786, when Sir William Jones acknowledged and explained the relationship between Sanskrit, Greek, Latin, Gothic, Celtic and Old Persian in a speech given to the Asiatic Society (Beekes 2011: 13-14). The discovery of the Indo-European language family can be considered to be the start of comparative historical linguistics. Major discoveries that were made include the discovery of the regularity of sound changes in the 1860s, the laryngeal theory as first hinted at by De Saussure in 1878, the discovery of what later became known as Hittite in 1887, and the realization that Hittite confirms laryngeal theory by Kuryłowicz in 1935 (Beekes 2011).

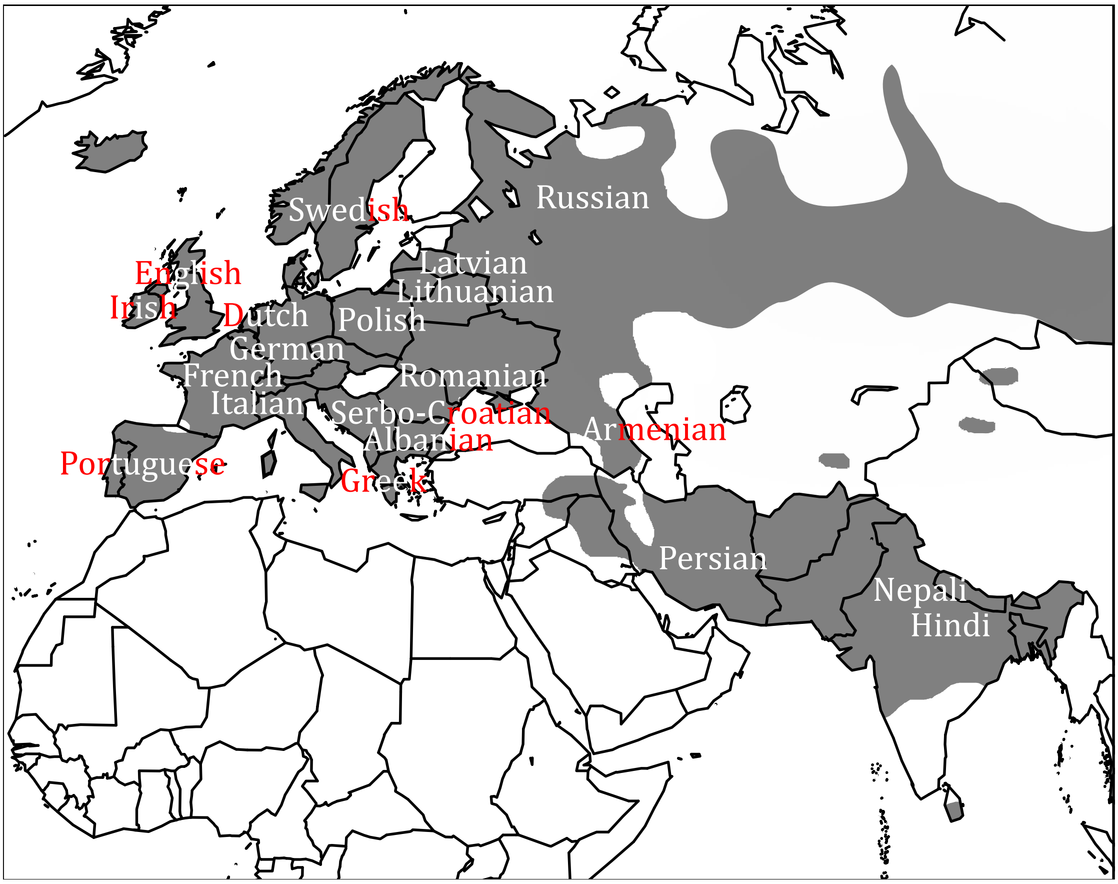


***Figure 1.1**: The spread of the Indo-European language family in Europe (based on Huffman 2013), with the languages sampled for this dissertation superimposed*

Some of the most hotly debated questions regarding the Indo-European language family center around the validity of the tree model, the suggested wider affiliations, and the subgrouping of the different subfamilies. Indo-European has 10 big subgroups: Italic (Romance), Celtic, Germanic, Balto-Slavic, Indo-Iranian, Hellenic (Greek), Anatolian, Tocharian Albanian, and Armenian. These subfamilies have been established in the early stages of study of the Indo-European family, however, the higher order subgrouping of Indo-European is still being actively debated. Most studies confirm that the Anatolian branch is the first branch to split up from the rest of Indo-European, and Tocharian is the second, but it is unclear how the remaining subgroups relate (Clackson 2007: 13; Beekes 2011: 30-31). Proposals that have been made include the unity of a Graeco-Armenian grouping (supported by Nakhleh et al. 2005a, 2005b; opposed by Clackson 1994) and a Italo-Celtic grouping (supported by Kortlandt 1981; Ringe et al. 2002; Nakhleh et al. 2005a; opposed by Watkins 1966). However, even if these higher-order subgroupings are accepted, it is not entirely clear how Italo-Celtic, Graeco-Armenian, Indo-Iranian, Germanic, Balto-Slavic, and Albanian relate, although Balto-Slavic and Indo-Iranian are often considered sisters (Nakhleh et al. 2005a). Some scholars, such as Garrett (2006), are very pessimistic about recovering the Indo-European family tree altogether.

Although the debate on the higher-order subgrouping of Indo-European is not likely to cool down in the near future, in this dissertation a set of phylogenetic trees that represent the history of the Indo-European language family will be used to study motion event encoding. This set of phylogenetic trees will be discussed in section 1.3.2.2. This approach is valid because higher-order groupings, although important, are not the only part of the trees that are relevant – in fact, for analyses that investigate correlated evolution the genealogical distance between closely related languages is more important than the distance between less closely related languages. In addition, the set of phylogenetic trees that is introduced in section 1.3.2.2 and used throughout this dissertation conforms well with other recent tree topologies (such as that of Nakhleh et al. 2005a) and therefore represents the state of the art of Indo-European tree topology.

1.3 Phylogenetic comparative methods

1.3.1 What are phylogenetic comparative methods?

Centuries of study have taught linguists that the majority of the world's languages can be placed into genealogical groupings of related languages called language families. Even though we do not know exactly how these language families relate to one another (Campbell 2008), we do know that language can be seen as a system that changes as it is passed on from generation to generation (Croft 2000; Mufwene 2001; Nettle 1999; Ritt 2004). As is the case for biological evolution, languages consist of heritable units. Biological evolution is concerned with various types of heritable units, as evolution does not only take place on the genetic level, but also on epi-genetic, behavioral, and symbolic levels (Jablonka and Lamb 2005). Likewise, language evolution takes place on various levels, resulting in heritable units on various levels: phonemes, words, syntactic constructions, and pragmatic conventions. These units are passed on from generation to generation and are subject to change over time. This process is easily observed in any pair of two closely related languages, as it is possible to reconstruct the changes that occurred since the two languages split. Even though we do not know how all the languages spoken today relate to one another, the genealogical relationships of languages within language families are relatively well studied for most languages. And most of the world's languages, almost 75%, belong to one of the ten biggest language families (Ethnologue 2013).

If the large majority of the languages spoken today are closely or more distantly related to other languages, then they cannot be considered to constitute independent data-points. This has been an issue in studies of motion event encoding in the same way as it has been an issue in all comparative or typological studies, as is illustrated by Wälchli (2009). Wälchli (2009) studies motion verb choice in motion event descriptions in five path domains: enter, exit, ascend, descend, and pass/cross. He uses a parallel corpus of translations of the *Gospel according to Mark* (a Bible text) in a world-wide sample of 117 languages. For each path domain, he makes an assessment of how often path verbs are used in the translations. Wälchli finds that about two-thirds of the languages in his sample use path verbs to encode all five path domains. This implies that the use of path verbs is the default and that satellite-framed languages, which do not encode path on their verbs, are typologically marked.

However, Wälchli of course knows that some of the languages in his sample are closely related to other languages in his sample: out of the 17 languages that do not use any path verbs for the five path domains, four are Finno-Urgic and two are Indo-European (Wälchli 2009: 215). These languages do not constitute independent data points: "Thus, instead of 15 to 17 languages with a low level of route encoding [use of path verbs, AV] in verb stems in the 109 language sample, there are rather only some 7-10 independent areas where this feature value is attested" (Wälchli 2009: 209). It seems that satellite-framed languages have only emerged independently around 7 to 10 times in the languages of the world. This suggests that the dominant use of the satellite-framed strategy is in fact more marked than was suggested by the frequency count, as the frequency count did not take into account genealogical relationships.

The notion that languages, cultures and species cannot be treated as independent from one another because of their shared history is now well-known in all disciplines that conduct comparative studies. In anthropology, this issue has come to be known as Galton's problem, as it was Galton who realized in 1889 that societies could not be seen as independent due to borrowing or common descent (Mace and Pagel 1994). Felsenstein (1985) recognized the validity of this issue for studies of evolutionary biology.

Linguistic typologists and anthropologists have dealt with this issue in the past by categorizing languages or cultures into smaller clusters, which are assumed to be independent, and then take only one culture from that cluster

(Bickel 2008; Dryer 1989, 1992; Mace and Pagel 1994; Rijkhoff and Bakker 1998). The most important problem with the application of these techniques, aside from issues relating to genealogical classification, is that typological samples that include more than 500 languages necessarily need sampling from within genealogical clusters, as the number of distinct genealogical clusters is limited (Bickel 2008). Another way to deal with non-independence is to remove the variation from the data points that is thought not to be independent. This technique has not been applied by linguistic typologists, but has been quite common in anthropology (Dow 1991). But both genealogical sampling as well as statistical approaches to remove non-independence discard information that could be valuable for comparative analysis: The first approach does not take into account the linguistic diversity evident within clusters that is relevant to understanding feature distributions, while the second approach only allows for the investigation of the small amount of variance that is left when the variance caused by shared descent and proximity is taken away (Mace and Pagel 1994). The use of phylogenetic comparative methods, as explained below, enables comparative researchers to look at different genealogical clusters and variance within those clusters at the same time (Levinson and Gray 2012).

Evolutionary biologists since the 1970s have taken the lead with the development of sound statistical methods to tackle non-independence by developing algorithms to build phylogenetic trees and use them for comparative studies (Cheverud et al. 1985; Felsenstein 1985; Harvey and Pagel 1991). These methods allow for the identification of independent instances of change of a feature on the branches of a phylogenetic tree (Mace and Pagel 1994: 550; Levinson and Gray 2012). The type of inferences that we can make with these methods are illustrated by Figure 1.2. Figure 1.2 shows a hypothetical example of change in motion event encoding that has been plotted onto a phylogenetic tree that represents the history of this set of languages. A simple count reveals that there are eight satellite-framed languages and also eight verb-framed languages. But simply counting them would over-estimate the number of independent evolutionary changes, as is revealed by the distribution of the satellite-framed and verb-framed motion event encoding systems on the phylogenetic tree. The evolution on the tree reveals that verb-framed languages emerged only once, at node X, while satellite-framed languages emerged only on two occasions, at node Y and Z. The ancestor of all languages represented on this phylogeny, A, was satellite-framed, and two of the contemporary satellite-framed languages retain this ancestral state. Satellite-framed languages do not emerge on eight occasions, but only on two (Y and Z), and verb-framed languages do not emerge eight times, but only once (X).

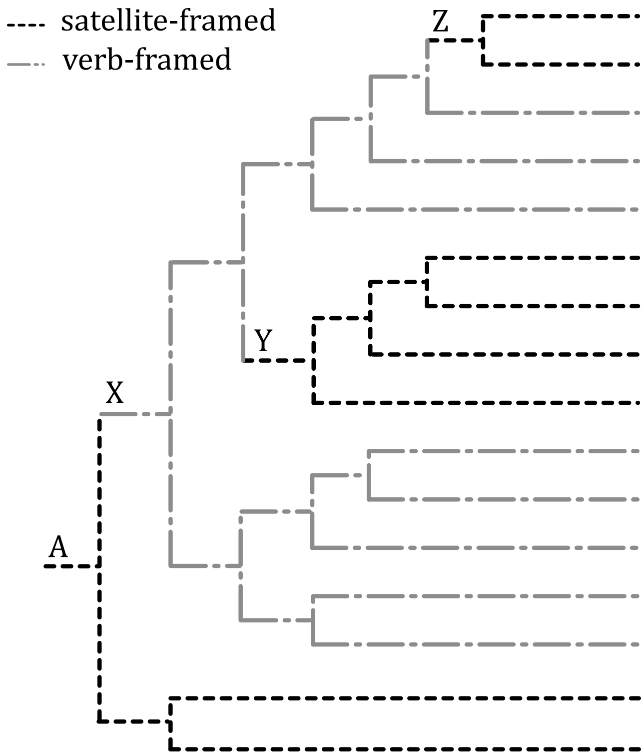


***Figure 1.2**: A hypothetical illustration of the evolution of satellite-framed and verb-framed languages on a phylogenetic tree*

Phylogenetic comparative methods, then, are a set of statistical tools that can be used to model the evolution of features like motion event encoding on a phylogenetic tree. The phylogenetic tree functions as a representation of the history of the languages in the sample that allows one to take into account the shared ancestry whilst investigating comparative questions. These questions include the type of questions that are answered by Figure 1.2: how many independent changes have occurred in the motion event encoding of this language family and what was motion event encoding like in the ancestor of these languages?

A further introduction to the types of questions that can be answered using phylogenetic comparative methods is presented in section 1.3.3, after phylogenetic trees are discussed further in section 1.3.2.

1.3.2 Phylogenetic trees

1.3.2.1 The construction of phylogenetic trees

Every phylogenetic comparative analysis is dependent on a representation of the historical relationships of the languages sample in the form of a single or a set of phylogenetic trees. In this section, the general principles of phylogenetic tree inference are explained.

The data that is used to generate phylogenetic trees can belong to different heritable units such as phonemes, lexical items, or syntactic constructions. The most common type of data used in historical linguistics is cognate-coded lexical data (Dunn et al. to appear), but typological characteristics or structural data have also been used (Dunn et al. 2005; Saunders 2005; Wichmann and Saunders 2007) as well as phonological similarities (Brown et al. 2008). Here, the process of building trees is illustrated by using an example dataset of cognate-coded lexical data. Such data is generated by finding cross-linguistic lexical data for a list of meanings and then coding this data for cognacy. Cognates are words from different languages that each language has inherited from a common ancestor. Some examples are provided in Table 1.3.

Table 1.3*: A sample cognate-coded dataset for five meanings across five Indo-European languages*[a]

Meanings	**crawl**	**run**	**wheel**	**high**	**bottom**
French	ramper[1]	courir[1]	roue[1]	haut[1]	fond[1]
Dutch	kruipen[2]	rennen[2]	wiel[2]	hoog[2]	bodem[1]
Russian	presmykay'sja[4]	bežat'[4]	koleso[2]	vysakij[3]	dno[3]
Lithuanian	lįsti[5]	bėgti[4]	ratas[1]	aukštas[5]	dugnas[3]
Irish	snámh[3]	rethim[3]	roth[1]	uasal[3], ard[4]	bun[2]

[a]Numbers in superscript indicate cognate sets for each meaning, taken from Buck (1949).

As is evident from Table 1.3, the cognates found for a single meaning say something about the history of the five languages involved, but different meanings can tell different stories. For instance, the cognates found for the meaning 'high' indicate that Irish and Russian have some shared history as opposed to French, Dutch, and Lithuanian. On the other hand, the cognates found for the item 'bottom' reveal that French and Dutch are closely related, and Russian and Lithuanian are closely related, while Irish is separate from both groups. These two different histories have been plotted in Figure 1.3.

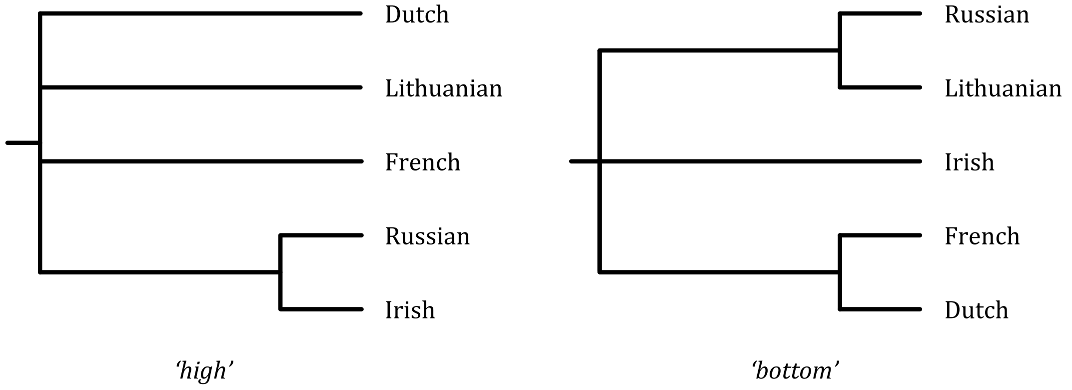


Figure 1.3: *Two phylogenetic trees representing cognate histories for different meanings.*

Phylogenetic tree building methods take large lists of meanings into account. Typically, Swadesh lists are used (Swadesh 1952, 1955). These are lists of meanings that are supposed to be very stable, very unlikely to change rapidly, and to be resistant to borrowing. Meanings of this type are selected because these have the highest change of recovering the genealogical or 'vertical' historical signal, rather than grouping languages together on the bases of borrowed or 'horizontally' transmitted similarities. Examples of words on Swadesh lists are words for body parts, substances, and natural objects, highly frequent verbs, and kinship terms.

As was evident from Table 1.3, different meanings may tell different evolutionary stories. Computers are used to find the phylogenetic tree that represents these evolutionary stories in the most optimal way, as this cannot be computed by hand even for a moderate number of languages. The most important reason to use computers for phylogenetic inference is that the possible number of phylogenetic trees for a given number of languages becomes extremely large very quickly: the number of possible unrooted trees for twenty languages exceeds Avogadro's Number (the number of atoms in twelve gram of pure carbon-12, 6.022×10^{23}, as noted by Felsenstein 1982). Phylogenetic inference therefore does not aim to calculate all possible phylogenetic trees and then select the most optimal one from the complete set of possibilities, but rather to comprehensively search the space of all possible phylogenetic trees and find the most likely tree.

There are two ways in which cognate data may be coded so that it can be used for phylogenetic inference. It can be coded as in Table 1.3, with each meaning having a variable number of character states that represent the different cognate classes. In Table 1.3, the meaning 'bottom' would have three possible states, while the meanings 'high' and 'crawl' would have five. This is called multistate coding. Multistate cognate-coded lexical data can also be transformed into a set of binary codes that represents the absence or presence of a cognate for each cognate set in each language (as done by Gray and Atkinson 2003; Bouckaert et al. 2012; and many others). This is called binary coding. A conversion of Table 1.3 into such a binary matrix is presented in Table 1.4. Atkinson & Gray (2006: 93-94) discuss some of the benefits and downsides of both multistate and binary coding. On the one hand, it could be argued that multistate cognate coding captures change within meanings, which can be seen as the fundamental units in which change takes place. The evolutionary models use account binary coding do not capture change of the cognate sets for a specific meaning in the same way as evolutionary models of multistate cognate data do. However, from a computational perspective, the analysis of binary cognate data is considerably easier than the analysis of multistate data, as the number of parameters that is required to model the process of evolutionary change is significantly lower.

Table 1.4*: A binary version of the sample cognate-coded dataset presented in Table 1.3*[a]

lang.	**crawl 1**	**crawl 2**	**crawl 3**	**crawl 4**	**crawl 5**	**run 1**	**run 2**	**run 3**	**run 4**	**wheel 1**	**wheel 2**	**high 1**	**high 2**	**high 3**	**high 4**	**high 5**	**bottom 1**	**bottom 2**	**bottom 3**
Fre.	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0
Dut.	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
Irish	0	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0	1	0
Rus.	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1
Lith.	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0	1

[a] '1' indicates presence of a cognate form in a cognate set for an individual language, '0' indicates absence. The columns represent different cognate sets for different meanings.

There are two general ways of inferring phylogenetic trees: character-based methods and distance-based methods. Distance-based methods convert matrices of the type presented in Tables 1.3 and 1.4 into distance matrices. Distance-based methods calculate distances between each pair of languages based on the proportion of shared cognates. Pairs of languages that have the highest proportion of shared cognates will be grouped together first, after which these groups will be connected to languages with which they share less and less cognates until all languages have been placed in the phylogenetic tree. Distance-based methods do not model change on an evolutionary pathway in the way character-based methods do, but rather take 'as the crow flies' distances between languages. Phylogenetic trees that are built using distance methods are therefore not based on a model of (cognate) evolution. These methods group languages together on the basis of similarity, but similarity does not necessarily imply a close genealogical relationship. Character-based methods directly use matrices of the type presented in Table 1.4 to infer phylogenetic trees. To do this they use algorithms that model the loss and gain of cognates. Character-based methods aim to find the most optimal scenario of gain and loss within each cognate class on the branches of a phylogenetic tree. In order to discover the true evolutionary history as a group of related languages, character-based methods are preferable to distance-based methods.

A commonly used character-based method to infer phylogenetic trees is maximum likelihood. Maximum likelihood methods explicitly attempt to find the tree and the model parameters that maximize the probability of producing the observed data set, given a certain model of evolution (Felsenstein 1981; Nichols and Warnow 2008: 774ff; Pagel and Meade 2005). The maximum likelihood algorithm describes the likelihood that a certain evolutionary process has given rise to the observed data as opposed to another process generating the observed data. For the cognate-coded lexical dataset in Table 1.4, the model of evolution is a model of cognate evolution. For each cognate set in Table 1.4, each language either has a cognate ('1') or it does not have a cognate ('0'). The types of evolutionary change that can take place, therefore, are very simple:

3) q_{10}: 1 → 0
q_{01}: 1 ← 0

A language can either lose a reflex of a cognate set (1 → 0) or a new lexical item can come into existence (1 ← 0). In evolutionary models used for the inference of linguistic phylogenetic trees, back mutation typically is not allowed. This means that languages cannot gain a member of a cognate set in any other way than inheriting it from an ancestral language, so each new lexical item that is introduced gives rise to a new cognate set. The stochastic Dollo model that is often used to model cognate evolution, for instance, restraints cognate sets to emerge only once (Nicholls and Gray 2008). The rates at which a reflex of a cognate is lost or a new cognate set is gained are denoted by q_{10} and q_{01}, respectively. The probability of each inferred phylogenetic tree is entirely dependent on the rates of change q_{10} and q_{01} along all of the individual segments of the tree. The probability of a change in each specific cognate set and in each individual segment of the tree are assumed to be independent from one another.

For the first cognate set in 'bottom', French and Dutch had a cognate in that set (*fond* and *bodem*), while Irish, Russian and Lithuanian did not have a cognate belong to that set (*bun*, *dno* and *dugnas*). Just using this single cognate set, two possible tree solutions with historical inferences on the presence (1) and absence (0) of a cognate in this set are depicted in Figure 1.4.

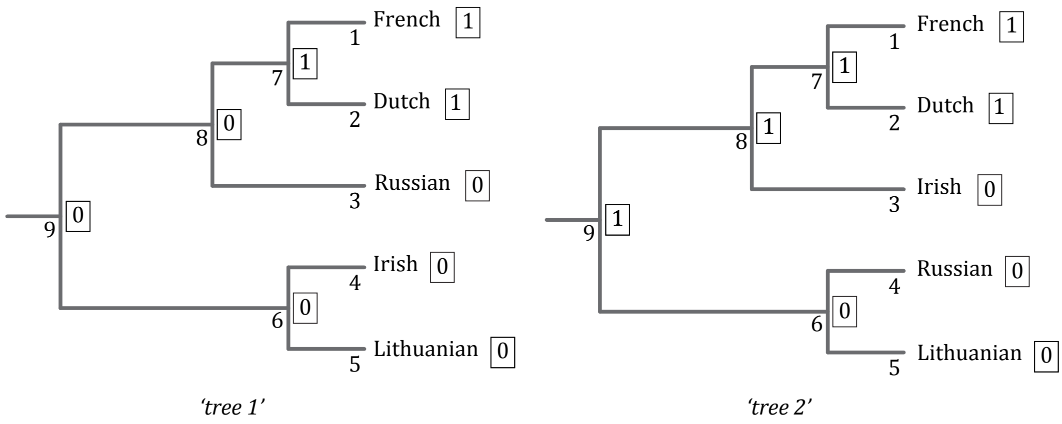


Figure 1.4*: Two possible phylogenetic trees for bottom*[1]

We can calculate the different probabilities for both trees as follows. There are three unknown factors in both trees: the state of the root (absence of cognate ‘0’ or presence of cognate ‘1’), the state of the internal nodes of the tree (0 or 1), and the probabilities of change on each segment of the tree. Tree segments are parts of branches that are intermediate between internal nodes or between internal nodes and tip nodes. These three unknown factors are all assigned probabilities (p), which are then multiplied to generate the likelihood of that particular tree. Examples for the two trees in Figure 1.4 are given in (4).

4) tree 1: $p(0)9 \times p(0)8 \times p(1)7 \times p(0)6 \times p(0 \rightarrow 0)96 \times p(0 \rightarrow 0)98 \times p(0 \rightarrow 0)_{83} \times p(0 \rightarrow 1)_{87} \times p(1 \rightarrow 1)_{72} \times p(1 \rightarrow 1)_{71} \times p(0 \rightarrow 0)_{64} \times p(0 \rightarrow 0)_{65}$

tree 2: $p(1)_9 \times p(1)_8 \times p(1)_7 \times p(0)_6 \times p(1 \rightarrow 0)_{96} \times p(1 \rightarrow 1)_{98} \times p(1 \rightarrow 0)_{83} \times p(1 \rightarrow 1)_{87} \times p(1 \rightarrow 1)_{72} \times p(1 \rightarrow 1)_{71} \times p(0 \rightarrow 0)_{64} \times p(0 \rightarrow 0)_{64}$

The probabilities of change $p(1 \rightarrow 0)$ and $p(0 \rightarrow 1)$ are described using Markov processes. This implies that the probability of a change of state is determined by the state (presence or absence of a cognate) at the beginning of the tree segment, and not by the presence or absence of a cognate in past history, i.e. earlier tree segments (Felsenstein 1981: 371). In addition, the probabilities of change are calculated for each individual tree segment, allowing for varying rates of change in different parts of the tree and for different cognate sets. The development of evolutionary models with a relaxed clock means that it is no longer necessary to assume a constant rate of evolution (Drummond et al. 2006).

Calculations as those in (4) are carried out on the complete cognate matrix to assess the likelihood of different phylogenetic trees, with different topologies and different branch lengths. These trees all have different likelihoods. This process continues until the likelihood cannot be improved.

Branch lengths in such trees are a reflection of the accumulated changes in each of the cognate sets used to estimate it. If two languages are connected by longer branches than two other languages, this indicates greater divergence between languages (Pagel 2000: 197; Pagel and Meade 2005: 239; see also Holden and Mace 2003: 2431). Branch length can represent the amount of time, i.e. the number of substitutions per site per year, if information on 'fossils', which in linguistics take the form of ancient languages such as Latin or Sanskrit, can be used to date certain internal nodes of the phylogenetic tree.

A popular and practical implementation of phylogenetic tree inference using Maximum Likelihood methods are the so-called Bayesian approaches. Given the incredibly large number of possible rooted phylogenetic trees for the twenty Indo-European languages that form the current sample (8.2×10^{21}), we can use the principles of Bayesian inference to explore the large universe of parameter combinations and find the most likely set of trees. Bayesian approaches estimate the probability that each possible tree is the true tree and therefore they do not produce a single tree, but a posterior probability distribution on the set of possible trees (Nichols and Warnow 2008: 774ff; Pagel and Meade 2004, 2005). In order to do this, the analysis requires a prior probability distribution that is informative with regard to all the parameters involved in tree building, for instance rates of change and variance in rates of change across different sites (Ronquist et al. 2009). One can then theoretically obtain the posterior distribution that specifies the probability of each tree, given the prior, the data, and the model. However, even with a small number of languages such an analysis is computationally infeasible to conduct. Computationally less demanding strategies are needed to collect a sample of trees that can function as an estimate of the true posterior probability distribution.

Such a tree sample may be acquired using Markov Chain Monte Carlo (MCMC) methods (Pagel and Meade 2005: 240ff; see Dunn 2009 for an introduction into Bayesian MCMC approaches for linguistic tree building). MCMC analysis can be viewed as a hill climbing procedure: The most likely phylogenetic trees can be found on the tops of hills, and the MCMC methods climb these hills by jumping from possible tree to possible tree using Markov chains, ultimately finding their way to the trees with the highest likelihoods. This procedure is illustrated in Figure 1.5. Markov chains are mathematical devices that jump from state to state within the parameter space that describes possible phylogenetic trees. The probability of each jump is determined only by the current state, not by any of the previous states. For phylogenetic purposes, the states in Markov chains are different phylogenetic trees. At each step in the chain, a slightly changed new tree is proposed - this could be a change in the topology of the tree, in the branch lengths, or in the parameters of the model of word evolution. Whether a new tree occurs is determined by the Metropolis-Hastings algorithm.

When the newly proposed tree is an improvement on the previous tree, the jump is always made, otherwise it is made with a probability that is dependent on how bad the newly proposed tree is compared with the previous tree (Pagel and Meade 2005: 241).

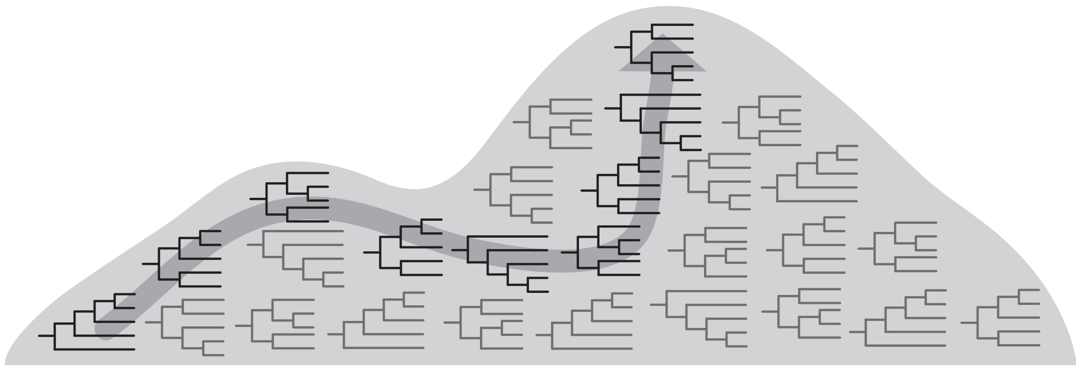


Figure 1.5: *An illustration of a MCMC chain that moves through the parameter space*

If such a Markov chain is run for enough time, it reaches stationary distribution. In stationary distribution, the chain jumps from tree to tree in the tree space, some of which are slightly better, some of which are slightly worse, but it no longer moves to better and better trees. In this stationary distribution, the chain samples trees from the tree space in proportion to their frequency of occurrence in the tree space. In this way it constructs a sample of trees that approximates the 'true' posterior probability distribution (Pagel and Meade 2005: 241).

1.3.2.2 The set of phylogenetic trees used in this dissertation

The set of phylogenetic trees that is used in this dissertation is an example of a set of trees that have been estimated using Bayesian methods. Using a sample of trees rather than a single individual tree allows one to account for the uncertainty that is a part of every phylogenetic estimate due to the different histories represented by different linguistic features such as cognate sets. The current set of trees was taken from Bouckaert et al. (2012), who collected cognate-coded lexical data (Swadesh lists) on 103 Indo-European languages (see Dunn et al. to appear for the cognate database). A binary conversion of this data resulted in a matrix that indicated absence (0) or presence (1) of a cognate for each of the 5047 individual cognate sets.

Bouckaert et al. (2012) used the Bayesian Markov Chain Monte Carlo approach (Huelsenbeck et al. 2001) available in the software BEAST (Drummond et al. 2012) to estimate a posterior distribution of phylogenetic trees. They used the Stochastic Dollo substitution model with an uncorrelated log-normal relaxed clock (Bouckaert et al. 2012, supplementary materials). This method did not enforce a constant rate of cognate replacement, rather the rate of cognate replacement was allowed to be different for different cognate sets as well as in different parts of the tree. A sample of 12,500 phylogenetic trees with high likelihoods was constructed using BEAST and constituted Bouckaert et al.'s (2012) definitive tree sample that was used to study the origins and expansion of the Indo-European language family.

A further random selection of 1000 trees from Bouckaert et al.'s (2012) tree sample was made by myself to use in this dissertation using LogCombiner (part of the BEAST package; Drummond et al. 2012). The reason for this random selection was purely one of manageability: running an analysis with 1000 phylogenetic trees takes less time than running it with 12,500 trees. The random selection of 1000 trees was made possible entirely by the high stratification that is present in Bouckaert et al.'s (2012) tree sample: as can be observed in Figure 1.6, the large majority of the nodes that are relevant for this dissertation are attested in all phylogenetic trees (see below). If higher levels of phylogenetic uncertainty would have been evident in Bouckaert et al.'s (2012) tree sample, the analyses presented in this dissertation would have been run over their full 12,500 tree sample.

Languages that are not studied in this dissertation were removed from the selection of trees, so that they included only the languages for which data was collected. Two tree samples are distinguished: one that includes all 20 Indo-European languages (Albanian, Armenian, Dutch, English, French, German, Modern Greek, Hindi, Irish, Italian, Latvian, Lithuanian, Nepali, Persian, Polish, Portuguese, Romanian, Russian, Serbo-Croatian, and Swedish) and one that includes a subset of 16 of these languages, excluding Albanian, Hindi, Nepali, and Persian. This second, more restricted tree sample is used for certain analyses in chapter 5, while the first tree sample that includes all languages is used for all other analyses in chapter 4, 5 and 6. To illustrate what these tree samples look like, maximum clade credibility trees of both the 1000 tree sample for 20 languages and the 1000 tree sample for 16 languages were calculated using TreeAnnotator v.1.6.1 (Drummond et al. 2012). These trees are presented in Figure 1.6 and Figure 1.7. Note however that the phylogenetic comparative analyses presented in this dissertation were conducted over all trees in both samples of 1000 phylogenetic trees. The maximum clade credibility tree summarizes the whole tree sample by taking and averaging over the tree which receives the most global support. It is built by evaluating each of the sampled posterior trees by scoring each clade within the tree based on the number of times that the clade appears in other sampled posterior trees. These scores are multiplied to give a total score for each tree, and the tree with the highest score is the maximum clade credibility tree. This tree then is adjusted to have median branch lengths based on the branch lengths of all the trees in the posterior sample. In this sense, the maximum clade credibility tree is different from the majority rules consensus tree, which includes clades that appear in more than 50% of the sampled posterior trees. The majority rules consensus tree can have a topology that is not present in the posterior sample, while the maximum clade credibility tree always has a topology that was sampled in the posterior distribution.

The support values presented for each internal node of the maximum clade credibility trees in Figure 1.6 and 1.7 indicate how often each clade is attested within the tree sample. A support value of 1 indicates that this internal node is attested in all the trees in the sample. A support value lower than 1 indicates that this internal node is only attested in a subset of the trees in the sample. The internal node that leads to the Romance, Celtic, Germanic and Balto-Slavic subgroups in Figure 1.6, for instance, has a support value of 0.99, indicating that it appears in 99% of the trees in the sample. The length of the branches of the phylogenetic trees are drawn in proportion to time. Since Bouckaert et al.'s (2012) tree sample was time-calibrated but the current sample only includes contemporary languages, the trees are ultrametric, meaning that all branch lengths leading from the root node to the different languages on the tips of the tree have equal length.

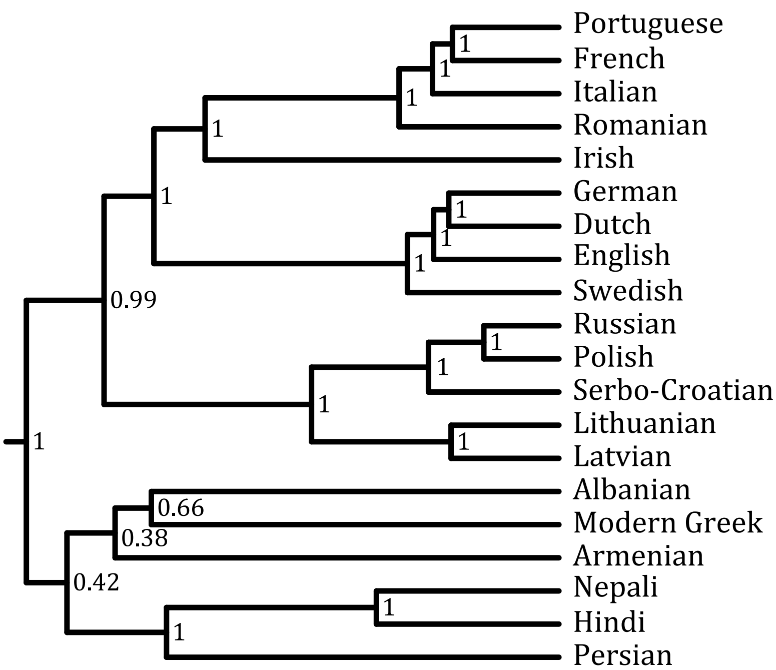


Figure 1.6: *The maximum clade credibility tree of 1000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 20 languages featured in this dissertation.*

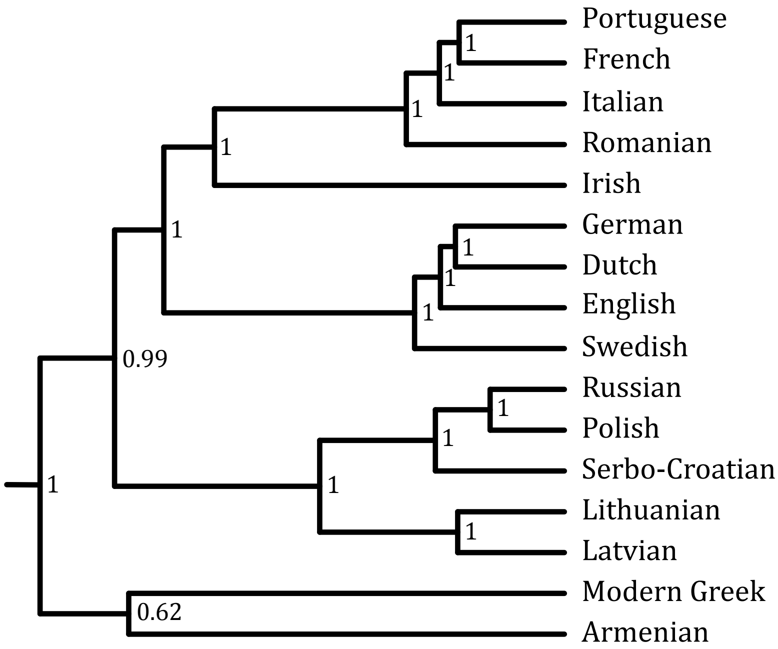


***Figure 1.7**: The maximum clade credibility tree of 1000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 16 languages featured in some analyses presented in this dissertation.*

The trees proposed by Bouckaert et al. (2012) have been part of a debate on the application of phylogenetic methods to study genealogical relationships between languages. Criticism has focused on their use of cognate-coded lexical data rather than regular sound changes, which are generally regarded to be more reliable to recover genealogical relationships. Also, some problems have been found with the dates used for some of the calibration points, and with the geographic location of certain languages. However, this dissertation is not concerned with Bouckaert et al.'s (2012) specific hypotheses regarding the geographical expansion of the Indo-European language family through space and time, rather it uses the structure of their phylogenetic trees to diachronic change in motion event encoding. Since Bouckaert et al.'s (2012) trees are the only published trees built using the much improved Dyen et al. (1992) dataset (Dunn et al. to appear), they are the best choice at this time.

1.3.3 Phylogenetic comparative methods and their applications

This section gives an overview of the types of questions that can be answered using phylogenetic comparative methods. The various types of analyses will be discussed in some mathematical detail, as the method sections of the empirical chapters of this dissertation are briefer due to space constraints. In addition, there exist conceptual and mathematical commonalities behind all of the methods used in this thesis, and explaining them here together will allow the reader to better understand these commonalities.

1.3.3.1 Models of evolutionary change

This section starts off with an explanation of the general models of evolutionary change that are employed by the different phylogenetic comparative methods. These models have close parallels with the models for phylogenetic tree inference discussed in section 1.3.2.1. Different evolutionary models exist for discrete data and continuous data. Discrete data are data for which there are only a finite number of values possible, such as data on the word order of subject, object and verb, for which six values are possible (SOV, SVO, VSO, OSV, OVS, VOS). Discrete data is typically categorical, and can be binary (yes, no) or multistate (SOV, SVO, VSO, OSV, OVS, VOS). Continuous data are quantitative data (interval or ratio) that can take any value within a range, such as the frequency of SOV word order in a given language corpus (which can range from 0% to 100%).

One of the ways to model discrete trait evolution is to use a continuous-time Markov model that plots evolutionary changes along the branches of a phylogenetic tree (Pagel 1994). A continuous-time Markov model assumes that the probability of change is independent in every segment of every branch of the tree, and that the probability of change is dependent only on the state of the feature at the beginning of the branch segment under consideration, and not on anything that has happened before that time (Pagel 1994: 38). As discussed in section 1.3.2.1, branch length is taken as an operational unit of time (Pagel 1999b: 613). Branch length can represent either the amount of evolution, i.e. the number of substitutions per site, or the amount of time, i.e. the number of substitutions per site per year. In linguistic phylogenetic trees built on lexical cognate data, substitutions are cognate gains and losses. Branch length can only represent time if information on 'fossils' such as Gothic or Ancient Greek can be used to date specific internal nodes of the phylogenetic tree. If two languages are connected by longer branches than two other languages, this implies a greater genealogical distance between them, i.e. a longer time of separate evolution.

If we take a single feature that has two possible states, 0 and 1, we can describe the evolutionary changes that can take place in this feature with four probabilities: P_{00}, the probability that over a certain period of time, a feature that started out in state 0 stays 0, P_{01}, the probability that a feature changes from 0 to 1, P_{11}, the probability that a feature that started out in state 1 stays in state 1, and P_{10}, the probability that a feature changes from 1 to 0 (Pagel 1994). These probabilities are illustrated in (5).

5) P_{00}: 0 → 0
P_{01}: 0 → 1
P_{11}: 1 → 1
P_{10}: 1 → 0

Because there are only two options for each starting state, P_{00} can be derived from P_{01}: P_{00} = 1 - P(0 ← 1); and P_{11} can be derived from P_{10}: P_{11} = 1 - P(1 → 0). Thus, the information that is necessary to assess the evolutionary change of the feature along the branches of the tree is the transition rate between state 1 and 0 and the transition rate between 0 and 1. We have encountered these transition rates earlier in section 1.3.2.1, but they are presented again in (6).

6) q_{10}: 1 → 0
q_{01}: 1 ← 0

Using only these rate parameters and the length of the branch, which is equal to time, Pagel (1994: 38-39) shows that it is possible to determine the probabilities of the feature states at each point in time. In his words:

> For example, P_{01} (t) will be a function of the rate of transition from 0 to 1, balanced over time *t* by the rate of transition from 1 to 0: the larger q_{01} is relative to q_{10}, the greater the probability that at the end of time *t* a character beginning in state 0 will be in state 1. (Pagel 1994: 39).

The rates q_{10} and q_{01}, which are vital for the resolution of any evolutionary question for discrete data, can be estimated using maximum likelihood or Bayesian methods. The maximum likelihood solution is estimated by searching the likelihood surface for the transition rates q_{10} and q_{01} that have the largest likelihood, given the length of each branch (Pagel 1994). Bayesian methods can be used to estimate the posterior probability distribution of rates q_{10} and q_{01} (Pagel and Meade 2005: 243). MCMC chains can be used to estimate values of the rate parameters, resulting in a stationary distribution of the chain that samples the posterior distribution of rates.

The transition rates q10 and q01 are in themselves interesting parts of the evolutionary model to look at. If the rates are approximately equal, that means that it is equally likely that the feature changes state from 0 to 1 as it is to change from 1 to 0. However, if q10 is bigger than q01, this means that changes from 1 to 0 are more likely than changes from 0 to 1, which indicates that state 1 is gradually being lost and getting replaced by state 0. This could, depending on the magnitude of the difference in rates, result in all languages having state 0 if evolutionary change would continue onwards in the same direction. This type of directional trend can be an interesting finding of a comparative phylogenetic analysis by itself.

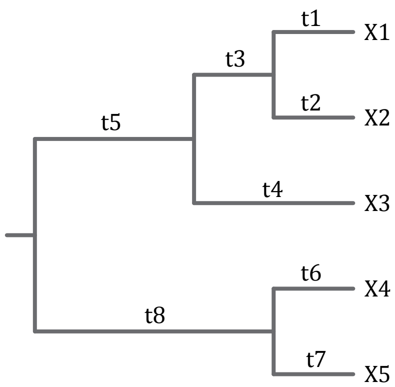
The evolution of continuous features can be studied with the constant-variance random walk model of evolution, which is also known as the Brownian motion model (Harvey and Pagel 1991: 115ff; Pagel 1999a: 878; Pagel 2002: 270ff; O'Meara et al. 2006: 922). In the simplest version of the Brownian motion model, the states of a trait can increase or decrease at each instant of time with a mean change of zero and a fixed variance. More complicated versions of the Brownian model allow for directional rates of change or accelerating and decelerating rates of change. Each of these changes is completely independent from the current state of the feature. The trait data is assumed to adhere to a normal distribution, although more complex methods can account for data with non-normal distributions. There are several different approaches to trait evolution of continuous features that employ the Brownian motion model. The most common are Phylogenetic Generalized Least Squares (PGLS) (Martins and Hansen 1997; Pagel 1997: 338ff; Pagel 1999a: 878ff; Garland and Ives 2000) and Phylogenetically Independent Contrasts (PIC) (Felsenstein 1985: 8; Garland and Ives 2000), which can both be implemented using maximum likelihood methods or Bayesian methods.

Phylogenetic Generalized Least Squares can be implemented as a regression analysis, in which each value for each language is predicted from the regression of the feature on the branch leading from the root to that individual language (Pagel 1997: 337ff). For a trait value in one specific language X, evolving on a branch with a total length of Y, the regression formula is as follows:

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

In this formula, a is the y-axis intercept of the trait values that are regressed on the total branch length. β is the slope of the regression line that relates the trait value X to operational time Y, measured by the lengths of the branches leading from the root of the tree to the languages at the tips of the tree. e is the error term. In effect, then, the trait value X evolves along the branch of the tree leading from the root of that tree to the language tip at a rate β per unit of branch length. In this model, variation in trait values among different languages evolves due to different branch lengths from the root to the tips of the tree.

If the language data were independent, the value β could be estimated using conventional regression techniques. However, the language data is not independent because of the shared history of the languages. The languages share some proportion of the branch lengths that lead from the roots to the tips. The expected variance (variability in a feature) of a trait value is proportional to the time it has been evolving, which is proportional to the length of the branch leading from the root of the tip. The expected covariance (lack of independence) of a trait value is proportional to the branch length shared between each of the languages. An example of a variance-covariance matrix that contains all this information on (shared) branch lengths is provided in Figure 1.8. The PGLS model takes into account the variance-covariance matrix in the error term e to estimate a phylogenetically adjusted value of β. This value of β is used in the regression formula to maximize the likelihood of observing the set of trait values across languages given the phylogeny. Aside from solving the regression formula while taking into account shared history, the algorithm also infers rates of change and ancestral states.



	X1	X2	X3	X4	X5
X1	t5+t3+t1	t5+t3	t5	-	-
X2	t5+t3	t5+t3+t2	t5	-	-
X3	t5	t5	t5+t4	-	-
X4	-	-	-	t8+t6	t8
X5	-	-	-	t8	t8+t7

Figure 1.8*: An example of a tree with a corresponding variance-covariance matrix*

Estimating the rates of change and looking at directionality in those rates is hardly ever the only question asked in comparative phylogenetics, as most studies of trait evolution investigate more complex questions, such as ancestral-state estimation or correlated evolution. However, often statements about the rates and directionality of change are very interesting in themselves. An example of this from evolutionary biology is the study by Barkman et al. (2008), who have shown that rates of change have increased over time in the family of *Rafflesia* flowers. The *Rafflesia* family is known for having the largest flowers on earth, with some species having flowers over one meter in diameter. The flowers smell like rotten meat to attract insects, which play a role in the reproduction of *Rafflesia*. The evolutionary process that led to these enormous flowers was one of continuously increasing rates of change. As ancestral flower size increased, the rate of change increased as well, leading to ever-bigger flowers in certain parts of the family. This illustrates that the rates of change that are inferred by evolutionary models are a very important aspect of correctly modeling evolution. In this dissertation, rates of change are used to compare classes of motion verbs in chapter 6.

1.3.3.2 Testing for historical signal

Before doing any type of phylogenetic comparative analysis, it is generally recommended to estimate whether history actually has influenced the current distribution of the states attested for the trait under investigation (Freckleton et al. 2002: 724). Phylogenetic signal is present when closely related languages exhibit similar traits and when trait similarity between related languages decreases as their most recent common ancestor is situated in a more remote past (Losos, 2008). There are several methods available that can be used to assess whether phylogenetic or historical signal is present or not.

One of the most extensively used tests has been developed by Pagel (1999a) and is called lambda (λ) (see Freckleton et al. 2002 for discussion). Lambda is a branch length scaling parameter and can be used to measure the extent of the dependency of the data on the model of Brownian evolution given the tree topology. The value of lambda can be optimized using maximum likelihood methods. The optimized lambda value indicates to what extent the data under consideration have been influenced by their shared history.

The optimization of lambda works as follows. First, the algorithm calculates the variance-covariance matrix based on the tree topology (see Figure 1.8). The covariance values in this matrix are measures of shared branch length between the different languages. Then, the algorithm determines which is the optimal value of lambda, a number between 0 and 1 with which the covariance values are multiplied. If the covariance values are multiplied by lambda = 0, all shared branch lengths and thus all evidence of shared history between the languages is erased. A low (0 or another low value) optimized lambda value implies that the genealogical relationships between the languages have not influenced the trait data to a large extent. If the covariance values are multiplied by lambda = 1, the information on shared history of course stays exactly as it is. A high (1 or another high value) optimized lambda value implies that the feature is evolving exactly along the branches of the phylogenetic tree under a random walk model of evolution (Pagel 1999a), and thus that there is evidence for shared history. Any number between 0 and 1 decreases the amount of shared history between the languages. These manipulations are illustrated in Figure 1.9.

After the optimized value of lambda is retrieved, it can then be tested whether the optimized lambda value is significantly different from a model in which lambda = 1 or lambda = 0 using a likelihood ratio test (Pagel 1997: 334). These tests assess 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. In short, the estimation of lambda can in this way provide statistically sound evidence for the presence or absence of phylogenetic signal.

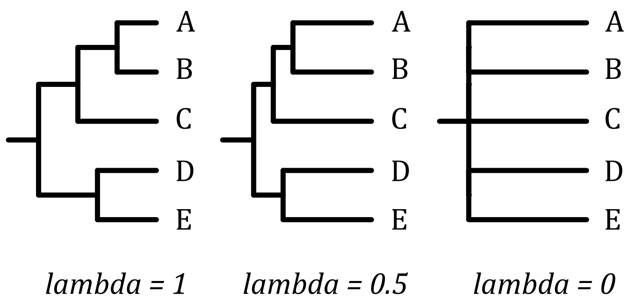


***Figure 1.9**: Three versions of the same tree with lambda modified*

In some cases, the optimized lambda can be higher than 1 (Freckleton et al. 2002: 715). Since the optimized lambda is defined as the transformation of the shared branch length that makes the comparative data best fit the phylogeny, lambdas higher than 1 may arise when the data is more similar than predicted by the model of Brownian motion on that particular phylogenetic tree. However, the optimized lambda values is restricted by the variance-covariance matrix (see again Figure 1.8): the off-diagonal branch lengths can never be longer than the diagonal branch lengths. For that reason, lambda can never be much larger than 1: Freckleton et al. (2002) cite a lambda of 1.24 for data on body size.

The interpretation of lambda estimates, for instance in comparing the evolution of different features on the same tree, is unfortunately not straightforward. Although it would be useful to be able to say that a low optimized lambda value indicates that the trait is influenced by other processes than shared genealogical descent, such as lack of variation in the feature, parallel change with other linguistic features, very rapid linguistic change, or borrowing, this would be misguided as different evolutionary processes can be the cause of similar lambda estimates. Revell et al. (2008) have used simulations to demonstrate that factors that act on the evolutionary process, including the rate of evolution, functional constraints, fluctuating selection, niche conservatism, and evolutionary heterogeneity, interact in complex ways and may affect the results of a phylogenetic signal analysis in a way that does not allow for disentangling the contribution of different factors. Even though not all of these processes will act on linguistic features, the interpretation of low lambda estimates remains quite limited. The most important thing that the estimation of lambda tells us is whether there is evidence for similarity among languages due to shared descent.

Strong phylogenetic signal is found in a range of biological and anthropological traits. Freckleton et al. (2002) show that body size in both invertebrate and vertebrate species often have large lambda values, while other measures, such as parasite density in fish and mammals, have low lambda values. Jordan and Currie (submitted) find high lambda values for both population size

(0.90) and population density (0.86) of Austronesian cultures. In this dissertation, tests for historical signal using the estimation of lambda are included in chapters 3, 4, and 5.

1.3.3.3 Estimating ancestral states

The aim of estimating ancestral states is to infer how ancestral languages might have behaved with regard to the trait under investigation. Ancestral states are automatically reconstructed during most phylogenetic comparative analyses, as they are as much a vital part of the modeling of the evolution of a trait as the rates of change are. The estimation of ancestral states allows for the investigation of the likelihood of any hypothesis that we may have about the behavior of ancestral languages.

Ancestral-state reconstruction using maximum likelihood methods for discrete data rely on the continuous time Markov model that was introduced in section 1.3.3.1 (Schluter et al. 1997; Pagel 1997, 1999a, 1999b). The Markov model estimates the rates at which a discrete trait changes its state along the branches of the tree (see (6), repeated below as (8) for convenience).

8) q_{10}: 1 $\rightarrow$ 0
q_{01}: 1 $\leftarrow$ 0

Using this information, it calculates the most probable ancestral states for the internal nodes in the tree. It does this by calculating the likelihood of observing the linguistic data, while fixing the internal node at each of the possible states of the trait. The state with the highest likelihood is the most probable state for that node. A possible outcome of this process is illustrated in Figure 1.10, where inferences for the internal node are placed above each node.

Ancestral-state estimation using maximum likelihood methods for continuous data most commonly relies on the constant-variance random walk model (or Brownian motion model) that was introduced in section 1.3.3.1 (Pagel 1999a: 878). Phylogenetic Generalized Least Squares (PGLS) and Phylogenetically Independent Contrasts (PIC) can both be used to estimate ancestral states and yield similar results (Pagel 1999a: 878; Garland and Ives 2000). PGLS methods using maximum likelihood approaches calculate ancestral states by minimizing the sum of squared changes across the branches (Schluter et al 1997: 1701). An example of a hypothetical outcome is given in Figure 1.10, in which distributions of the inference for each internal node are plotted over each node. This approach infers the most likely ancestral states based on an evaluation of each internal node in terms of the sum of squared changes in the clade connected by that node for each possible state that node can take. Going towards the root of the tree, it attempts to find the minimum sum of squared changes in the whole tree and thus estimate the most likely ancestral states of all internal nodes (Maddison 1991: 305). However, a range of different methods to estimate ancestral states is available (Martins and Hansen 1997: 661; Webster and Purvis 2002).

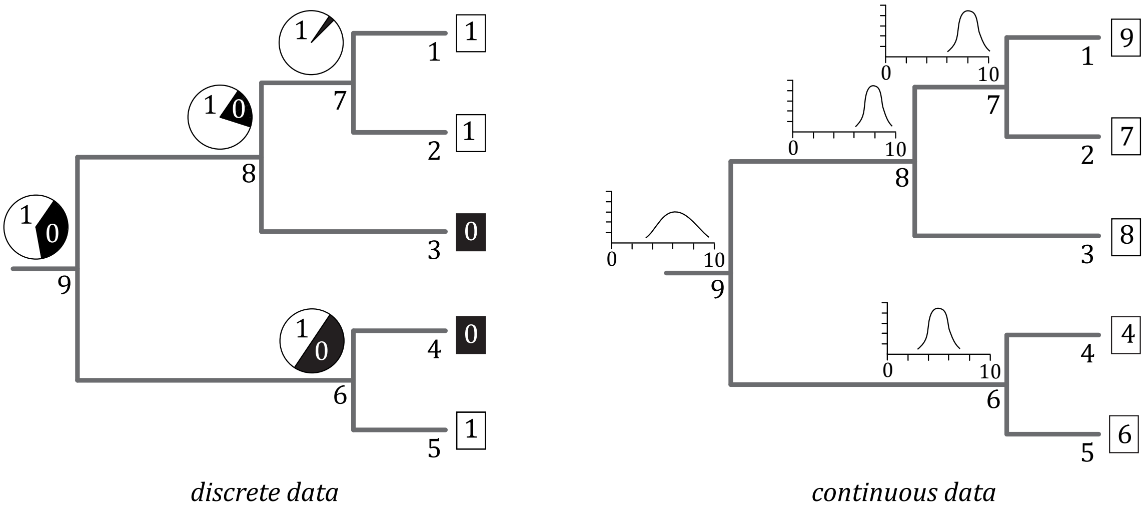


***Figure 1.10**: Ancestral-state reconstruction (based on Haun et al. 2010's Figure 2). The inference for each internal node is plotted over each node; inference distributions are different for discrete data and continuous data. The most likely ancestral states are determined by taking into account the amount of evolutionary change within each clade for each possible value of an internal node, minimizing the total amount of evolutionary change given all nodes.*

Ancestral-state estimation is a popular analysis in evolutionary biology and comparative anthropology. Martins and Lamont (1998) estimate ancestral-states of headbob displays of *Cyclura* iguanas, using several different measures such as the number of headbobs and the duration of headbobs. Ancestral states were estimated using the maximum likelihood approach outlined in Martins and Hansen (1997). Martins and Lamont (1998) found that headbob displays have evolved frequently and dramatically, although some measures seemed to be more stable then others. Jordan et al. (2009) infer the ancestral type of residence in the Austronesian language family using Bayesian methods. They infer that ancient Austronesian societies are most likely to have been matrilocal. Currie et al. (2010) infer the ancestral state of political complexity in Island South-East Asia. Proto-Austronesian societies are inferred to have an acephalous system. In this dissertation, ancestral-state estimation analysis is used in chapter 4 to investigate changes in motion event encoding systems and infer the behavior of Proto-Indo-European.

1.3.3.4 Testing for co-evolution

The aim of testing for co-evolution is to investigate whether two or more characteristics of languages are correlated or can be said to have evolved together. It is especially important to use phylogenetic comparative methods to investigate this type of question, because it is easy to over-estimate a correlation between two features if historical dependencies have not been taken into account. This type of analysis is of particular importance for language evolution, because of the many dependencies between linguistic features that have been found since the advent of linguistic typology. Many of Greenberg's (1966) 45 universals of word order and morphology had a conditional form: if a language has feature A, it also has feature B. The study of the mechanisms that have generated these dependencies using phylogenetic comparative methods will ultimately reveal whether the dependencies are universal (Dunn et al. 2011; Levinson et al. 2011) and through what kind of processes they emerge.

Correlated evolution of discrete binary features can be modeled by the Markov model that was described in section 1.3.3.1 (Pagel 1994: 38, 1999a: 882). The algorithm tests whether it is more likely that the two traits have evolved together or that they have evolved separately. The likelihood of the data assuming that both feature X and feature Y have evolved separately is calculated by taking the product of the likelihood that the tree and the model have generated the data set of feature X and the likelihood that the tree and the model have generated the data set of feature Y. These two likelihoods are estimated by searching the likelihood surface for the probabilities of the state of each node in the tree (P_{x1} and P_{x0} for feature X, and P_{y1} and P_{y0} for feature Y), given transition rates of states given in (9) and the length of the branch. In (9), 1 indicates presence of the feature, while 0 indicates absence of the feature.

9) X: q_{x10}: 1 → 0
q_{x01}: 1 ← 0
Y: q_{y10}: 1 → 0
q_{y01}: 1 ← 0

The calculation of the likelihood of the data assuming that the two traits are evolving together requires a more complex model, which allows that the changes in feature X to be dependent on the state of feature Y and vice versa. In this analysis, one is calculating the probabilities of change for both feature X and feature Y at the same time, for instance the probability that feature X changes from absent (0) to present (1) and feature Y changes from present (1) to absent

(0). The transitions that are assumed in such a dependent model are depicted in Figure 1.11.

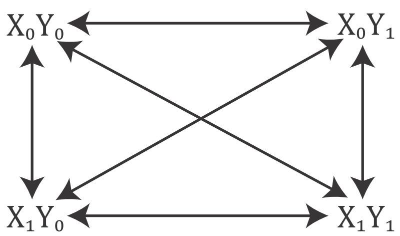


Figure 1.11*: Transitions in a dependent model of evolution of feature X and Y*

If the changes in feature X and Y are dependent upon one another, the probability of change in feature X and Y cannot simply be described as the product of the probability of change in X and the probability of change in Y. To compare the results of the independent and dependent analysis, the likelihoods of both the dependent model and the independent model are calculated. Whether the two models are significantly different from each other can be tested with the likelihood ratio statistic (Pagel 1994: 41) or Bayes factors (Pagel and Meade 2005; Pagel and Meade 2006: 813ff). In addition, it is possible to test for directionality of change, for example whether changes from 1 to 0 in Y are more likely when X has state 1.

Correlated evolution of continuous features can be modeled by the constant-variance random walk model (or Brownian motion) that was described in section 1.3.3.1. There are two types of statistical analyses that can be used to test for correlated evolution of continuous features: regression analysis, which predicts the variance in a dependent variable as a function of one or more independent variables, or multivariate analysis of correlations between two or more dependent variables (Rohlf 2006: 1509).

The Phylogenetic Generalized Least Squares (PGLS) regression works as follows (Pagel 1997: 340ff). In the formula in (10), a feature X is dependent on a feature Y, a is the intercept of the regression line, β is the regression of the X feature on the Y feature, and e is the measure of error. Specifically, β specifies the amount of change in X given change in Y, i.e. how X changed as Y evolved over time.

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

This model is slightly different as the PGLS model in which the evolution of a single trait is modeled, although the principle remains the same. In section 1.3.3.1 was described how a single trait X is regressed against operational evolutionary time Y. In a PGLS regression analysis that tests the correlation of two linguistic features, feature X is regressed against feature Y. As was the case for the single feature model in section 1.3.3.1, the error term e is adjusted for phylogenetic relatedness using the variance-covariance matrix (see again Figure 1.8).

The multivariate analysis of correlations between two or more dependent variables works as follows (Rohlf 2006: 1509). The model for this analysis is given in (11), where Z is an n × p matrix of n observations of p dependent variables, 1_n is an unit vector of length n, μ is an 1 × p vector of means, and E is an n × p matrix of multivariate normally distributed errors.

11) $Z = 1_n\mu + e$

Similar as the PGLS regression, a PGLS correlation is calculated by incorporating information on the amount of phylogenetic relatedness in the error terms of the formula (see again Figure 1.8). The PGLS regression and the PGLS correlation can be calculated using both maximum likelihood and Bayesian approaches. The results of these analyses can be tested for significance using likelihood ratio tests or Bayes factors, in which one compares the dependent model to an independent model, or the results of a test in which β or the correlation is forced to be zero.

Dunn et al. (2011) investigated correlated evolution of eight word-order features: the order of genitive and noun, adposition and noun, numeral and noun, adjective and noun, demonstrative and noun, relative clause and noun, subject and verb, and object and verb. They showed, contra the predications made by Greenberg (1966) and Dryer (1992), that dependencies between these word order features are not the same in different language families. The correlations differ in each of the four different language families that were investigated. In this dissertation, co-evolution analyses are employed in chapter 5 to investigate the correlation between the size of the motion verb lexicon and the motion event encoding system.

1.3.4 Discussion

The use of statistical methods developed for analyzing biological evolution to model cultural and linguistic change is subject to debate (Bateman et al. 1990; Boyd et al. 1997; Claidière and Andre 2012; Moore 1994). The most common criticism is that phylogenetic comparative methods only take into account vertical transmission and do not account for the acquisition of cultural or linguistic features through horizontal transmission (Borgerhoff-Mulder 2001; Borgerhoff-Mulder et al 2006; Boyd et al. 1997; Tëmkin and Eldredge 2007). Horizontal transmission takes place when change in one community is induced by contact with another community. An example of this is the borrowing of words from a different language, such as English *mammoth* (from Yakut via Russian), *taboo* (from Hawaiian), and *hooligan* (from Irish). All kinds of linguistic elements (phonemes, words, morphology, syntactic constructions, pragmatic conventions etc.) can be borrowed from one language community to another, depending on the extent of the contact between the two communities. This implies that there are two lines of transmission, one vertical across generations of language learners and one horizontal through interaction between language communities, that need to be distinguished when building phylogenetic trees and when investigating linguistic features.

Horizontal transmission of cultural or linguistic features poses a problem when one wants to study the vertical evolution of languages, as a focus on vertical transmission might not always be appropriate because of large amounts of horizontal transmission in the data. This is both the case when we want to infer phylogenies (Nelson-Sathi et al. 2010; Nichols and Warnow 2008: 789, 809ff) or when we want to study trait evolution. Note that in both cases, borrowing is not an issue if a single cognate or feature is borrowed in a single language, as it is simply considered to be an innovation. However, borrowing can become a problem when a borrowed cognate or feature is subsequently inherited by descendent languages or cultures. This issue has led to a big debate in anthropology on the applicability of these methods to cultural and linguistic data (Atkinson and Gray 2006; Borgerhoff-Mulder et al. 2006; Gray et al. 2007: 365ff; Gray et al. 2010). However, much of this debate has been fueled by a priori philosophical objections and anecdotes rather than rigorous empirical investigation (but see below). Recent advances in Bayesian methods can address uncertainty with regard to the phylogeny and the evolutionary model more adequately than has been possible in the past (Gray et al. 2010: 3924).

However, it is possible to simulate the impact of horizontal transmission on the inference of phylogenetic trees and the application of phylogenetic comparative methods (Collard et al. 2006; Currie et al. 2010; Greenhill et al. 2009; Nunn et al. 2010). Such simulations point out that there are certain conditions in which comparative phylogenetic methods behave well and seem to recover the signal from vertical transmission, even when certain amounts of horizontal transmission are present. In other conditions, the signal of vertical transmission may be partly lost because of the horizontal transmission that is present. In practice, the presence of an influence of horizontal transmission on the results of the application of phylogenetic comparative methods on a particular dataset will have to be assessed for that particular dataset. Solutions can then be found to take into account such dependencies between the languages in the dataset.

Phylogenetic comparative methods have only been developed during the last four decades. Although results always need to be interpreted with care, for instance by taking into account potential influences from horizontal transmission, great advances in their development are still being made. Most of these developments will regard the calculation and use of phylogenetic networks for comparative analyses. The advantage of inferring phylogenetic networks alongside phylogenetic trees is being recognized increasingly (Chen et al. 2013). Although especially the inference of rooted phylogenetic networks is computationally difficult (Huson and Scornavacca 2011), their development will subdue much of the debate surrounding the use phylogenetic methods. In addition, even though phylogenetic comparative methods may be flawed, methods that do not take into account phylogenetic dependencies are flawed as well, albeit in different ways (see section 1.3.1). The chance of finding a correlation between two linguistic features that is in fact an accident of shared history, for instance, becomes much larger if one does not control for phylogenetic dependencies, even if these are accounted for by controlled sampling. Removing the variance that can be attributed to phylogenetic dependencies results in throwing away information about differences between languages in different genealogical clusters.

The take away message here is that phylogenetic comparative methods can be used to answer questions about language change in a sophisticated and statistically sound way. They can be employed to infer ancestral states, directional rates of change, correlations between linguistic features, homelands, and dates of divergence events (Gray et al. 2007). Therefore the use of phylogenetic comparative methods is not just about incorporating history in comparative analyses: it allows for a more informed outlook on linguistic diversity and the mechanisms generating that diversity.

Having now explained what phylogenetic comparative methods are and why they are useful to answer questions about diachronic change in linguistics, an outline of the rest of the dissertation is given in the next section.

1.4 Outline of the dissertation

Chapter 1 has introduced the topic of this dissertation and the methodologies that have been used to investigate the main questions. The methodological section is chapter 1 are meant as a guide to the non-specialist reader; subsequently, the methodological sections in the empirical chapters are more brief.

Chapter 2 is an introduction to the dataset that will be used throughout the dissertation. It discusses in some detail the parallel corpus that is the source of the data, the rationale behind the coding of the dataset and how the coding differs from alternative coding schemes used in the motion literature. It also introduces the various sentence samples that are distinguished and the main aggregation method (principal components analysis) that is used to aggregate the data on motion event encoding.

Chapter 3 presents an introduction of the motion event encoding systems of the twenty Indo-European languages included in the sample. A contribution to the general motion event literature is made when results on aggregation analyses are presented that show that the diversity encountered in the sample cannot be captured in a simple dichotomy of verb-framed and satellite-framed languages that has traditionally been proposed.

Chapter 4 examines the diachronic changes that have taken place in the motion event encoding systems of twenty Indo-European languages. Data on ancient Indo-European languages as well as ancestral-state estimation analyses are used to investigate diachronic change as well as the behavior of Proto-Indo-European. The influence of the important diachronic process in which preverbs merged with verb roots on the motion event encoding systems of the different Indo-European languages is discussed. Contra earlier claims, both qualitative data on ancient Indo-European languages as well as quantitative results of ancestral state estimates point towards a mixed motion event system for Proto-Indo-European.

Chapter 5 tests correlated evolution between the size of motion verb classes and motion event encoding systems. It is shown that there exists some evidence for a phylogenetically controlled correlation between the size of the manner of motion verb class and the use of the satellite-framed construction. It also presents evidence for a correlation between the size of the path of motion verb class and the use of the verb-framed construction. This chapter also includes overviews of the lexicons of manner verbs and path verbs of the Indo-European languages.

Chapter 6 investigates the emergence of larger manner of motion verb lexicons in satellite-framed languages and larger path of motion verb lexicons in verb-framed languages more closely. Results from an investigation into the etymological origins of these motion verbs suggested that manner of motion verbs typically have different types of etymological origins as path of motion verbs. A study of the rates of change in different branches of the Indo-European tree suggested that manner of motion verbs evolve faster in the satellite-framed subgroups of Indo-European, while path of motion verbs evolve faster in verb-framed subgroups.

Chapter 7 brings together the themes researched in this thesis by providing general discussion and future directions. It provides further discussion on language-internal diversity, diachronic change of the use of motion event encoding constructions, correlations between syntactic and lexical features of motion encoding, and rates of lexical evolution by relating the results of the empirical chapters to a broader literature. It also reflects on the future of the application of phylogenetic comparative methods in linguistics.

Chapter 2: Materials and coding

Parts of this chapter have been taken from:
Verkerk, Annemarie. (2014c). Where Alice fell into: Motion events in a parallel corpus. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), Aggregating dialectology, typology and register analysis: Linguistic variation in text and speech *(pp. 324-354). Berlin: Walter de Gruyter.*

This chapter is devoted to an introduction to the dataset that was used for the analyses presented in the following chapters. It starts with an introduction to the parallel corpus (section 2.1), continues with a discussion of the coding scheme used to analyze motion events (section 2.2), and concludes by introducing the data reduction methodology (section 2.3).

2.1 The parallel corpus

2.1.1 Deciding on source material

The dataset that is employed in this dissertation comes from a parallel corpus. Parallel corpora consist of parallel texts, texts that are all translations of a single original text, which is also included in the corpus. The most obvious parallel text is the Christian Bible, of which parts have been translated into over a thousand languages (Cysouw and Wälchli 2007).

Using a parallel corpus to study the encoding of motion events has several advantages (see Wälchli 2001a; Slobin 1996b, 2005b; Baicchi 2005 for similar approaches). First of all, since motion events constitute a mostly lexical topic that is prevalent in natural language, using a parallel corpus is very suitable (Wälchli 2007: 128). In other words, parallel texts provide a bountiful source of motion descriptions. Secondly, the original text restricts the semantic primitives under study; i.e. the corpus consists of a finite set of linguistic expressions that are more or less equivalent. Thirdly, parallel texts are also highly adequate for investigating language-internal variability (Wälchli 2007: 129) – which is one of the aims in this dissertation. Finally, using a parallel corpus approach allows one to cover a larger set of languages than would be possible with experimental methods that have been used traditionally, since it is much less time demanding.

However, there are also a number of disadvantages. The original text might influence the translations in some ways. Patterns that would normally be uncommon might be used more often to accommodate certain features of the original. In addition, only the written register can be researched. However, Wälchli (2007: 132) mentions that many typological studies based on reference grammars might have the same focus on written language sources and are therefore not better off.

The parallel texts that have been chosen are three novels: *Alice's Adventures in Wonderland*, *Through the Looking-Glass and what Alice found there* (both by Lewis Carroll) and *O Alquimista* ['The Alchemist'] (by Paulo Coelho). These books were chosen to have different original languages, English and Portuguese, which have different typological patterns with regard to motion (as will be discussed in detail in chapter 3). This allows for an assessment of whether it makes a difference whether the translation is based on a satellite-framed or a verb-framed original text. In addition, these novels have been translated into a wide range of Indo-European languages, and their continuing popularity enabled their easy acquisition. The choice of three different books with different original languages should also make author and translator specific biases less pronounced. A list of all translations is included in Appendix 1.

The language sample consists of: French, Italian, Portuguese, Romanian [Romance], Irish [Celtic], Dutch, English, German, Swedish [Germanic], Latvian, Lithuanian, Polish, Russian, Serbo-Croatian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Iranian], Modern Greek [Hellenic], Albanian, and Armenian. The sample includes languages from all major Indo-European subgroups except Anatolian and Tocharian, for which only limited data is available and certainly not translations of modern novels. Unfortunately, the sample is biased towards European, non-Indo-Iranian languages. This is mostly due to the ability to acquire translations of the three novels in Indo-Iranian languages, which proved to be rather difficult. The sample is also biased towards languages that follow the major trend in their subgroup. Languages such as Rhaeto-Romansh, Breton, Bulgarian and Ossetic are not included here, while these are known to be different from the major trends in Romance, Celtic, Balto-Slavic, and Indo-Iranian respectively (Wälchli 2009: 215). This could mean that the current sample gives an underrepresentation of rapid diachronic change of motion event encoding in Indo-European – however, the availability of translations restricted the number of choices. The glossed parallel corpus (see below for details on glossing) and the complete set of translations, including translations into several other Indo-European languages not featured in this dissertation, are available upon request.

Given the focus of this dissertation on the Indo-European language family and the availability of data from ancient Indo-European languages such as Latin, Ancient Greek, Vedic Sanskrit, Old Church Slavonic, Gothic, and Classical Armenian, it would have been possible to investigate motion event encoding using corpora that include these ancient languages. This would have enabled a comparison between inferences made by the phylogenetic comparative methods and the attested patterns in the ancient Indo-European languages. The best parallel corpus of both ancient and modern Indo-European languages that is available is the *New Testament*, which is available in Latin, Ancient Greek, Old Church Slavonic, Gothic, and Classical Armenian (Haug and Jøhndal 2008) and of course many modern languages. However, the choice was made not to use this corpus because the *New Testament* is not a particularly rich source for Talmian motion data. The *New Testament* has very few instances of the satellite-framed construction with a manner verb and a path satellite (see example (12) and (14) below) that is diagnostic for the study of the Talmian theory of motion event encoding (see also Beavers et al. 2010: 332; Croft et al. 2010: 221). The satellite-framed construction is diagnostic because it is typically translated with a satellite-framed construction in a satellite-framed language, but translated with a verb-framed construction in a verb-framed language (Slobin 2005b). Manner is typically added in about 25% of the constructions when translating a text in a verb-framed language to a satellite-framed language, whereas manner is deleted in about 50% of the constructions when translating a text in a satellite-framed language to a verb-framed language (Slobin 1996b: 212). Examples (12)–(14) illustrate this process: the English original in (12) is translated with a verb-framed construction in verb-framed Romanian in (13), and with a satellite-framed construction in satellite-framed German in (14).

12) English

and (she) hurried of to the garden door.

13) Romanian

și	*se*	*îndreaptă*	*în*	*grabă*	*spre*
and	REFL.3SG	go.toward.PRS.3SG	in	rush	toward

uș-a	*grădin-ii.*
door-F.ACC.SG.DEF	garden-F.GEN.SG.DEF

‘and (she) went toward the garden door in a rush.’

14) German

und	*eilte*	*fort*	*zu*	*der*	*Gartentür.*
and	hurry.PST.3SG	off	to	DEF.ART.F.DAT	garden.door.F.DAT

‘and (she) hurried off to the garden door.’

Given this discrepancy in translating motion events, a parallel text needs instances of both the satellite-framed construction as well as instances of the verb-framed construction if it is to be used for the optimal characterization of motion event encoding in a given language. The *New Testament* unfortunately cannot offer both. This is in part due to a lack of manner verbs, as well as to a tendency to use manner verbs in descriptions of activities (i.e. ‘The man walked’) instead of descriptions of motion events (i.e. 'The man walked into the room'). An illustration of these issues in the *New Testament* and a comparison with a corpus of the two English *Alice* novels by Lewis Carroll is given in Appendix 2.

Another problem with the use of Bible texts is that they are written in a religious register that restricts translational freedom to at least some extent. The study of cross-linguistic motion event encoding has been done using parallel corpora in the past (Slobin 1996b, 2005b; Baicchi 2005), and one of the findings has been that in translations from a verb-framed language to a satellite-framed language, manner information may be added, while in translations from a satellite-framed language to a verb-framed language, manner information is often deleted – this is done in order to approximate the native motion encoding patterns. The religious convention to translate the Bible in such a way that it stays close to the original text might interfere with the artistic freedom that would be needed to translate motion events in the most natural way. Given these disadvantages associated with the use of the *New Testament* for motion event encoding, the decision was made to use the parallel corpus consisting of the three modern novels mentioned above.

2.1.2 Building the parallel corpus

The parallel corpus was built using translations of *Alice's Adventures in Wonderland*, *Through the Looking-Glass and what Alice found there* and *O Alquimista* ['The Alchemist']. First, all descriptions of motion events were extracted from these three novels. Motion events were defined as "situations in which an animate being moves from one place to another" following Özçalışkan and Slobin (2003: 259), although inanimate entities were included as well. Each motion extract that was picked constituted a single sentence in which (approximately) a single situation (event or activity) was being described (Berman and Slobin 1994: 657). Such a sentence could consist of several clauses, as will be seen in (19) below. However, there was always a single clause, i.e. a single combination of a subject and a predicate, which functioned as the main motion predicate of that sentence. In the case of (19), this was *floated*. Examples of these motion extracts can be found throughout this dissertation.

This selection procedure resulted in a set of 1270 motion event descriptions in the three novels. From this set, a smaller set of motion event descriptions was picked out. As including all descriptions of motion events found in the three novels would have resulted in a far too large dataset, a balanced sample of motion event descriptions was taken. This sample was constructed taking into account the type of motion event encoding construction and the main motion verb. At least one instance of each attested motion verb was included in order to maximize the lexical diversity present in the sample. In addition, a large and balanced set of different motion event encoding construction types was taken. This selection was informed by what we know about motion events from Talmy (1985, 1991, 2000) and Slobin (1996b, 2004): care was taken to select verb-framed and satellite-framed constructions, as these are diagnostic for the study of motion event encoding (see section 2.1.1). Even though subsequent research has indicated that languages typically use a diverse range of motion event encoding constructions (see section 1.1 and 3.2), these two constructions, although they may be rare in spoken discourse, remain very important for the study of motion event encoding. In addition, we expect that the usage of these constructions correlates with the use of other constructions and aspects of motion event encoding. The aim was to capture an broad and informed picture of motion event encoding as defined by Özçalışkan and Slobin (2003: 259).

Because of the emphasis on including all the attested variation, the choice of the motion sentences was not done on a randomized basis. The resulting picture that emerges from this smaller set does therefore not give a complete picture of the encoding of motion in each language, but a biased one. However, it does serve to provide a picture of the encoding of motion for each individual language relative to each of the other languages in the sample. The main aim of this study was to be able to draw exactly such a picture for each language and to assess as much verb variability as possible.

The smaller set of selected motion sentences amounted to 215 sentences that encode voluntary (non-causative) motion that were distributed among the three novels. Unfortunately, *Through the Looking-Glass and What Alice Found There* is not available in four languages: Albanian, Hindi, Nepali and Persian. To accommodate this, there are two different sentence samples listed in Table 2.1 that are used for the analysis of syntactic patterns of motion event encoding. The 118-sentence sample includes only sentences from *Alice's Adventures in Wonderland* and *O Alquimista*, and is available for all 20 languages. The 192-sentence sample also includes sentences from *Through the Looking-Glass and What Alice Found There*, and is available for 16 languages (excluding Albanian, Hindi, Nepali and Persian). The number between brackets in Table 2.1 gives the number of sentences taken from each of the novels.

***Table 2.1**: Sentence samples used for syntactic motion event encoding*

Sample	**Alice's Adventures in Wonderland**	**O Alquimista**	**Through the Looking-Glass and What Alice Found There**
118-sentence sample (20 languages)	yes (73)	yes (45)	no
192-sentence sample (16 languages)	yes (73)	yes (45)	yes (74)

In addition to these two samples, additional sentences were added for the study of manner of motion verbs. These sentences did not include a path of motion, and therefore fall outside of the definition of 'motion event' used in this dissertation (see section 2.2). However, since the addition of so-called manner only sentences (see again section 2.2) allowed for the collection of a larger set of manner of motion verbs, two more samples are distinguished in Table 2.2. The 132-sentence sample is created by adding 14 manner only sentences (6 from *Alice's Adventures in Wonderland* and 8 from *O Alquimista*) to the 118 sentence sample, and is available for all 20 languages. The 215-sentence sample is created by adding 23 manner only sentences (the same 6 from *Alice's Adventures in Wonderland* and 8 from *O Alquimista*, plus 9 from *Through the Looking-Glass and What Alice Found There*) to the 192 sentence sample, and is available for 16 languages (excluding Albanian, Hindi, Nepali and Persian).

***Table 2.2**: Sentence samples used for manner of motion verbs*

Sample	Alice's Adventures in Wonderland	O Alquimista	Through the Looking-Glass and What Alice Found There
132-sentence sample (20 languages)	yes (73+6)	yes (45+8)	no
215-sentence sample (16 languages)	yes (73+6)	yes (45+8)	yes (74+9)

In total, the dataset consists of 3968 motion sentences. The complete set of original motion extracts that were used in these samples is listed in Appendix 3. Table 2.3 lists the usage of the sentence samples in each of the chapters.

***Table 2.3**: Usage of sentence samples per chapter*

Chapter	Sentence samples used
Chapter 3	118-sentence sample
Chapter 4	118-sentence sample
Chapter 5	118-sentence sample, 192-sentence sample, 132-sentence sample, 215-sentence sample
Chapter 6	132-sentence sample

After the sample of motion event descriptions was decided upon, the original and translated sentences were glossed with the help of either native speakers or language specialists (see Acknowledgements). The Leipzig glossing rules were used as guidelines for the glossing. The glossing was done in order to understand the translation and as a starting point for an analysis of motion encoding in these languages. In addition, a native speaker helped to explain verb semantics. This person helped to categorize each motion verb that was attested as a manner verb, a path verb, a deictic verb, or a manner plus path verb. The coding of other motion event elements, such as prepositions, adverbs, and case markers, was done using grammars. The originals and translations were coded for the motion event encoding features that will be discussed in section 2.2.

2.2 Motion events

2.2.1 Introduction

The current approach to the analysis of motion expressions is heavily dependent on the ideas developed by Leonard Talmy (Talmy 1985, 1991, 2000), which will be discussed to the extent that they are relevant for the purposes of this dissertation. Specifically, it focuses on aspects of the theory and terminology developed most succinctly in Talmy (1991: 486-490) on motion events, while not taking into account the other types of framing event proposed by Talmy (1991). Talmy's framework for studying motion centers around the idea that abstract semantic concepts are encoded by different linguistic surface structures in different languages. This idea is illustrated by the difference between (15) and (16). In the English sentence in (15), the way in which the Knight is moving is indicated by the main verb of the sentence, *ride*, while it is indicated by the adverbial-like gerund *cavalgando* in the Portuguese translation in (16). The same semantic information, namely the specific way in which the Knight is moving, is encoded by different types of linguistic elements in the two languages. In principle, then, different semantic components may be expressed with a set of different lexical expressions, which in turn are combined to form a range of different syntactic motion event constructions.

15) …, and then the Knight rode slowly away into the forest.

16) Portuguese

e	*o*	*Cavaleiro*	***afastou-se,***
and	DEF.ART.M.SG	knight.M	move.away.IND.PFV.3SG-REFL
cavalgando		*lentamente*	*pela*
ride.horseback.PRS.PTCP		slow.F . ADV	through . DEF.ART.F.SG
floresta.			
forest.F			

'And the Knight moved away, riding slowly through the forest.'

The current approach is also influenced by research on motion events subsequent to Talmy's work, which has moved away from the strict dichotomy between satellite-framed and verb-framed languages proposed by Talmy. Slobin and Hoiting (1994: 498-499) and Slobin (2004, 2005b, 2006) set up a continuum of manner salience in which manner salience is defined as "the level of attention paid to manner in describing events" in actual language use (Slobin 2006: 64). This approach leads to an understanding of manner expression in motion event encoding in terms of a gradient or scale. The placement of each language on the scale depends on the linguistic tools, i.e. the constructions, the language has available. The idea that motion event encoding is more varied than can be accounted for using a dichotomy is developed further by Croft et al. (2010) and Beavers et al. (2010): "Talmy's typological classification applies to individual complex event types within a language, not to languages as a whole." (Croft et al. 2010: 202). This section introduces the coding scheme used to analyze motion event encoding in this dissertation, which is based on a review of best practices in the motion event encoding literature.

2.2.2 Conceptual elements of motion events and their lexical expression

There are four main components of motion that were distinguished by Talmy (1985, 1991, 2000) and that are taken into account in this thesis as well: figure, path, ground and manner.[3] In summary: we observe a person or object that moves (figure), the path or direction that he takes (path), reference objects in the environment (ground), and the way in which he moves (manner). Languages may choose to encode these components in different ways, and they may choose not to encode some of these components at all. Each of these semantic components of motion and their possible lexical encodings will be considered in turn.

THE FIGURE can be defined as the entity that moves. In example (15) and (16), the Knight is the figure. The figure can be human, animal or inanimate, and it can be linguistically encoded in many different ways (by proper nouns, noun phrases, pronouns, etc.). In my sample, most figures are human, while there is a small subset of animal and inanimate figures.

THE PATH is the trajectory the figure follows while moving. In example (15) and (16), the path is the trajectory of the movement of the Knight, who is moving from an undefined place towards and into the forest. In my framework,

---

[3] Cause is not listed as one of the categories here, because caused motion is a related but different domain of inquiry that will not be discussed in this dissertation. Motion, also one of Talmy's primary concepts, is not listed here either because Talmy only needs this concept to differentiate motion events from 'stative' placement events. Since this dissertation focuses exclusively on motion events that describe transitional motion, it is not necessary to include it.

path (or deixis, see below) should always be encoded linguistically for a motion expression to count as a motion event.[4]

In Talmy's framework, path can either be expressed in the verb or in the satellite. Talmy (1985: 102) defined (path) satellites as "certain immediate constituents of a verb root other than inflections, auxiliaries, or nominal arguments." Several researchers, including Filipović (2007: 35), Beavers et al. (2010: 337), and Croft et al. (2010: 205-206), take issue with Talmy's (1985) criterion to distinguish path satellites from prepositions in English. Talmy (1985) states that if the ground can be left out, as in (17) below, the path element is a satellite, and if it cannot be left out, as in (18), the path element is not a satellite. Beavers et al. (2010: 338) point out that the sentences in (17) and (18) are functionally equivalent. Both "indicate the goal of motion and often they are apparently alternate expressions of the same semantic content" (Beavers et al. 2010: 338). In (17), as Filipović (2007: 35) also points out, even if no ground is mentioned, one would be inferred from the context. Talmy's (1985) diagnostic therefore does not seem justified from a functional semantic perspective.

17) John ran in (the house).

18) John ran to *(the store). Beavers et al. (2010: 338)

Following Filipović (2007) and Beavers et al. (2010), the strict definition of satellite as put forward by Talmy (1985) is rejected for the purposes of this dissertation. In its place, a broader definition is used: path satellites are all non-predicative elements that indicate (a part of) the path of the movement of the figure. This includes adpositions, adverbs, case markers, verbal prefixes, etc.

Aside from the use of path satellites, path can be expressed by two different types of verbs. It can be expressed in path verbs (such as English *enter* and *exit*), and manner plus path verbs (such as Lithuanian *kopti* 'climb up' and French *escalader* 'climb up'). The category of manner plus path verbs will be further discussed below.

A category of verbs that is often subsumed under the category of path verbs are the deictic motion verbs (Berthele 2006: 108). Deictic motion verbs are

---

[4] Note that viewing path as an obligatory component of motion is theory dependent. However, it is a useful idea because it allows for differentiation between movement that occurs at approximately the same place, such as movement of a squirrel on a treadmill, or movement of a baby around the room, and movement that results in a change of location. Even though the movement of a baby that is crawling around the room clearly has a path, when we say 'the baby crawled around the room' we are not specifying the path that the baby had, but only the location of the motion ('the room'). In other words, we are saying that the baby crawled inside the room, and no change of location has occurred. Such expressions are not part of the definition of a motion event that is used in this thesis. Positing path as an obligatory component of motion allows one to distinguish between this type of expression and an expression that refers to a change of location through movement, i.e. a motion event.

verbs that refer to motion with respect to a deictic center, rather than motion that has a certain path. The English verbs *come* and *go* are examples: *come* implies movement towards the speaker (*Lisa came home*) whereas *go* implies movement not towards the speaker (*Lisa went home*). Berthele (2006) points out why deictic verbs should be separated from path verbs: deictic verbs have very different semantics from path verbs, and since in many languages deictic verbs are the most common motion verbs, to count them as path verbs would skew the analysis. Following his lead, deictic verbs are separated from path verbs in this dissertation. Deixis is a complicated issue and is characterized by very different solutions cross-linguistically (Wilkins and Hill 1995; Wälchli 2009: 230ff). Even among related languages, the semantics of deictic verbs can be quite different (Ricca 1993) and a full inquiry would therefore take up too much space here. Therefore, a simple list of deictic verbs encountered in the sample is provided in Table 2.4. In this Table, and in the other Tables and most Figures in this dissertation, the languages under discussion are ordered first according to subgroup and then in alphabetic order. The order of subgroups is Romance, Celtic, Germanic, Balto-Slavic, Indo-Iranian, Hellenic, Albanian, and Armenian, reflecting the order on which they are presented on the phylogenetic tree in Figure 1.7 as well as a rough west-to-east division.

Table 2.4 lists the Balto-Slavic languages Russian, Polish, and Lithuanian as having no deictic verbs. It would be possible to list Russian *idti*, Polish *iść*, Latvian *iet*, and Lithuanian *eiti*, which are often translated as 'go', as deictic verbs. However, these verbs are in fact neutral with respect to deixis. Specific deixis can be added using verbal prefixes, such as Russian *pod*- and *ot*-. This is also true for Serbo-Croatian, in which *doći* 'to come' has lexicalized from the combination of *do-ići*. Consultations with native speakers suggest that the verbs *idti*, *iść*, *iet*, and *eiti* express some kind of 'prototypical' or 'general' motion, which is most often used in the context of human agents and is then interpreted as expressing walking motion. However, most of these verbs can also be used in other contexts, for instance for the movement of trains. In the current dataset, these verbs most often occur in the context of moving human agents, and can therefore be said to mean 'walk' in those contexts. Therefore these verbs have been classified as manner verbs. See for some discussion of the Russian verb *idti* as a generalized motion verb Nesset (2009) and Dickey (2010) and for more general discussion on the identification of deictic verbs Wälchli (2009: 230ff, 2001a: 311ff).

THE GROUND is defined as an explicitly indicated object that serves as a reference point for the motion in which the figure is engaged. In example (15) and (16), the forest functions as the ground of motion. It can be any type of object, from buildings to forests, and from people and animals to household objects. The ground can also be an extended area or place, such as the air or the sea.

***Table 2.4**: Deictic verbs encountered in the sample.*

Language	Deictic verbs	Reference
French	venir, aller	Ricca (1993)
Italian	venire, andare	Ricca (1993)
Portuguese	vir, ir	Ricca (1993)
Romanian	veni, merge	no reference
Irish	tar, gabh, téigh	Ó Baoill (1975)
Dutch	komen, gaan	Ricca (1993)
English	come, go	Ricca (1993)
German	kommen, gehen	Ricca (1993)
Swedish	komma, gå	Viberg (2006)
Latvian	nākt	Wälchli (2001b: 414)
Lithuanian	no deictic verbs	
Polish	no deictic verbs	
Russian	no deictic verbs	
Serbo-Croatian	doći, ići	Gathercole (1978)
Hindi	ānā, janā	Kachru (2006: 86-87)
Nepali	aunu, jānu	Gathercole (1978)
Persian	āmadan, raftan	Feiz (2011)
Modern Greek	erchomai, pigaino	Ricca (1993)
Albanian	vij, shkoj	Ricca (1993)
Armenian	gal, gnal	no reference

THE MANNER is defined as the way in which the action can be carried out. In example (15) and (16), the verb *ride* indicates the manner of motion. Manner is a component of motion that can be explicitly encoded or not. Different languages pay different amounts of attention to encoding manner of motion, as has been pointed out by Slobin (2004) and others.

For manner, a broad definition is employed that includes every linguistic element that indicates something about the way in which the figure is physically moving. Manner can be expressed by four different categories. First, there are manner verbs such as English *trot*, *run* and *fly*. Second, there are manner plus path verbs such as such as Lithuanian *nudrožti* 'move away speedily' and French *escalader* 'climb up'. This category will be further discussed below. Third, there are adverbial manner expressions. These can be adverbs or other adverbial expressions that directly indicate aspects of manner, such as *gently* in (19). These adverbial manner expressions can also be descriptive phrases that encode aspects of manner. An example of the latter type is given in (19), where the phrase *without even touching the stairs with her feet* indicates the manner in which the agent floated down.

19) … and [she] floated gently down without even touching the stairs with her feet

Fourth, manner participles may be used to encode manner. Manner participles are used in the verb-framed strategy that consists of a path verb plus a manner participle. An example was given in (16): The main verb *afastar-se* 'move away' indicates the path of the movement while the participle of the verb *cavalgar* 'ride horseback' indicates the manner of motion.

The last lexical category to be discussed is that of the manner plus path verb. This type of verb expresses both manner and path at the same time. Slobin (2004: 230-231) discusses the Turkish manner plus path verb *tırmanmak* 'climb up' and points out that it is readily used in contexts that require expression of both manner and path. This Turkish verb is semantically different from English *climb*, since English *climb* can also be used for downwards motion. Likewise, Zlatev and Yangklang (2004: 167-168) distinguish a class of path plus manner verbs in Thai. Some of the languages in the current sample have manner plus path verbs. An example is Modern Greek *skarfalono* 'climb up', which expresses both upward motion and a climbing manner.[5]

Although the differences between path verbs, manner verbs, and manner plus path verbs seem clear when they are defined in the previous discussion, classifying motion verbs can be difficult. Many verbs are intermediate between the two categories, such as English *climb* and Dutch *klimmen*, which can be used for all kinds of paths, including up, down, into, and out of, but which without further specification of direction indicate movement upwards. Indeed, Fillmore (1982: 32) and Taylor (1989: 105-109) write that English *climb* has two distinct attributes to its meaning: 'clambering' and 'ascending'. The sometimes complex semantics of motion verbs make it difficult to assign them to delimited categories. In many languages, the classification of the verb meaning 'fall' is also very problematic, since in some way it specifies a manner of descending (in the sense that it is involuntary), but at the same time it can often be specified for speed (slowly, quickly) and other aspects of manner. In my classification, the definition of a manner verb is that is can be used with different types of path - English *climb* and Dutch *klimmen* are therefore classified as manner verbs. The

[5] The existence of manner plus path verbs has also been questioned. Jones (1983) and Beavers et al. (2010: 357ff) argue that manner plus path verbs do not exist. Beavers et al. (2010) posit that verbs may only lexicalize manner or path, but not both at the same time. They support this with a range of theoretical arguments, but do not consider any empirical data for this claim. Jones (1983: 178) writes the following: "The idea is that there are general limitations on the possible combinations of semantic components which can define the meaning of a verb and that, in particular, if a verb expresses movement, it may also contain either a vectorial feature or a feature (or set of features) describing the manner in which movement took place, but not both." However, Jones (1983: 179) immediately runs into problems with several French verbs that do seem to express both path and manner. The existence of manner and path verbs therefore seems a question that needs empirical scrutiny rather than more theorizing.

definition of a path verb is that it can be specified for different types of manner - most verbs meaning 'fall' are therefore classified as path verbs. The definition of a manner plus path verb, correspondingly, is that it is specified for a single manner and a single path - such as Modern Greek *skarfalono* 'climb up', Persian *goriḵtan* 'run away', Dutch *duiken* 'dive into', and Italian *arrampicarsi* 'climb up'. By looking at the possible usages of these verbs in different contexts, a 'core' meaning can often be discerned, although polysemy between several (slightly) different meanings will continue to be a problem. Another problem that is difficult to solve is that manner verb and path verb classes seem to adhere to prototype theory: some manner verbs and some path verbs are more central than others. For manner verbs, these are probably RUN, FLY, and SWIM, whereas for Talmian path verbs these are probably ENTER, EXIT, ASCEND and DESCEND. Less prototypical members of the manner verb and path verb classes will be more difficult to classify than more central members.

An overview of the surface structures that were discussed in this section is presented in Table 2.5.

Table 2.5: *Motion event encoding components distinguished in this dissertation*

Nature	Component	Semantics	Examples
verbal	manner verb	indicates the way in which a person or an object moves	*walk, run*
	path verb	indicates the path or trajectory of motion	*enter, descend*
	deictic verb	indicates the path of motion as seen from a deictic center	*go, come*
	manner plus path verb	indicates both manner and path of motion	Greek *skarfalono* 'climb up' Persian *goriḵtan* 'run away'
	neutral verb	indicates none of the above	*move, travel, find oneself, continue*
non-verbal	path satellite	indicates the path or trajectory of motion	*back, from, into*
	manner expression	is an adverb or manner verb participle that signifies manner, i.e. the way in which a person or an object moves	*slowly, quickly, running, swimming*

To summarize: in the theoretic framework employed in this dissertation, there are four semantic motion elements, namely figure, path, ground and manner. Of course, this is a huge simplification, as a much more fine-grained semantic coding of motion event encoding components would be possible (Frawley 1992).

In any case, languages make different choices with regard to the lexical coding of these features in their surface structures. The choices that they make with regard to the linguistic encoding of manner and path result in different motion encoding constructions. These will be discussed next.

2.2.3 Motion event encoding constructions

The combination of the motion event components presented in Table 2.5 resulted in a set of motion event encoding constructions that are featured in this section. Most of these constructions are familiar from the literature (Talmy 1985, 1991; Croft et al. 2010). The approach taken here is inspired by construction grammar (Goldberg 2006) and distributed spatial semantics (Sinha and Kuteva 1995). I consider the motion event encoding constructions discussed in this section to be salient entities that speakers draw on again and again to encode translational motion. Part of their meaning and pragmatics arises from the combination of the motion event encoding components within constructions. However, identifying specific constructions also means abstracting away from some of the peculiarities caused by the fact that spatial semantics is often distributed over a set of elements: the path in a sentence can be encoded by 1 or by 5 elements, if the main and only verb of the sentence is a manner verb, it is a satellite-framed construction.

The starting point will be motion event encoding strategies that encode motion in a single clause, with a single main verb. The two most often discussed strategies of this type are the satellite-framed construction and the verb-framed construction (Talmy 1985; Slobin 2004; among others). In the satellite-framed construction, manner is encoded by the main verb of the sentence, while path is encoded by a path satellite. Examples are provided in (15) and (20). In verb-framed constructions, path is encoded by the main verb of the sentence, while manner is encoded by an adverbial expression or participial verb form. Examples are provided in (16) and (21).[6]

[6] Note that the verb-framed construction may feature path satellites, such as Romanian *după* 'after' in example (21). The same applies to the path-only construction, deictic verb construction and the deictic verb-framed construction that will be introduced below. For the purposes of this thesis, the semantics of the main verb has been regarded as the most relevant criterion for classifying motion event encoding constructions. Although path-satellites may occur in the majority of the constructions distinguished here, their classification is determined most importantly by which motion element (deixis, path, manner, both manner and path, or none) is encoded by the main verb.

20) German

Die	*Königin*	*breitete*		*ihre*
DEF.ART.F.NOM	queen.F.NOM	spread.3SG.PST		3SG.F.POSS
Arme	*wieder*	*aus*	*und*	***segelte***
arm.M.ACC.PL	again	out	and	glide.3SG.PST
hinterher				
after				

'The Queen spread her arms out again and sailed after [it].'

21) Romanian

Regin-a		*întin-se*	*iarăși*	
Queen-F.NOM.SG.DEF		stretch-PRET.3SG	again	
braț-ele	*și*	***plecă***	***în***	***zbor***
arm-N.ACC.PL.DEF	and	leave.PRET.3SG	in	flight.N.ACC.SG
după	*ea*			
after	3SG.F.OBJ			

'The Queen spread her arms again and sailed after it.'

Then the next two strategies leave out either path or manner. If manner is not present, we have a path-only construction in which path is encoded on the main verb. Examples are provided in (22) and (23).

22) Armenian

Na	*viravor-v-ac*	*otk'-i*	*el-av*
3SG.SBJ	insult-ANTIC-RES.PTCP	foot-DAT	stand-AOR.3SG
u	***heṙ-ac'-av***.		
and	go.off-AOR-3SG		

'Insulted, she got up and went off.'

23) Nepali

dosro	*din*	*keto-le*	*sivir*	*najikai-ko*	*thulo*
second	day	boy-ERG	camp	close-GEN	big
bhir-ko	*māthi-tira*	***ukli-yo***			
cliff-GEN	top-DIR	climb.up-PST.3SG			

'On the second day, the boy climbed to the top of the cliff near the camp'

If path is not present and only manner is encoded, we have a manner-only construction in which manner is encoded on the main verb. In the current analysis, these examples do not count as motion events. However, since they are encountered occasionally, they are included in the discussion here. An example is provided in (24).

24) Persian

ālis	*ham*	***be***	***sor'at-e***	***bād***	***harekat***	***kard***
Alice	also	to	speed-of.EZ	wind	movement	do.AUX.PST.3SG

'Alice also moved as rapidly as the wind.'

If a manner plus path verb is the main verb of the sentence, and there is no other verbal indication of path or manner, the manner plus path verb strategy is employed. Examples are provided in (25) and (26).

25) French

Alice	*contempl-a*	*le*	*Roi*	*Blanc*
Alice	watch-PRET.3SG	ART.DEF.M	king.M	white.M
qui	***escalad-ai-t***	*pénible-ment*	*la*	*grille*
that	climb-IPFV-3SG	with.difficulty-ADV	ART.DEF.F	bar.F

'Alice watched the White King as he climbed the fender with difficulty'

26) Italian

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

'and stepped over the first of the six small brooks with one step.'

When a deictic verb is the main verb, we have an instance of the deictic verb strategy, exemplified in (27). Since deictic verbs can be used with manner expressions, a special class of verb-framed patterns with a deictic verb as the main verb was distinguished from verb-framed patterns with a path verb as the main verb. An example of such a deictic verb-framed construction, which has a deictic verb as the main verb and either an adverbial or a participial manner expression, is provided in (28).

27) Irish

arsa		*Eilís*	*go*	*han-mhúinte*	*agus*	*í*	
say.PST		Alice	ADJ.PART	polite	and	3SG.F.OBJ	
ag	***dul***	*trasna*		*an*	*tsrutháin*	*bhíg*	*i*
at	go.INF	over.DEF.ART		DEF.ART	brook	little	in
ndiaidh		*na*		*Banríona*			
pursuit		DEF.ART.GEN		Queen			

‘Alice said politely and she went over the small brook after the Queen’

28) Dutch

Het	*was*	*het*	*Witte*	*Konijn*	*dat*	*weer*
3SG.N	COP.PST.SG	DEF.ART	white	rabbit	that	again
langzaam	***kwam***	***aan-getrippeld***				
slowly	come.PST.SG	towards-patter-PTCP				

‘It was the White Rabbit that was coming back slowly trotting’

There are also two constructions attested in the current sample that employ two clauses to encode motion events. The first of these is the coordinate strategy (Croft et al. 2010: 207-208). An example from Albanian is included in (29). This is a translation from the English: ‘and then [the soldiers] quietly marched off after the others.’ In the English original there is a single manner verb, *march*, while in the Albanian translation, there are two verbs that are coordinated with *e* ‘and’, *iki* ‘to go’ and *bashkohem* ‘to join’. Note that all reference to the manner of motion has been removed in the Albanian translation.

29) Albanian

pastaj		***ikën***	*të*		*qetë*
afterwards		go.PST.3PL	DEF.M.NOM.PL		quiet.M.DEF.NOM.PL
e	***u***	***bashkuan***	*me*	*të*	
and	REFL	join.PST.3SG	with	DEF.M.NOM.PL	
tjerët.					
other.M.DEF.NOM.PL					

‘afterwards, they went quietly and followed the others’

Another example of the coordinate strategy is included in (30). This is the Hindi translation of the English original: ‘and the whole party swam to the shore.’ In the Hindi translation, the manner of motion is preserved by the verb *tairnā* ‘to swim’. The second verb gives the path of motion: *baṛhnā* ‘to advance’.

30) Hindi

…sabse	*āge*	***tair***	***rahī***	
…of.all.SUP	in.front.ADV	swim.INF	PROG.F.SG	
thī	*aur*	*yah*	*pūrī*	
be.AUX.PST.F.SG	and	3SG.PROX	complete.ADJ.F	
jamāt	*tālāb*	*ke*	*kināre*	*kī*
party.F	pond.M	GEN.M.OBL	bank.M.OBL	GEN.F
taraf	***baṛh***	***rahī***	***thī***	
side.F	proceed.INF	PROG.F.PL	be.AUX.PST.F.PL	

'…she was swimming in front of everyone and the whole party was proceeding to the bank of the pond'

From the current sample, a construction emerged that has not been discussed in the motion event literature as of yet. This is the subordinate strategy, in which there is one main verb and one subordinate verb that both encode aspects of the motion that is involved. An example from Modern Greek is provided in (31).

31) Modern Greek

… to	*Leyk-o*	*Vasilia*	*poy*
DEF.ART.M.ACC.SG	White-M.ACC.SG	King.M.ACC.SG	who
paley-e	*sig-a-sig-a*	*na*	
struggle-PST.IPFV.3SG	slowly-ADV-slowly-ADV	to	
skarfalos-ei			
climb.up.DEP-3SG			

'the White King, who was struggling slowly to climb up [a fire fender, AV]'

In this subordinate construction, two verbs are involved that do not have equal status, i.e. there is one 'main' verb and one 'subordinate' verb. In (31), the main verb is *paleyo* 'struggle' and the subordinate verb is *skarfalono* 'climb up'. In this example, both verbs express aspects of the manner of motion, while path is encoded by the second verb. This second verb, *skarfalono* 'climb up', carries dependent verb marking, marking it as having a different status from the main verb *paleyo* 'struggle'.

This strategy differs from the equipollently-framed strategy that was identified by Zlatev and Yangklang (2004) and Slobin (2004). In such constructions, both manner and path are expressed by elements that are "equal in formal linguistic terms, and appear to be equal in force or significance" (Slobin 2004: 228). This strategy is also different from the verb-framed strategy in that it is not necessarily the manner component that is in the subordinate clause. An example would be the English sentence 'he hurried to leave', where the path verb is located in the subordinate clause. In addition, languages like Greek make use of both the verb-framed strategy and the subordinate strategy at the same time – a verb-framed example from Greek is included in (32). The verb-framed strategy in Modern Greek are characterized by using a participle form of the verb, which is different from the dependent verb marking in (31).

32) Modern Greek

… ***e-fyg-e***	***alafropat-ontas:***
PST-go.away.PST.PFV-3SG	walk.delicately-PTCP.ACT

'she left walking gently'

Aside from these motion event encoding constructions, a category that contains 'other' construction types is included as well. This category includes translations with verbs that cannot be classified as a manner verb, path verb, deictic verb, or manner plus path verb. Examples are verbs like 'move' or 'travel', which do not encode deixis, manner, or path. It also includes translations that are very deviant and do not contain the motion event as encoded by the original sentence, or translations that do not include a verb.

A note with regard to both the lexical classification of verbs and path satellites and the constructions built from them concerns the problem of cross-linguistic identification: how does one know whether a satellite-framed construction in Albanian can be compared with a satellite-framed construction in Irish? The only solution for this problem is to base the analysis on semantics and morpho-syntactic function. The semantic verb classifications are based on consultation with native speakers. The morpho-syntactic function of the various elements involved in the motion encoding constructions can be assessed as well: verbs are able to function as predicates by themselves, while participles, adverbs, and path satellites cannot (Croft et al. 2010: 205ff). The different semantics of participles, adverbs, and path satellites serve to distinguish them from each other as well. Taking a perspective grounded in semantics and cross-linguistic functional equivalence helps to diminish the problem of cross-linguistic identification, making sure that constructions are cross-linguistically comparable.

To summarize this section, an overview of the constructions distinguished in this dissertation is presented in Table 2.6.

***Table 2.6**: Motion encoding constructions distinguished in this dissertation*

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

2.3 Motion event encoding measures

As mentioned in section 2.1, the complete dataset consists of 3968 motion event sentences in 20 different Indo-European languages. In order to analyze evolutionary change in this dataset, it was necessary to reduce the dimensionality of this dataset. Two data measures are employed that characterize the motion event encoding system used in each language in a comprehensive way. The measures introduced in this section will be used throughout this dissertation, and there will be references to this section throughout the thesis.

The first measure that is used is the proportion of usage of each motion event encoding strategy on a scale from 0 to 1. As explained earlier, all 3968 motion event sentences were coded for one of the nine motion event encoding strategies introduced in section 2.2. Figure 2.1 presents an overview of the usage of these nine motion event encoding strategies in each of the sampled 20 Indo-European languages. This barplot gives the proportion of usage of each strategy on a scale from 0 to 1 (on the x-axis) for each of the twenty languages (on the y-axis) in the 118-sentence sample. In Figure 2.1, the top bar labeled 'originals' gives the proportion for the original sentences, combining the English sentences from *Alice's Adventures in Wonderland* and the Portuguese sentences from *O*

Alquimista. These are provided to give the reader a sense of the starting point of the parallel corpus, i.e. of the set of motion event encoding constructions that was used in the original texts.

The main finding that emerges from Figure 2.1 is that there exists much variance in strategy usage: some languages use certain construction types often, while others do not. This finding will be discussed at length in chapter 3 – the purpose of introducing the usage proportions in this section is purely to present them as relevant data measures that will be used throughout this dissertation.

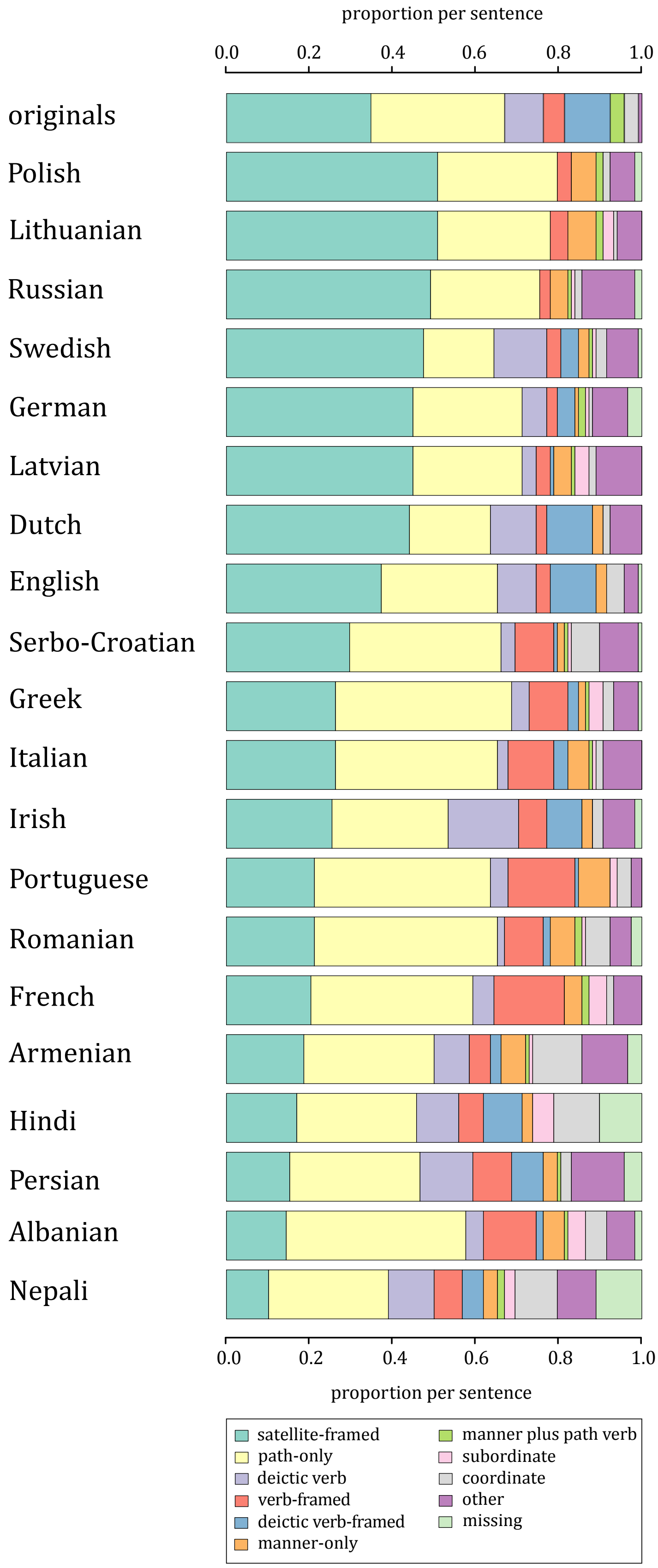


***Figure 2.1**: The frequency of the usage of nine different motion encoding strategies in twenty Indo-European languages*

The second set of data measures that will be used throughout this thesis are the first and second principal components from a principal components analysis conducted on the proportions of usage of each motion event encoding construction as displayed in Figure 2.1. Principal components analysis is a ubiquitous data reduction technique that is used in comparative analyses in biology (Collar et al. 2009), anthropology (Harris and Bailit 1988), and linguistics (Baayen 1994). Studies of morphology in biology often use a large amount of highly correlated measurements - see for examples studies of fish skulls (Collar et al. 2009) or behavior of *Anolis* lizards (Losos 1990). These studies use principal components analysis as a tool to reduce the dimensionality of these measures, and it is used here for the same reasons.

The most important reason for using a principal components analysis to derive a meaningful data measure is the fact that the proportions of motion event encoding strategy usage appear to be heavily correlated. This can already be observed in Figure 2.1: languages that use the satellite-framed strategy often, use the path-only strategy less often, and vice versa. Although the individual proportions of usage are relevant to characterize the motion event encoding system, they can quite easily be represented much more elegantly by the first two principal components of a principal components analysis, as explained below. A third and last reason is that for the co-evolution analyses conducted in chapter 5, a single holistic data measure is required. These analyses simply cannot take into account the nine separate usage proportions in a single analysis, and doing that would not be useful given the covariation between the proportions of usage.

The alternative to principal components analysis would be to classify languages into classes that can be discerned with the naked eye. However, it can already be seen in Figure 2.1 that motion event encoding construction usage is highly variable: each language makes use of each motion event encoding construction to a different degree. For this reason, assigning languages to distinct classes would not be very useful: although Polish and Russian behave similarly, assigning them to the same class would not capture what is different about them. In addition, no natural boundaries that would classify each language to a meaningful distinct class emerge from Figure 2.1 – at least not if the usage of all motion event encoding constructions needs to be taken into account. More discussion of this will be presented in chapter 3.

The principal components analysis was performed on the proportion of usage of each of the nine motion event encoding constructions in the 118-sentence sample and the 192-sentence sample (see section 2.1.2). Since the genealogical relationships between these languages are likely to explain part of the variance present in the data, the phylogenetic principal components analysis proposed by Revell (2009) was used. This analysis removes only a small portion of the variance that can be attributed to these relationships, and still requires the data to be further analyzed with phylogenetic methods:

> "…phylogenetic size-correction and principal components provide estimates of the allometric coefficient and eigenstructure that will have lower variance relative to nonphylogenetic procedures, thus reducing type I error to its nominal level when residuals and scores are subsequently analyzed using phylogenetic methods. If phylogeny is instead ignored in the preliminary transformations, then variance and type I error of our statistical estimators and hypothesis tests can be substantially increased" Revell (2009: 3259-3260).

The results of the phylogenetic principal components analysis are graphically depicted in Figure 2.2 and Figure 2.3.

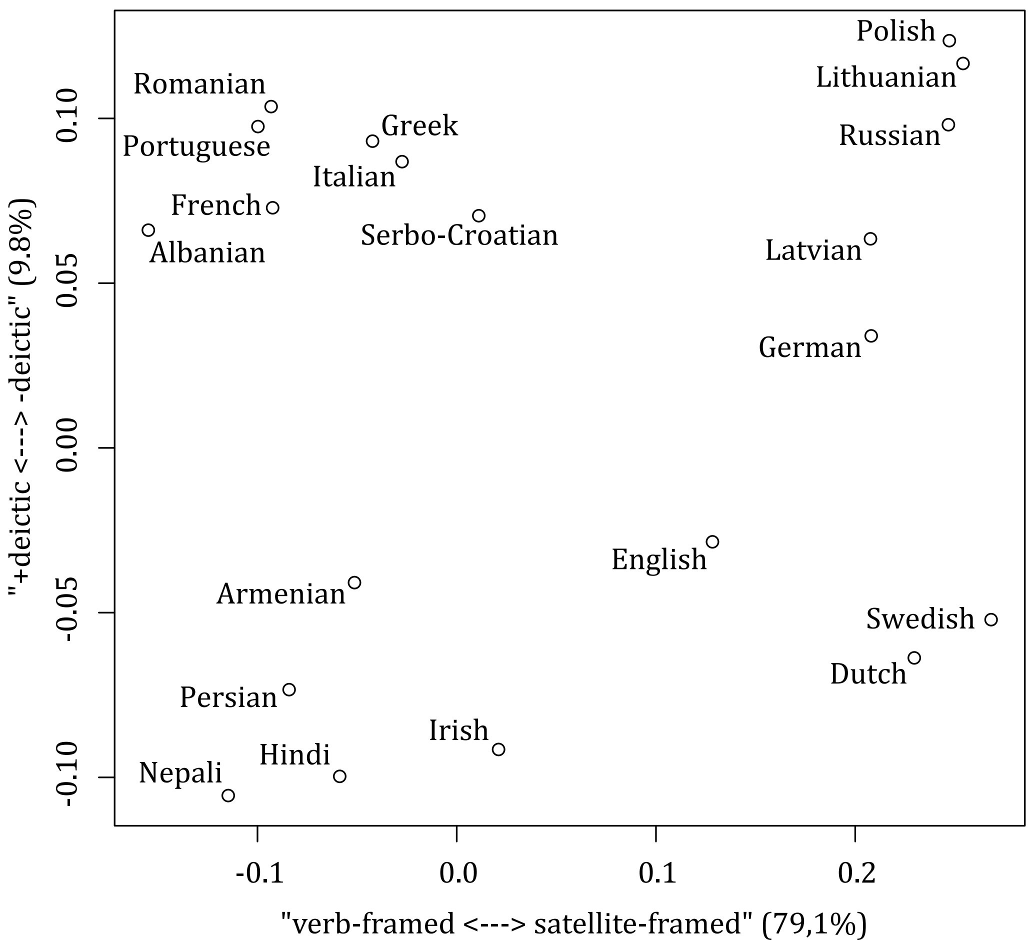


***Figure 2.2**: A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 118-sentence sample for 20 Indo-European languages*

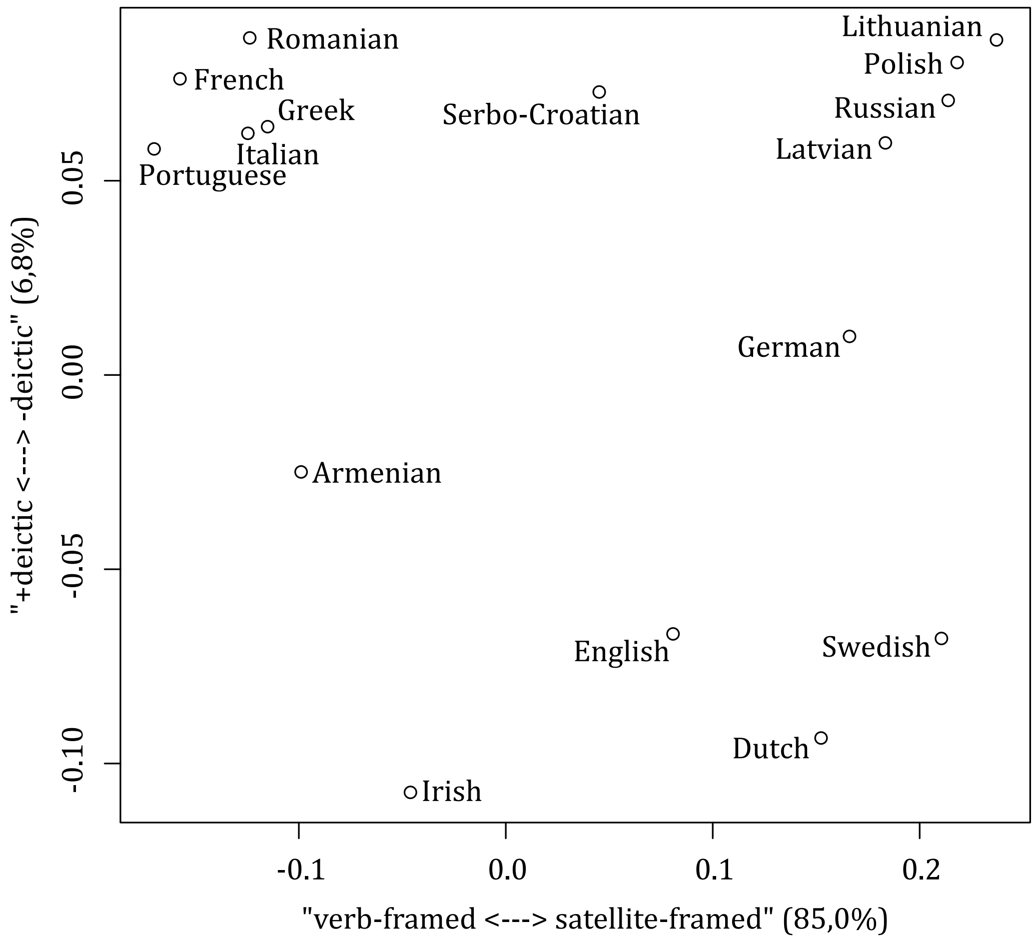


Figure 2.3*: A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 192-sentence sample for 16 Indo-European languages*

In both principal components analyses, the first and the second principal component are the most important and together, they account for a large proportion of the variance. For the principal components analysis conducted on the 118-sentence sample (depicted in Figure 2.2), the first principal component (PC1) given on the x-axis accounts for 79.1% of the variance and can be interpreted to relate the Talmian scale: languages situated in the far right of Figure 2.2 are the most satellite-framed, while languages situated in the far left of Figure 2.2 are the most verb-framed. The second principal component (PC2) given on the y-axis accounts for 9,8% of the variance and relates the amount 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 Figure 2.2. For the principal components analyses conducted on the 192-

sentence sample depicted in Figure 2.3, the PC1 explains 85.0% of the variance, while PC2 explains 6.8%. The same interpretations of the principal components apply.

These two dimensions, interpreted to reflect verb-framed vs. satellite framed character on the PC1 and +deictic vs. -deictic character on the PC2, seem to capture the diversity of the twenty languages under consideration quite well, as together they explain 88.9% (118-sentence sample) and 91.8% (192-sentence sample) of the variance. Note that groups of languages are found in each quadrant of Figure 2.2: the Romance languages, Greek, Albanian, and Serbo-Croatian are verb-framed and -deictic; German, Latvian, Lithuanian, Polish, and Russian are satellite-framed and -deictic; Dutch, English, and Swedish are satellite-framed and +deictic, and Armenian, Hindi, Irish, Nepali, and Persian are verb-framed and +deictic. These dimensions simply capture which strategies are used most and least frequently by the twenty languages, and they are in line with the frequency plot in Figure 2.1. Note that the +deictic vs. -deictic dimension is meant differently than the difference between strictly deictic, mainly deictic, and non-deictic languages found by Ricca (1993). Ricca (1993: 79-91) focuses on the semantics of the contexts in which verbs for ‘go’ and ‘come’ are allowed and restricted, while the current PC2 simply reflects how often a deictic verb is used as the main verb of the motion event description.

The score of each language on the PC1 for both principal components analyses was used as the position of that language on a Talmian scale that reaches from a maximally verb-framed character on the left side of Figure 2.2 and 2.3 and a maximally satellite-framed character on the right side of Figure 2.2 and 2.3. Since they provide a holistic characterization of the motion event encoding system in each language, the PC1 scores for the 118-sentence sample and for the 192-sentence sample have been used for further analyses in this dissertation. For convenience, the PC1 and PC2 scores and the proportions of usage of the nine motion event encoding constructions for the 118-sentence sample have been listed in Table 2.7, and those for the 192-sentence sample have been listed in Table 2.8.

***Table 2.7**: Principal component scores and proportions of construction usage in the 118-sentence sample for 20 Indo-European languages*

Lang.	**Score on first PC**	**Score on second PC**	**satellite-framed**	**path-only**	**deictic verb**	**verb-framed**	**deictic verb-framed**	**manner only**	**manner+path verb**	**subordinate**	**coordinate**
French	-0.09	0.07	0.20	0.39	0.05	0.17	0	0.04	0.02	0.04	0.02
Italian	-0.03	0.09	0.26	0.39	0.03	0.11	0.03	0.05	0.01	0.01	0.02
Port.	-0.10	0.10	0.21	0.42	0.04	0.16	0.01	0.08	0	0.02	0.03
Rom.	-0.09	0.10	0.21	0.44	0.02	0.09	0.02	0.06	0.02	0.01	0.06
Irish	-0.02	-0.09	0.25	0.28	0.17	0.07	0.08	0.03	0	0	0.03
Dutch	0.23	-0.06	0.44	0.19	0.11	0.03	0.11	0.03	0	0	0.02
Engl.	0.13	-0.03	0.37	0.28	0.09	0.03	0.11	0.03	0	0	0.04
Germ.	0.21	0.03	0.45	0.26	0.06	0.03	0.04	0.01	0.02	0.01	0.01
Swed.	0.27	-0.05	0.47	0.17	0.13	0.03	0.04	0.03	0.01	0.01	0.03
Latv.	0.21	0.06	0.45	0.26	0.03	0.03	0.01	0.04	0.01	0.03	0.02
Lith.	0.25	0.12	0.51	0.27	0	0.04	0	0.07	0.02	0.03	0.01
Polish	0.25	0.12	0.51	0.29	0	0.03	0	0.06	0.02	0	0.02
Rus.	0.25	0.10	0.49	0.26	0	0.03	0	0.04	0.01	0.01	0.02
S-C	-0.01	0.07	0.30	0.36	0.03	0.09	0.01	0.02	0.01	0.01	0.07
Hindi	-0.06	-0.10	0.17	0.29	0.10	0.06	0.09	0.03	0	0.05	0.11
Nepali	-0.11	-0.11	0.10	0.29	0.11	0.07	0.05	0.03	0.02	0.03	0.10
Pers.	-0.08	-0.07	0.15	0.31	0.13	0.09	0.08	0.03	0.01	0	0.03
Greek	-0.04	0.09	0.26	0.42	0.04	0.09	0.03	0.02	0.01	0.03	0.03
Arm.	-0.05	-0.04	0.19	0.31	0.08	0.05	0.03	0.06	0.01	0.01	0.12
Alb.	-0.15	0.07	0.14	0.43	0.04	0.13	0.02	0.05	0.01	0.04	0.05
mean	0.05	0.02	0.31	0.32	0.06	0.07	0.04	0.04	0.01	0.02	0.04

***Table 2.8**: Principal component scores and proportions of construction usage in the 192-sentence sample for 16 Indo-European languages*

Lang.	**Score on first PC**	**Score on second PC**	**satellite-framed**	**path-only**	**deictic verb**	**verb-framed**	**deictic verb-framed**	**manner only**	**manner+path verb**	**subordinate**	**coordinate**
French	-0.16	0.08	0.23	0.35	0.04	0.19	0	0.06	0.03	0.03	0.01
Italian	-0.12	0.06	0.25	0.35	0.03	0.14	0.04	0.08	0.02	0.01	0.02
Port.	-0.17	0.06	0.21	0.36	0.04	0.16	0.04	0.07	0	0.01	0.03
Rom.	-0.12	0.09	0.26	0.37	0.02	0.11	0.02	0.09	0.02	0.01	0.04
Irish	-0.05	-0.11	0.28	0.24	0.16	0.08	0.10	0.05	0.01	0	0.02
Dutch	0.15	-0.09	0.46	0.16	0.10	0.03	0.12	0.04	0	0	0.01
Engl.	0.08	-0.07	0.42	0.25	0.09	0.03	0.14	0.03	0	0	0.03
Germ.	0.17	0.01	0.50	0.22	0.06	0.02	0.05	0.04	0.01	0.01	0.01
Swed.	0.21	-0.07	0.52	0.14	0.12	0.04	0.06	0.03	0.01	0.01	0.02
Latv.	0.18	0.06	0.52	0.23	0.03	0.04	0.01	0.05	0.01	0.02	0.01
Lith.	0.24	0.09	0.57	0.20	0	0.03	0	0.09	0.02	0.02	0.02
Polish	0.22	0.08	0.55	0.21	0	0.04	0	0.06	0.02	0	0.02
Rus.	0.21	0.07	0.53	0.20	0	0.03	0	0.07	0.01	0.01	0.02
S-C	0.05	0.07	0.41	0.30	0.03	0.07	0.01	0.04	0.01	0.01	0.04
Greek	-0.12	0.06	0.27	0.36	0.04	0.13	0.03	0.04	0.01	0.03	0.03
Arm.	-0.10	-0.02	0.22	0.25	0.08	0.07	0.02	0.09	0.01	0.01	0.11
mean	0.04	0.02	0.39	0.26	0.05	0.07	0.04	0.06	0.01	0.01	0.03

Chapter 3: Motion events from a parallel corpus

This chapter is a revised version of:
Verkerk, Annemarie. (2014c). Where Alice fell into: Motion events in a parallel corpus. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), Aggregating dialectology, typology and register analysis: Linguistic variation in text and speech *(pp. 324-354). Berlin: Walter de Gruyter.*

The way in which different languages encode motion has been an important topic of investigation in the last few decades. As more data from typologically different languages has become available, the strict dichotomy between satellite-framed and verb-framed languages proposed by Talmy (1985, 1991, 2000) has come under fire (Croft et al. 2010; Beavers et al. 2010). Drawing on a parallel corpus with data from twenty Indo-European languages, this chapter investigates the validity of these categories. Aggregation measures are used to present visual representations of the relationships between the languages in order to show that although some languages fit the category of 'satellite-framed' or 'verb-framed' language very well, others clearly do not. In line with these and other results, the proposal is made that the Talmian classifications only have limited use, and motion research should take into account all motion event construction types used by an individual language when describing motion event encoding.

3.1 Introduction

Scholars of Germanic and Romance languages have reflected on the following types of sentences for many years now:

33) It was the White Rabbit, **trotting slowly back** again, …

34) Portuguese

Era	*o*	*Coelho*	*Branco,*
be.IND.IPFV.3SG	DEF.ART.M.SG	rabbit.M	white.M

regressando	*com*	*pul<inh>o-s*	*vagaroso-s,*
return.PRS.PTCP	with	hop<DIM>-PL	slow.M-PL

'It was the White Rabbit, returning with slow hops.'

In (33) and (34), why is the manner of motion, i.e. the 'trotting' way in which the White Rabbit is moving, expressed by the main verb in English, while it is expressed by an adverbial expression in Portuguese – *com pulinhos vagarosos* 'with small hops'? Why doesn't the Portuguese translator simply translate the English sentence by using the verb *trotar* 'to trot'?

In chapter 2, it was explained that languages may encode motion by means of different lexical elements and motion event encoding constructions. The English sentence in (33) is an example of a satellite-framed construction, while the Portuguese translation in (34) is an example of a verb-framed construction. Talmy (1991) originally classified languages into a set of types, based on whether languages used the satellite-framed construction or the verb-framed construction most naturally and frequently. In this chapter, the central question is whether motion event typology is better framed in terms of types of languages, as proposed by Talmy (1991), or in terms of the range of motion construction types that are used within the language, as has been proposed by Slobin (2004, 2005b, 2006) and later by Croft et al. (2010) and Beavers et al. (2010) (see section 2.2.1). The suggestion is made that looking at rates of usage of motion construction types is the most viable approach. It will be shown that a set of motion event encoding constructions is used to different extents by the languages included in the sample, demonstrating that the Talmy typology is not sufficient to explain all the attested variation in motion event encoding. The suggestion is also made that a first step in analyzing the variability that is encountered in motion event encoding can be to use aggregation methods. These methods provide a visual presentation of the relationships between the different languages, and can be used as hypothesis generators for further inquiry into explanations of these relationships. At the same time, they can be used as tools to discover whether there are distinct typological classes in motion event encoding.

The data are from a parallel corpus of translations of 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). The languages under consideration are French, Italian, Portuguese, Romanian [Romance], Irish [Celtic], Dutch, English, German, Swedish [Germanic], Latvian, Lithuanian, Polish, Russian, Serbo-Croatian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Iranian], Modern Greek [Hellenic], Albanian, and Armenian. Seven of these languages, namely Albanian, Armenian, Irish, Latvian, Lithuanian, Nepali and Romanian, have not been studied before in the motion event literature concerned with regard to the satellite-framed or verb-framed nature of languages. In this chapter, the 118-sentence sample was used (see section 2.1). The total set of data available for this chapter thus consisted of 118 original motion extracts and their translations in a total of twenty languages.

Section 3.2 presents an overview of the usage of the motion event encoding strategies in the twenty Indo-European languages. In section 3.3, some results from different aggregation methods used to provide a more sophisticated view on the data are discussed. Section 3.4 presents the conclusion of this chapter.

3.2 Motion event encoding strategy usage

In Figure 3.1, Figure 2.1 is repeated for convenience. Figure 3.1 presents an overview of the usage of the nine motion event encoding strategies that were discussed in section 2.2. This barplot gives the frequency of usage of each strategy (on the x-axis) for each of the twenty languages (on the y-axis) in the 118-sentence sample. In both Figures 3.1 and 3.2, the twenty Indo-European languages were specifically ordered in decreasing order of use of the satellite-framed strategy. In both Figures, the first bar labeled 'originals' gives the proportion of usage for the original sentences, combining the English sentences from *Alice's Adventures in Wonderland* and the Portuguese sentences from *O Alquimista*. These are provided to give the starting point of the parallel corpus, i.e. the original set of constructions that was used.

From Figure 3.1 it becomes clear that all languages use most of the motion encoding strategies available to them, but do so to different extents.[7] The use of the satellite-framed strategy is most variable, and the use of the path-only strategy is quite substantial in almost all languages. In Figure 3.1, the twenty languages under investigation have been ordered so that a cline with regard to the use of the satellite-framed strategy becomes visible. The satellite-framed strategy is used most often in the Polish and Lithuanian samples, in over half of the sentences attested in this corpus. It is used the least in the Nepali sample, in only 10% of the sentences. The cline in usage of the satellite-framed strategy is paralleled partly by a cline in usage of the path-only strategy, which becomes more common as one moves from the upper part to the lower part of Figure 3.1.

[7] Note again that these ratios cannot be considered to be a full account of motion encoding in the individual languages, as the selection of the originals was not done on a randomized basis. This is especially relevant for the use of the deictic strategy. The deictic verbs, i.e. English *come* and *go* and Portuguese *ir* 'go', were among the most commonly used motion verbs in the original books. However, only a restricted subset of these verbs were selected for this study, and the size of this subset did not take into account the proportion of the deictic verbs with respect to the other types of verbs. If the selection of the sentences would have taken this proportion into account, the deictic strategy would have been much more common.

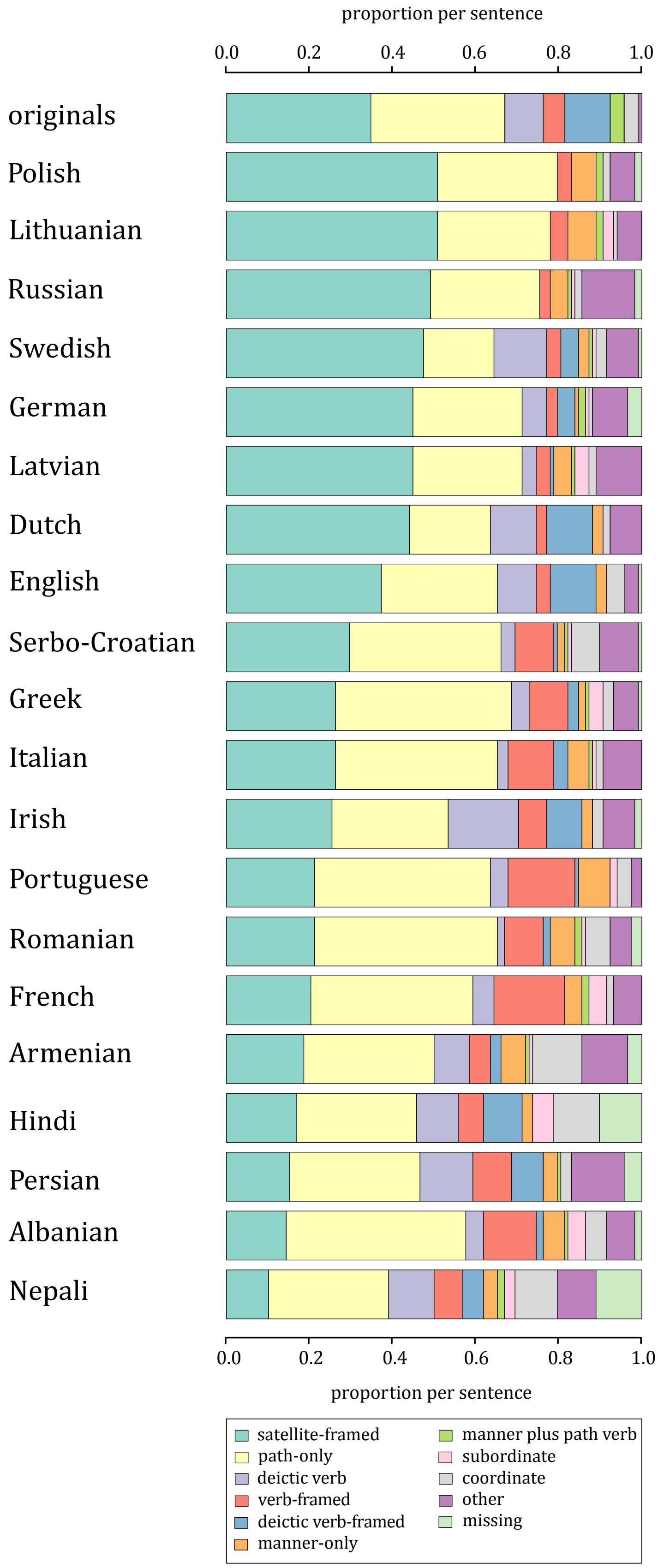


***Figure 3.1**: The frequency of the usage of nine different motion encoding strategies in twenty Indo-European languages*

The use of the deictic verb strategy seems more variable, some languages hardly using deictic verbs at all (Italian), while other languages use them quite often (Irish, Persian). The use of the two types of verb-framed strategies (verb-framed strategies using path verbs or deictic verbs as the main verb in the sentence) is more common on the right side of the plot. This is especially the case for Greek, Italian, Irish, Portuguese, French, Hindi, Persian and Albanian, but not as much for Armenian, Romanian and Nepali. The coordinate strategy is quite often used by Armenian, Hindi and Nepali, while the remaining strategies are less common.

The encoding patterns that are found in the current data set agree with what is known about motion descriptions in these languages. In Table 3.1, an overview of classifications made in the literature on motion events is presented. Several languages in the sample, namely Albanian, Armenian, Irish, Lithuanian, Latvian, Nepali and Romanian, have not been described in the literature on motion encoding before, and are therefore not listed in Table 3.1.

Table 3.1*: Motion encoding classifications made in the literature*

Language	**Classification**	**Source**
French	verb-framed	Jones (1983); Kopecka (2006); Pourcel and Kopecka (2005)
Italian	verb-framed	Folli (2008); Iacobini and Masini (2006)
Portuguese	verb-framed	Slobin (2005b)
Dutch	satellite-framed	Slobin (2005b, 2006); Croft et al. (2010)
English	satellite-framed	Talmy (1985)
German	satellite-framed	Berthele (2006)
Swedish	satellite-framed	Viberg (2006)
Polish	satellite-framed / mixed	Slobin (2005a); Kopecka (2009b)
Russian	satellite-framed	Slobin (2005a)
Serbo-Croatian	satellite-framed / mixed	Filipović (2007); Slobin (2005a, 2005b)
Hindi	verb-framed	Narasimhan (2003)
Persian	mixed	Feiz (2011)
Modern Greek	verb-framed / mixed	Papafragou et al. (2006); Talmy (2007: 105); Hickmann et al. (to appear)

On the basis of Talmy's (1991) dichotomy and the classifications made in the literature, we would expect a strong, categorical difference between Russian, English, German, Polish, Swedish and Dutch on one hand and Portuguese, French, Italian and Hindi on the other, with Greek, Serbo-Croatian and Persian somewhere in between. However, this is not what we observe in Figure 3.1. There is a steady decline in the use of the satellite-framed strategy and an increase in the use of the path-only strategy if we move from the top-most language to the bottom-most language. This suggests that languages cannot simply be said to be 'satellite-framed' or 'verb-framed' – they all make use of a subset of the same nine strategies, but do so to different extents.

In spite of this variability, it seems to be possible to identify the two traditional classes of languages, even though it is clear there are some differences between the languages within these classes. On the upper side of the plot we find languages that use the satellite-framed strategy more often than the 'originals' (the strategy usage in the original sentences taken from the English *Alice's Adventures in Wonderland* and the Portuguese *O Alquimista*). In the remainder of this chapter, these languages will be called 'satellite-framed', as is conventional in the Talmian literature, but note that internal diversity does exist within this class. Clear satellite-framed languages are Russian, Dutch, Polish, Lithuanian, Swedish, German, English, and Latvian. On lower side of the plot we find languages that use the satellite-framed strategy less often than the 'originals', and that use the path-only strategy and the verb-framed strategy more often. These languages will be called 'verb-framed', again reflecting traditional terminology, with the provision that the languages in this class are not exactly the same. Clear verb-framed languages are Greek, Italian, French, Portuguese, Romanian, and Albanian.

However, there are also languages that do not really fit one of these two traditional classes. Irish seems to follow a satellite-framed pattern easily[8] and more often than the verb-framed languages, but uses the deictic verb strategy quite often. Serbo-Croatian uses the path-only strategy far more often as the other Balto-Slavic languages, and seems to be shifting from a satellite-framed to a verb-framed system. Hindi and Nepali, unlike other languages traditionally classified as verb-framed, do not use the path-only strategy as much, but especially use the deictic verb strategy, the coordinate strategy, and the subordinate strategy. Persian also deviates from the verb-framed languages by using a fair amount of the deictic verb strategy. Armenian likewise uses the deictic verb strategy and the coordinate strategy. These languages show that a dichotomy cannot be used to classify all possible language types. Since Irish, Hindi, Nepali, Armenian and Persian are different from satellite-framed and verb-framed languages in different ways, it seems more useful to classify languages with regard to their usage of the different motion encoding strategies.

---

[8] With 'easily' is meant that Irish freely uses satellite-framed patterns in boundary-crossing situations, unlike verb-framed languages that often have difficulty with the use of satellite-framed patterns in those contexts:

rith	*sí*	*amach*	*as*	*an*	*teach*
run.PST	3SG.F	away	out	DEF.ART	house.GEN

'She ran out of the house.'

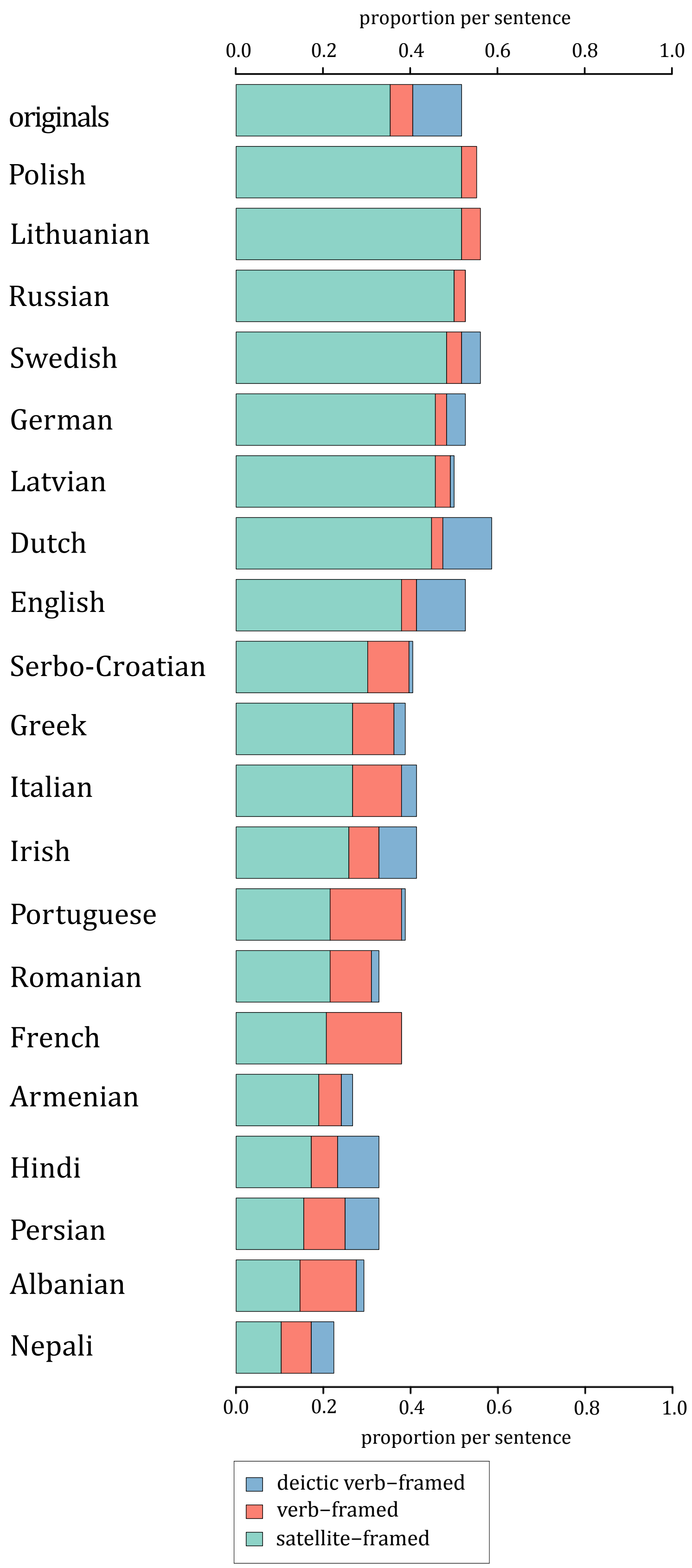


Figure 3.2: *Means of manner expression in twenty Indo-European languages*

In Figure 3.2, the usage of the three most common strategies to encode manner (the satellite-framed strategy, the verb-framed strategy, and the deictic verb-framed strategy) are shown separate from the other strategies.[9] The variation depicted in Figure 3.2 seems to be mostly due to the rates of use of the satellite-framed strategy, which declines as we go from upper part to the lower part, as was shown in Figure 3.1. Verb-framed and deictic verb-framed strategies are used to the same extent both by some of the satellite-framed languages (Dutch, English) and some of the verb-framed languages (the Romance languages, Greek, and Albanian). Languages which make use of the deictic verb strategy relatively often, also make more use of the deictic verb-framed strategy. This is especially true in English and Dutch, where the deictic verb-framed strategy is used much more often than the regular verb-framed strategy.

An interesting finding that emerges from Figure 3.2 is that the Balto-Slavic languages Russian, Polish, Lithuanian and Latvian seem to avoid the usage of the verb-framed strategy. There are some instances of the use of the verb-framed strategy with manner adverbials, but the verb-framed strategy with manner verb participles are quite rare (Russian: none; Polish: 2; Lithuanian: none; Latvian: 2; Serbo-Croatian: none). There seems to be a large pressure for these languages to encode manner on the main verb, as is evident from Figure 3.2 and illustrated by the examples (35) – (39). In these translations, the English original is a deictic verb-framed construction ('came running'), which is translated with a satellite-framed construction in Russian, Polish, Lithuanian and Latvian, and with a adverbial verb-framed construction in Serbo-Croatian.

35) Russian

kak	*vdrug*	***iz***	***les-u***	***vy-bež-a-l***
when	suddenly	from	forest-SG.M.GEN	out-run-VF-PST.3SG.M

livrejn-yj	*lakej*
liveried-SG.M.NOM	footman.SG.M.NOM

'when suddenly a footman in livery ran out from the forest'

[9] Only the three most common strategies to encode manner are included in this graph. Note that the use of the coordinate strategy and the subordinate strategy which feature a manner verb is not included. Languages which make use of these strategies, such as Armenian, Hindi and Nepali, therefore encode slightly more manner as is depicted here.

36) Polish

gdy	*nagl-e*	***z***	***las-u***		
when	sudden-ADV	of	wood-M.GEN.SG		

wy-bieg-ł	*lokaj*
PRFX-run.IPFV-PST.3SG.M	footman.M.NOM.SG

odzi-an-y	*w*	*liberi-ę*
attire.PFV-PST.PTCP.PASS-M.NOM.SG	in	livery-F.ACC.SG

'when suddenly a footman in livery ran out of the forest'

37) Serbo-Croatian

kad	*najednom*	***is-pad-e***	***trk-om***
when	suddenly	PRFX-fall.PFV-AOR.3SG	run-M.INS.SG

iz	***šum-e***	*jedan*	*dvoranin.*
out.of	wood-F.GEN.SG	one.M.NOM.SG	footman.M.NOM.SG

'when suddenly a footman fell running out of the forest'

38) Lithuanian

kai	*staig-a*	***iš***	***mišk-o***	***iš-bėg-o***
when	sudden-ADV	out.of	wood-SG.M.GEN	out-run-3.PST

liokaj-us	*su*	*livrėj-a*
footman-SG.M.NOM	with	livery-SG.F.INST

'when suddenly a footman in livery ran out of the forest'

39) Latvian

pēkšņi	***no***	***mež-a***	***iz-skrēj-a***
suddenly	from	wood-SG.M.GEN	out-run-PST.IND.SG

livrej-ā	*tērp-ies*	*sulainis*
livery-SG.F.LOC	dress-PTCP.SG.M	footman.SG.NOM

'when suddenly a footman in livery ran out from the forest'

Figure 3.2 also shows that the Romance languages, Greek, and Albanian, languages that tend to express path in the verb, do not reach the same amount of manner encoding as is present in the Balto-Slavic and Germanic languages, languages that tend to express manner in the verb. This is probably due to the fact that the (deictic) verb-framed strategy is quite 'heavy' with regard to processing load (Slobin 2004: 229). The native pattern for the expression of manner information in verb-framed languages is that information on manner of motion is often not explicitly coded. However, manner information can sometimes be inferred from context. Adding the same amount of manner information as is present in satellite-framed languages by using the verb-framed strategy would give too much prevalence to the manner information, and would make the text clumsy and difficult to read.

In the end, languages that do not make much use of the satellite-framed strategy simply end up encoding less manner, as is illustrated in Figure 3.2. The use of manner verbs as the main verb of the clause (or as one of the main verbs in one of the clauses, see footnote 9) therefore seems to drive much of the variation within motion typology: it controls both the satellite-framed pattern and the expression of manner in a clause per se. Since the use of the satellite-framed strategy varies from language to language, it is difficult to make a clear dichotomy between 'satellite-framed' and 'verb-framed' languages. For some languages we can say that they are verb-framed or satellite-framed, for other languages different classifications have to be made.

3.3 Aggregation analysis: a demonstration

Figure 3.1 and 3.2 give an indication of how often a strategy is used in each language. However, it does not take into account the relationships between different languages with regard to strategy choice for individual sentences. We can look at these relationships using Neighbor-Net, (Bryant and Moulton 2004), a distance based method for constructing phylogenetic networks. This method calculates the difference between each language in the sample using Hamming distances, aggregating all the differences and correspondences between the languages into a single distance measure. The analysis was conducted with the software SplitsTree4 (Huson and Bryant 2006).

In Figure 3.3, the results of a Neighbor-Net analysis on the usage of the nine motion encoding constructions distinguished in this chapter is presented.[10] A picture emerges that overlaps with the frequency bar plot in Figure 3.1. Three groupings emerge: Russian, Lithuanian, Latvian, and Polish (Balto-Slavic); Irish, English, German, Dutch and Swedish ('Germanic+Irish'); and Greek, Albanian, French, Portuguese, Italian, Romanian, and Serbo-Croatian ('Romance+Balkan'). Persian is situated close to the Romance+Balkan group, while Hindi, Armenian and Nepali appear between the Germanic+Irish and the Romance+Balkan group.

It is clear from Figure 3.3 that a phylogenetic signal can be found in these data: languages that we know to be closely related appear closer together in the graph. This means that languages that are closely related show similar motion event encoding patterns. This is corroborated by phylogenetic tests conducted in

[10] For this analysis and the other analyses reported in this section, constructions coded as 'other' were recoded as 'missing'. This was done to prevent the algorithms used in the analyses from interpreting the category 'other' as a unified, meaningful category.

chapter 4. However, there are also divergences from the phylogenetic pattern: English patterns closely with German and Dutch, as expected, but also seems to be pulled in the direction of Irish; and Serbo-Croatian, a Slavic language, is placed in the Romance+Balkan group.

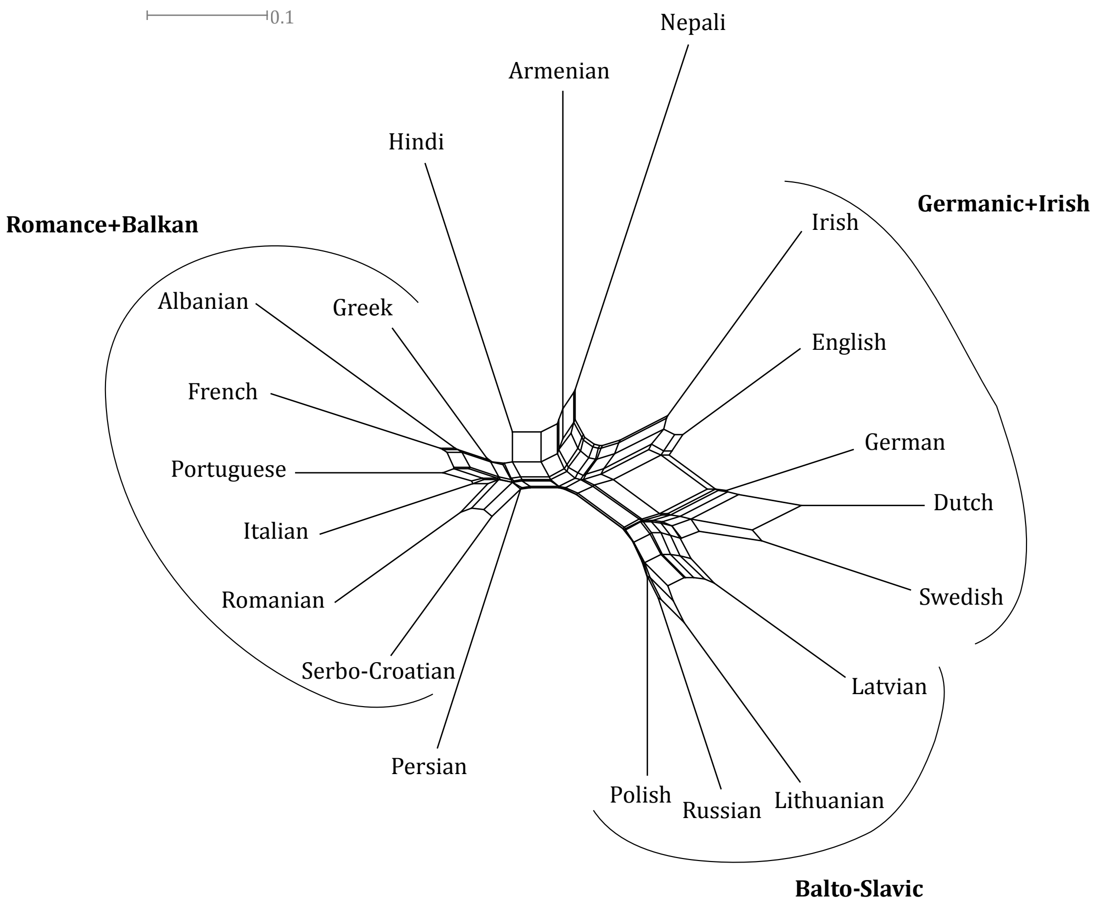


***Figure 3.3**: A split graph showing the results of a Neighbor-Net analysis of motion encoding constructions used in 20 Indo-European languages*

A first interpretation of Figure 3.1 could be that divergences from the phylogenetic pattern are due to language contact: maybe English is situated more closely to Irish because of English-Irish contact? It is possible to assess where such conflicting, non-tree like signal in a Neighbor-Net analysis arises by looking at the delta scores, which can also be calculated by SplitsTree4 (Gray et al. 2010). The delta score for each language gives a measure to what extent each language is involved in conflicting signal. It ranges from 0 to 1, and equals zero if the language is not involved in any conflicting signal.

A prototypical example of a language that generates reticulations of this type is the creole language Sranan, as shown by Gray et al. (2010). Sranan is an English-based creole, but has been spoken in close contact with Dutch for most of its history. As a result of this mixed history, Sranan is positioned between English and Dutch in a Neighbor-Net analysis of vocabulary data of the Germanic languages. Consequently, Sranan has a higher delta score than the other Germanic languages (Gray et al. 2010).

For the current analysis presented in Figure 3.3, the delta scores are given in Table 3.2 below. The average delta score is 0.36. Languages that have a higher delta score are Armenian (0.42), Albanian (0.41), and Serbo-Croatian (0.41). Languages that have a lower delta score are Polish (0.32), Swedish (0.33), German (0.33) and Portuguese (0.33). In this particular case, it seems that these numbers should not immediately be interpreted as indications of conflicting history, as was done by Gray et al. (2010) for Sranan. While Armenian has been influenced by contact with both Indo-European and non-Indo-European languages for centuries, Serbo-Croatian would normally not be characterized as heavily influenced by other languages (although the recent codification into Serbian and Croatian might influence motion event encoding). Also, a contact language like Modern Greek does not have a very high delta score (0.38).[11]

Since language contact does not provide a ready explanation for these patterns, it seems that the higher delta scores for Armenian, Albanian, and Serbo-Croatian suggest a mixed pattern in the type of motion event encoding constructions that are being used. This means that, for a part of the 118-sentence sample, Albanian, Armenian and Serbo-Croatian languages are similar to certain languages, while for another part of the 118-sentence sample, they pattern similarly to other languages. The Neighbor-Net analysis presented in Figure 3.3 can therefore in the first place be interpreted as a map of typological types: a verb-framed group (the Romance languages, Greek and Albanian) a satellite-framed group which doesn't use the deictic verb strategy (Russian, Lithuanian, Latvian, Polish), and a satellite-framed group which does use the deictic verb strategy (Irish, English, German, Dutch and Swedish). The rest of these languages do not immediately belong to one of these groups. Note that if Talmy's (1991) dichotomy was a good classification of motion typology, we would expect two clear groups, and not the crescent shaped continuum that can be observed in Figure 3.3. The Neighbor-Net plot in Figure 3.3 therefore also supports the suggestion that Talmy's (1991) dichotomy is a reduction of the actual variation that is present in motion encoding.

[11] One might suspect that the high delta score for Armenian is caused by the fact that this language constitutes a single Indo-European subgroup and thus is the only language from this subgroup included in the Neighbor-Net analysis. However, this seems not to be the case here. In a Neighbor-Net analysis that included only one, randomly chosen language from each subgroup (included were Dutch, French, Polish, Latvian, Irish, Hindi, Persian, Armenian, Albanian, and Greek), the average delta score was 0.40, with Armenian having a score of 0.46.

Table 3.2: *Delta scores for the Neighbor-Net analysis of motion encoding constructions used in 20 Indo-European languages*

Language	**Delta score**
French	0.35
Italian	0.36
Portuguese	0.33
Romanian	0.36
Irish	0.39
Dutch	0.33
English	0.38
German	0.33
Swedish	0.33
Latvian	0.35
Lithuanian	0.34
Polish	0.32
Russian	0.35
Serbo-Croatian	0.41
Hindi	0.38
Nepali	0.39
Persian	0.37
Modern Greek	0.38
Albanian	0.41
Armenian	0.42

Figure 3.3 shows that Neighbor-Net analysis is not only useful as a method to get a first impression about the phylogenetic signal or geographical signal in the data, it is also useful as a tool to test whether there are any inherent groupings in the data, which may correspond to typological types. It shows (mixed) dependencies between the languages that cannot be assessed from a frequency plot, and cannot easily be inferred from looking at the data matrix with the naked eye. The groups of languages that emerge can then be further investigated, giving rise to specific hypotheses about the specific patternings of motion encoding strategy usage that can be explored further. In this particular case, it seems useful to investigate whether it is possible to find out what is causing Armenian, Albanian and Serbo-Croatian to express this mixed typological pattern.

In order to compare the Neighbor-Net analysis with another aggregation analysis, the results of a classic multidimensional scaling analysis (MDS) are presented in Figure 3.4. This analysis was performed on a Euclidian distance matrix based on the usage of the nine motion encoding strategies. Multidimensional scaling computes a spatial representation of the similarities between the languages. The more similar two languages are, the closer they are placed together on the plot, and the more distinct two languages are, the further away they are placed. Multidimensional scaling can be done using a number of dimensions, ranging from 1 to the number of data points minus 1. The appropriate number of dimensions was assessed by looking at the eigenvalues, which become smaller as newly added dimensions explain less and less variance. For the current dataset, an analysis with 5 dimensions seemed appropriate, but since the first three dimensions already present a clear picture, these first three dimensions have been depicted in Figure 3.4. The first dimension gives part of the Talmian cline, with Armenian, Hindi and Nepali far removed from the other languages. The second dimension gives part of the Talmian cline as well, with Romanian and Serbo-Croatian far removed from the rest of the languages. The third dimension removes Albanian and Armenian from the rest of the languages. The numbers on the axes represent the distances between the languages.

The results in Figure 3.4 are similar to those in Figure 3.3. There seems to be a cline from satellite-framed languages (dashed line) to verb-framed languages (dotted line) in the middle of the plot, with the satellite-framed languages to the left of the middle and the verb-framed languages in the middle. Hindi, Nepali, Albanian and Armenian are removed furthest from the cline of satellite-framed and verb-framed languages. As becomes clear from Figure 3.4, Hindi, Nepali and Armenian are not in fact very similar: they are actually quite different and positioned at quite large distances from each other on all three dimensions. The scale from satellite-framed to verb-framed languages is not as clear-cut as it was in Figure 3.3: Italian is situated quite close to the satellite-framed languages, while Lithuanian is situated quite close to the verb-framed languages.

By using the Neighbor-Net and the MDS analysis, conflicting typological signals were found in the following languages: Albanian, Armenian, and Serbo-Croatian (which had the highest delta scores), English (which is situated in between the other Germanic languages and Irish in Figure 3.3), and Hindi and Nepali (which are located far away from the main group of languages in Figure 3.4). Some reasons for this are presented below.

As has become clear in Figure 3.1, Figure 3.3 and Figure 3.4, Armenian, Hindi and Nepali cannot be said to belong to either the group of satellite-framed or verb-framed languages. None of these languages employs the satellite-framed strategy very often, but all three employ the coordinate and deictic verb strategy relatively often. Even though Armenian, Hindi and Nepali are still quite dissimilar (as illustrated by their locations on the MDS plot in Figure 3.4), there are a few sentences where they match exactly, which seems to give rise to the placement of these three languages close together in Figure 3.3. Both the Neighbor-Net and the MDS analysis give a clear assessment of Armenian, Nepali and Hindi as not belonging to either of the two typological types.

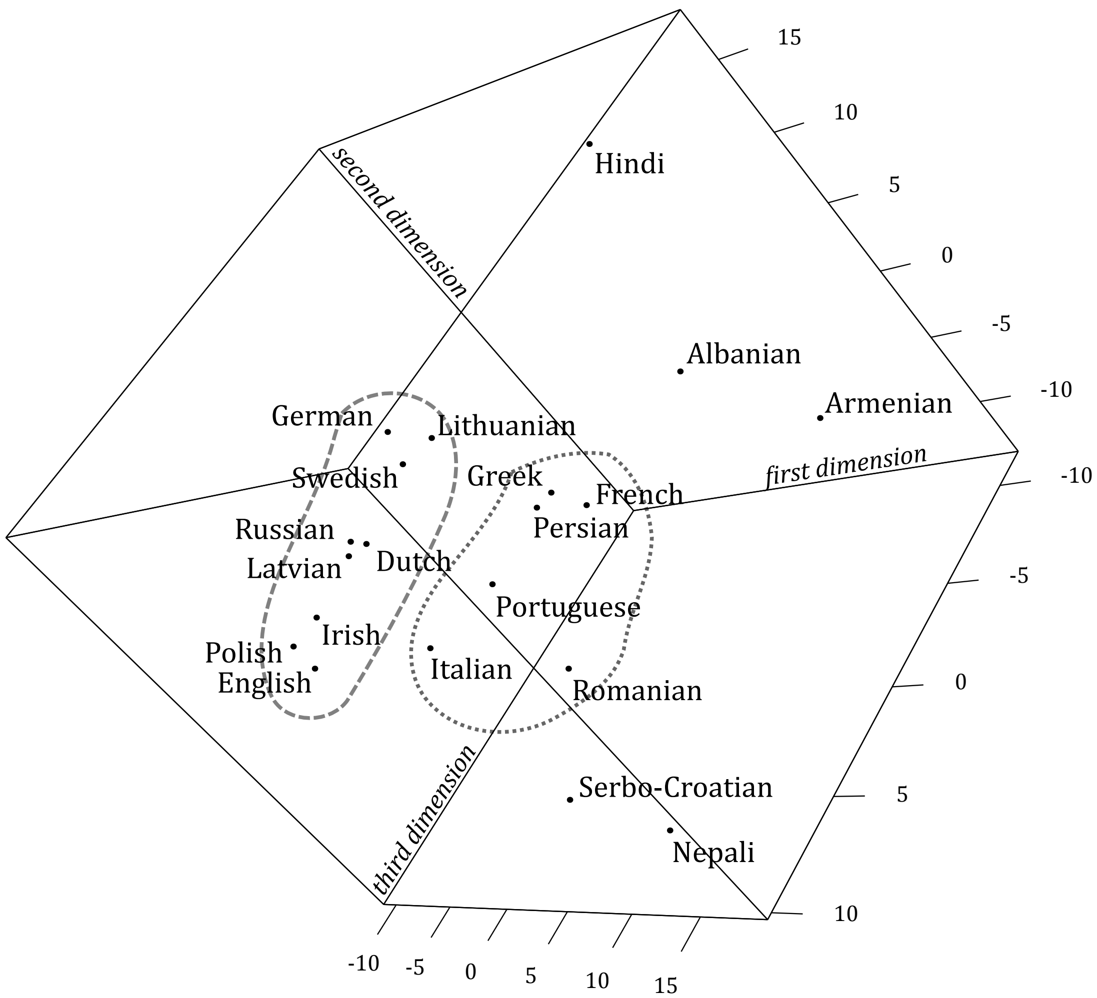


Figure 3.4*: A 3D map of the first three dimensions of a classical multidimensional scaling analysis on motion encoding constructions used in twenty Indo-European languages. The dashed line circles the satellite-framed languages, the dotted line circles the verb-framed languages.*

English is pulled away from the other Germanic languages to a position closer to Irish, as was illustrated in Figure 3.3. English uses the satellite-framed strategy less often than the other three included Germanic languages. This is mostly due to the fact that a part of the sentences was originally in English and because of the sampling techniques that were used, the strategy use of English in the current sample includes a relatively larger proportion of instances of the path-only strategy than would normally be found in an English text. The result is that in Figure 3.3, English has a position closer to Irish, which also employs the satellite-framed strategy less often than might be expected.

Albanian is situated in between French and Greek in Figure 3.3, and outside of the satellite-framed versus verb-framed scale in Figure 3.4. Albanian is clearly verb-framed, as it uses the path-only strategy often, and is therefore placed together with the Romance languages in Figure 3.3. However, Albanian is placed on the edge of this group, close to Greek, as Albanian behaves similar to Greek with respect to the use of the coordinate strategy and the deictic verb-framed strategy. The high delta score of Albanian might be explained by this mixed nature of Greek and Albanian, which are not quite as verb-framed as the Romance languages, and not completely similar to non-satellite-framed languages like Armenian, Hindi, and Nepali either.

Serbo-Croatian has a high delta score, is placed with the Romance languages in Figure 3.3, and is situated outside the main two groups of languages in Figure 3.4. In both Figure 3.3 and 3.4, an affinity with Romanian emerges (this is unfortunately not very clear in Figure 3.4, but Romanian is basically situated in between Serbo-Croatian and the rest of the Romance languages). Out of all Romance languages, Romanian uses the coordinate strategy most often and the verb-framed strategy the least. Serbo-Croatian behaves similar to Romanian with regard to at least these two strategies. Serbo-Croatian is the only Balto-Slavic language in the current sample that is not clearly satellite-framed. The reason for this is that the spatial verbal prefixes present in all Balto-Slavic languages are merging with the deictic verb *ići* 'to go' in Serbo-Croatian, thus creating a range of path verbs. Serbo-Croatian currently seems to be changing from a satellite-framed language to a verb-framed language, as it is using the satellite-framed construction less often and the path-only construction more often. Serbo-Croatian's high delta score and placement in Figure 3.3 and 3.4 is therefore entirely in line with the mixed typological nature of this language.

Although all of the patterns that were discussed here are very tentative and require further investigation in larger corpora, it is clear that aggregation analyses such as Neighbor-Net and MDS analysis are very useful for the discovery of typological patterns and for determining whether there are languages which do not belong in any of the main typological groups. Using different types of analysis and doing the same analysis with subsets of the data is useful to get a better picture of the relations between the languages. Especially because the distance matrices employed by these methods are calculated on a sentence-by-sentence basis, a fine-grained perspective on the variation becomes possible.

3.4 Conclusion

In this chapter, data has been presented on motion encoding from twenty Indo-European languages, seven of which have not been described in the motion event literature that focuses on the classification of languages as 'satellite-framed' or 'verb-framed' before. The data that were gathered for this study come from a parallel corpus. The use of a parallel corpus has proved immensely useful for this typological study, since parallel corpora allow for a full exploration of a typological domain. It provided a full view of the variability of strategy usage in the different languages. Parallel corpora are extremely suitable for all kinds of typological studies (see for instance Wälchli 2009) but are also of great value to dialectologists. Many popular novels have been translated into a range of European dialects. The most interesting novel for this purpose would probably be *Le Petit Prince* by Antoine de Saint-Exupéry, which has been translated into Pennsylvanian German, Platt, Provençal, Gascon, and other dialects. A parallel corpus of translations of any novel into a range of dialects could be used for quantitative study of many different linguistic features.

The theoretical framework of this study relies heavily on Talmy's groundbreaking work on motion in that it employs many of the same concepts (path, figure, ground, manner). There are some important differences, too, for instance the different conceptualizations of what a path satellite is (see section 2.2.2). However, the biggest difference between Talmy's (1991) approach and the current one is that Talmy proposes a dichotomy of language types, while this chapter has shown that languages employ a whole range of different encoding patterns. Classifying languages in terms of the traditional Talmy dichotomy does not take into account this variability. It disregards the variation attested within the verb-framed class and within the satellite-framed class, and cannot account for languages that do not belong to either of these classes.

The aggregation methods employed in section 3.3 support the claim that the variability present in motion encoding cannot be captured in a straightforward dichotomy of verb-framed and satellite-framed languages. These methods clearly show that some languages, such as Irish, Armenian, Hindi, Nepali, and Persian, show a mixture of construction usage that prevents inclusion of these languages in one of these two classes. The characteristics of these languages give rise to the potential identification of new classes or to new hypotheses concerning the mixing or change of typological types. Potential areas of investigation of change in motion encoding could be internal mechanisms (linguistic change) or external mechanisms (contact-induced change). The causal factors behind the motion encoding patterns that were discussed in this chapter with the help of aggregation methods will be the focus of the following chapters. In chapters 4–5, the emphasis will be on the discovery of the mechanisms that have changed the encoding strategies used by Indo-European languages throughout the history of the Indo-European language family.

The Neighbor-Net analysis and the MDS analysis conducted in section 3.3 are not only useful for an assessment of Talmy's (1991) dichotomy. Generally, these methods can be used to identify groupings in any dataset. These groups can be geographical, phylogenetic, and/or typological. For typological studies that use sets of typological features or that use large amounts of empirical data, these methods are very useful for a first assessment of typological groupings. For dialectologists, this type of aggregation method is also very useful to gain an overview of the relationships between different dialects. Explanations for these relationships can then be sought using different methods, for instance using the multivariate spatial analysis proposed by Grieve (2014) to identify regional variation in a set of features, or if the phylogeny of the dialects is known, using the methods proposed by Pagel (1997) to study the evolution of certain features throughout the history of the dialect group. Aggregation methods are therefore valuable tools for scientists involved in cross-linguistic studies, which includes both typologists and dialectologists alike.

Chapter 4: Diachronic change in motion event encoding

This chapter is a slightly revised version of:
Verkerk, A. (2014a). Diachronic change in Indo-European motion event encoding. Journal of Historical Linguistics *4(1): 40-83.*

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

4.1 Introduction

The modern cognitive tradition of the study of motion event encoding originates with Talmy (1985, 1991), who postulated the now well-known difference between so-called 'verb-framed' and 'satellite-framed' languages. Central concepts in Talmy's (1985, 1991) framework are path and manner. Path refers to the path or trajectory that a person or an object has while moving, while manner refers to the way in which a person or an object moves (for instance by flying, swimming or walking). Path is encoded on the verb in verb-framed constructions as exemplified by the Albanian example in (40), where it is encoded on the verb *kaloj* 'to pass'. The manner of motion is not overtly expressed in this sentence. Path is encoded on the satellite in satellite-framed constructions, as exemplified by the Swedish example in (41), where it is encoded on the preposition *genom* 'through'. The manner is expressed by the verb *krypa* 'to crawl' in this sentence. These two examples are taken from the parallel corpus that was introduced in chapter 2.

40) Albanian

nëse	*ende*	*do*	*të*	*vazhdoj*	*të*	*zvogëlohem*
if	still	FUT	FUT	continue.PRS.1SG	to	melt.1SG.PRS

do	*provoj*	*të*	***kaloj***	***nën***	*derë.*
FUT	try.PRS.1SG	to	pass.PRS.1SG	under	door.F.INDF.ACC.SG

‘If I continue to melt, I can try to pass under the door ’

41) Swedish

gör	den	mig	mindre	kan	jag
make.PRS	3SG.N	1SG.OBJ	small.COMPA	can.PRS.AUX	1SG.SBJ

krypa	***genom***	*springa-n*	*under*	*don-en*
crawl.INF	through	slot-SG.DEF.UT	below	door-SG.DEF.UT

‘And if it makes me grow smaller, I can creep under the door’

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

Nevertheless, a few hypotheses on change in motion event encoding have been put forward. Croft et al. (2010: 236), for instance, extensively discuss the two grammaticalization pathways that lead to unified lexical items, most often verbs that express path (path verbs). These are presented in (42). Both pathways are concerned with complex motion events in which both path and manner are expressed.

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

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

b. Coordination > Verb-framed > Verb-Adverb fusion

Croft et al. (2010) describe (38a) as a process in which a complex motion event starts with coordination of two clauses (as in ‘John walked and crossed the street’). This construction gives rise to a serial verb construction (‘John walked crossed the street’). After this step, one of the verbs of the SVC becomes grammaticalized into a satellite (‘John walked across the street’). Then, these satellites fuse again with their verb root to form single roots, most often creating path verbs (‘John crossed the street’). In (38b), the coordinated construction changes to a verb-framed construction with a subordinate manner element (‘John crossed the street walking’). Then, the verb and the subordinate manner verb or manner adverb merge again into a compound ('John crossed-walking the street') and ultimately merge into a single, semantically bleached root ('John crossed the street'). Both of Croft et al.'s (2010) pathways end in fusion of verb roots and particles, resulting in unified lexical items, most often path verbs. These unified lexical elements can then be coordinated again to express more complex motion events, and the beginnings of the grammaticalization pathways in (42) are reached once more.

Another study that discusses diachronic change in motion event encoding is Wälchli (2009). Wälchli (2009: 183-221) studied lexicalization patterns in motion events in a world-wide sample of 117 languages. He focused on the use of path verbs in five different path domains (enter, exit, ascend, descend, and pass/cross) in a set of 56 motion clauses in a parallel corpus of the *Gospel according to Mark* (a Bible text). An assessment of whether a path domain is predominantly encoded by path verbs or by other types of verbs (including deictic verbs and manner verbs) was made for each domain. Wälchli (2009: 214-215) found that certain languages encode all path domains predominantly with path verbs, such as French, Italian, and Spanish, while others encode all path domains predominantly with other types of verbs, such as Latin, Irish, and German. Other languages encode some but not all path domains predominantly with path verbs, such as English, Greek, and Hindi.

Wälchli (2009: 214) finds that there is little genealogical stability for the choice of verb type in his world-wide sample. The same holds for Indo-European. None of the subgroups of the Indo-European language family behave in a completely unified manner, and the languages of the Indo-Iranian subgroup even range across the entire scale of path lexicalization - some languages use path verbs for all path domains, while others do not use path verbs for any of the path domains. To explain these results, Wälchli (2009: 214) proposes the existence of an areal pattern rather than a genealogical one: "Languages in Northern and Central Europe (including Finnic and Hungarian) as well as in the Caucasus (except Armenian) tend to lack route verbs [path verbs, AV]." Even though languages that do not use path verbs are found all across the world, they seem to be most common in the area mentioned above. Wälchli (2009: 216) also finds that no language family bigger than a subfamily such as Germanic lacks path verbs altogether. He suggests that languages tend to acquire new path verbs with relative ease due to a variety of mechanisms, including the univerbation of adverbial path markers with verb stems (Croft et al. 2010's first grammaticalization pathway, see (37a)) and the borrowing of path verbs.

Croft et al. (2010) indicate that there are clear grammaticalization pathways for diachronic change in motion event encoding constructions. Wälchli (2009) suggests that languages may move through these grammaticalization pathways at a steady pace, as even closely related languages may behave quite differently. However, none of these claims have been tested on a full-scale study of motion event encoding in a single language family. This chapter aims to fill that gap.

The current chapter is an investigation of diachronic change in motion event encoding in the Indo-European language family. Specifically, it investigates whether phylogenetic comparative analysis of contemporary language data can provide support for one of the following two hypotheses: 1) the hypothesis that Proto-Indo-European was satellite-framed, as proposed by Talmy (2007) and Acedo Matellán and Mateu (2008), and 2) the hypothesis that Proto-Indo-European was typologically mixed, as suggested by a review of the comparative work on the ancient Indo-European languages presented in section 4.2. The question of whether Proto-Indo-European was satellite-framed or mixed was investigated using the available comparative historical data from ancient Indo-European languages such as Latin, Homeric Greek and Vedic Sanskrit as well as phylogenetic comparative methods. In particular, a method called ancestral-state estimation that was introduced in section 1.3.3.3 is used to infer the behavior of the now extinct ancestors of contemporary languages on the basis of the behavior of modern languages by finding the evolutionary model that has the best fit to the modern data. For the ancestral-state reconstruction, comparable usage data on motion event encoding from a sample of 20 contemporary Indo-European languages was used (French, Italian, Portuguese, Romanian [Romance], Irish [Celtic], Dutch, English, German, Swedish [Germanic], Latvian, Lithuanian, Polish, Russian, Serbo-Croatian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Iranian], Modern Greek [Hellenic], Albanian, and Armenian). This dataset is taken from a parallel corpus of two novels, *Alice's Adventures in Wonderland* (by Lewis Carroll) and *O Alquimista* [The Alchemist] (by Paulo Coelho), that was introduced in chapter 2. In this chapter, the 118-sentence sample described in section 2.1.2 was employed.

In the current chapter, ancestral-state estimation analyses are used to infer the motion event encoding system of Proto-Indo-European, the ancestor of all Indo-European languages. The use of ancestral-state estimation analyses on data from contemporary languages allows one to draw a comprehensive picture of typological change from (for instance) the reconstructed Proto-Indo-European language to the reconstructed language Proto-Germanic, and from Proto-Germanic to the contemporary Germanic languages. The information on motion event encoding in the ancient Indo-European languages is incorporated in the current analysis as much as possible. However, note that the information on motion event encoding in ancient languages is mostly concerned with qualitative descriptions of different motion constructions. As was presented in chapter 3, a quantitative analysis of construction usage is needed if the motion event encoding system is to be captured accurately. The data presented in this chapter is therefore concerned with quantitative analysis of the use of different motion constructions in a corpus, and goes beyond stating which constructions are attested.

The comparative historical data is presented in section 4.2. The phylogenetic comparative methods that are employed are described in section 4.3. The results of these analyses are presented in section 4.4. A general discussion of the results is given in section 4.5, while future directions and a general conclusion are provided in section 4.6.

4.2 The comparative historical evidence

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

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

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

The most ancient Indo-European languages (Greek, Hittite, Vedic Sanskrit, Avestan and Old Persian) were characterized by a relatively free word order, in which these adverbial particles moved freely within the sentence as modifiers of verbs and as modifiers of nouns (Delbrück 1888, 1893; Speyer 1896; Whitney 1879; Kuryłowic 1964; Watkins 1964; Hofmann and Szantyr 1965; Lehmann 1974: 116ff, 212ff, 228ff, 233ff; Hewson and Bubenik 2006: 358ff; Luraghi 2010). These adverbial particles are called 'preverbs' when they function as modifiers of verbs. Examples of their different functions are provided in (43), in which *epí* is a preverb that is separated from its verb (this process is called 'tmesis') and in (44), in which *epí* is an adposition, which could have been placed in several different places in the sentence.

43) Homeric Greek

kai	*epì*	*knéphas*	*hierón*	*élthēi*
and	on	darkness	sacred	come.3SG

'and the sacred darkness closes in' [*Il*. 1 1.209]

44) Homeric Greek

pléōn	*epì*	*oínopa*	*pónton*
sailing	over	wine.dark	sea.ACC

‘sailing over the wine-dark sea’ [*Il*. 7.88]

Preverbs were used to encode path in satellite-framed constructions such as (43), in which *epí* encodes that the darkness is coming towards a reference point, presumably the speaker. This was true for most ancient Indo-European languages, such as Homeric Greek, Hittite, Vedic Sanskrit, Avestan and Old Persian, and therefore the preverb system is reconstructed to be Proto-Indo-European. Some of these preverbs encoded deictic reference rather than path, and were used by the ancient Indo-European languages to encode deixis. Most ancient Indo-European languages (except Hittite) lacked deictic verbs, a situation that can still be observed in the modern Balto-Slavic languages.

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

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

However, not all Romance languages seem to be completely verb-framed. Italian is reported to have satellite-framed constructions with a limited set of verbs (Folli and Ramchand, 2001, 2005). Italian can also make use of a satellite-framed construction using a set of post-verbal particles that are used both with path verbs and with manner verbs (Masini 2005; Iacobini and Masini 2006, 2007). Brucale et al. (2011) and Brucale (2011), who study Classical Latin, show that Latin might not be completely satellite-framed either. Their studies suggests that even though it was possible to use the satellite-framed construction in Classical Latin, it might not have been used very often at all. Ferrari and Mosca (2010: 320) seem to support this view when they describe Latin as an ‘unstable’ language in which path is distributed over three linguistic elements (preverbs, prepositions, and cases). Even though it is clear that the satellite-framed construction was productive in Latin, corpus studies are needed to assess how often it was used in order to assess the exact magnitude of the change from Latin to the Romance languages. The same applies to most other ancient Indo-European languages discussed in this chapter, for which there exist virtually no quantitative corpus studies of motion event encoding.

A similar change took place in the Indo-Aryan languages: the preverbs that were present in Vedic Sanskrit (see Danesi 2013 for an overview) became more fixed and developed into a system of postpositions (Speyer 1896; Bloch 1965). They also became more closely associated with the verb, and in the end merged with verb roots altogether: “Preverbs are shown by etymology to exist at the beginning of many modern verbs commencing with o- or u- (apa-, ava-, ud-) or by p- (pra-, prati-), v-/b- (vi-), sam-.” (Bloch 1965: 158). Speyer (1896: 47) writes that this univerbation process took place during the transition from Vedic Sanskrit to Classical Sanskrit. Hindi, the only Indo-Aryan language to be studied from a Talmian perspective, is verb-framed according to Narasimhan (1998), but has a tendency to use different types of non-satellite-framed motion encoding strategies as demonstrated in chapter 3.

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

In Homeric Greek, the occasional instance of free word order for the path encoding adverbial particles is already considered to be a poetic archaism (Schwyzer 1950: 425). The fixation of adverbial particles in front of nouns leads to the development of the prepositional phrase in Classical Greek (Hewson and Bubenik 2006: 59ff). The preverbs also underwent processes of univerbation with the verb root (Skopeteas 2002: 164ff, 349ff). For Homeric Greek, Imbert (2010) lists fourteen adverbial path particles that were productive as adverbs, preverbs, and adpositions (*amphí*, *aná*, *apó*, *diá*, *eis*, *ek*, *en*, *epí*, *hupér*, *hupó*, *katá*, *pará*, *perí*, and *pró*). For Modern Greek, Holton et al. (1997: 180) list nine of these that may function as prefixes on verbs and as prepositions (*ana*, *apó*, *dia*, *ek*, *huper*, *hupo*, *katá*, *pará*, and *pró*). It is clear that many Modern Greek verbs are the result of univerbation between preverbs and verb roots that took place in the development from Homeric Greek to Modern Greek, (i.e. see the range of verbs derived from *bállo* 'to throw, to put' listed by Holton et al. 1997: 180). Many of the preverb-verb combinations no longer have compositional meanings, suggesting that these have become unified lexical elements (Skopeteas 2002). Modern Greek is said to be verb-framed (Papafragou et al. 2002, 2006) or mixed (Talmy 2007: 105; Skopeteas 2002: 34; Hickmann et al. to appear).

Although not much is known on motion event encoding in Hittite, work by Brosch (2014) and Junghänel (in preparation) suggests that Hittite mostly preserved the Proto-Indo-European system of free path particles and thus was satellite-framed. However, Brosch (2014: 442) notes that even though Hittite is mostly satellite-framed, there are at least two path verbs, which may suggest an ongoing change from a satellite-framed to a verb-framed system. In addition, he also claims that Hittite has a very restricted manner verb lexicon, while it has a rich lexicon of verbs that encode deictic information (Brosch 2014: 326-327). Junghänel (in preparation) tentatively notes that Lydian (an Anatolian language like Hittite attested in 100 BCE) employs a system of path prefixes, suggesting a potential change towards a verb-framed system for the Anatolian languages.

There is some information on Armenian from Wälchli's (2009) study of lexicalization patterns that suggests that Classical Armenian was already verb-framed to the same extent as the modern Romance languages (Wälchli 2009: 215). This suggests that a change from a hypothetical satellite-framed system in Proto-Indo-European to a verb-framed system must have been completed before the emergence of Classical Armenian between 400-1100 CE. This is also suggested in Schmitt (1981: 86), who points out that preverb-verb combinations only rarely feature in Classical Armenian. Rather than fusing its preverbs with verbs in a later stage, the productive use of preverbs was already obsolete in Classical Armenian.

The shift from a preverb system to a prepositional system is attested in Albanian as well, suggested by several adverbs and prepositions that are cognate with Proto-Indo-European preverbs (*jashtë* 'out', *ndë* 'into', *nga* 'from', *para* 'before', *për* 'for', etc.). It also has a range of verbal path prefixes that are cognate with the preverbs found in Homeric Greek, Latin, and Sanskrit (Orel 2000: 167). Orel (2000) provides evidence that, also in parallel to Modern Greek and the

Romance languages, a process of univerbation of verbal prefixes with verb roots has taken place. Motion verbs in which this process can be detected are *ndjek* 'to follow', *ngre* 'to lift', and *përshkoj* 'to go through'. Prefixation is still a productive means for verb derivation, although the majority of the prefixes used today are innovations. None of these prefixes have spatial meanings (Camaj 1984: 208ff).

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

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

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

Taking this data on diachronic change in Indo-European languages, it seems clear that the last stage of Croft et al. (2010)'s first grammaticalization pathway (see (37a)) is attested in all Indo-European subgroups, albeit to different extents. The preverbs that were used in satellite-framed constructions merged with verb roots to create path verbs. In the Romance languages and the Indo-Iranian languages this change has lead to the emergence of path verbs and the increased use of verb-framed constructions, while in Germanic and Balto-Slavic it has resulted in some univerbated or grammaticalized verbs (such as Dutch *binnen-vallen* lit. inside-fall, i.e. 'to visit unexpectedly'), but it has not led to a shift away from using satellite-framed constructions.

The data presented in this section lend support to the claim that Proto-Indo-European could make use of the satellite-framed construction, as there is evidence for the existence of the satellite-framed construction in Latin, Hittite, Ancient Greek, Old Irish, and Vedic Sanskrit. This is also why Talmy (2007) and

Acedo Matellán and Mateu (2008) have proposed that Proto-Indo-European was satellite-framed. However, it remains unclear whether the satellite-framed construction was the most frequently used motion encoding construction in these languages. As we have seen, Brucale et al. (2011) and Brucale (2011) point out that in Latin, the satellite-framed construction was not used as often as suggested previously, and Brosch (2014) points out that in Hittite, the manner verb lexicon was quite small (path satellites were most often used in combination with deictic verbs). Therefore, the overview of comparative historical data on motion in the ancient Indo-European languages presented in the current section gives rise to the alternative hypothesis that Proto-Indo-European was typologically mixed, rather than strictly satellite-framed (see also Ringe 2006: 58-59 on the productivity of preverb-verb combinations in Proto-Indo-European).

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

4.3 Methodology

As measures of the motion event encoding system, the score on the PC1, the score on the PC2, and the proportions of usage of the five most frequent motion event encoding constructions were used. The evolutionary behavior of these measures along the branches of the Indo-European tree will be investigated in this chapter. The dataset used for the analyses conducted in this chapter was the 118-sentence sample. Figure 2.1, which presented an overview of the usage of each motion event encoding construction in the 20 languages included in the 118-sentence sample, is repeated below as Figure 4.1 for convenience. Figure 2.2, which gives the position of each of the 20 languages on the first and second principal components analysis (PC1 and PC2) conducted in section 2.3, is repeated below as Figure 4.2 for convenience.

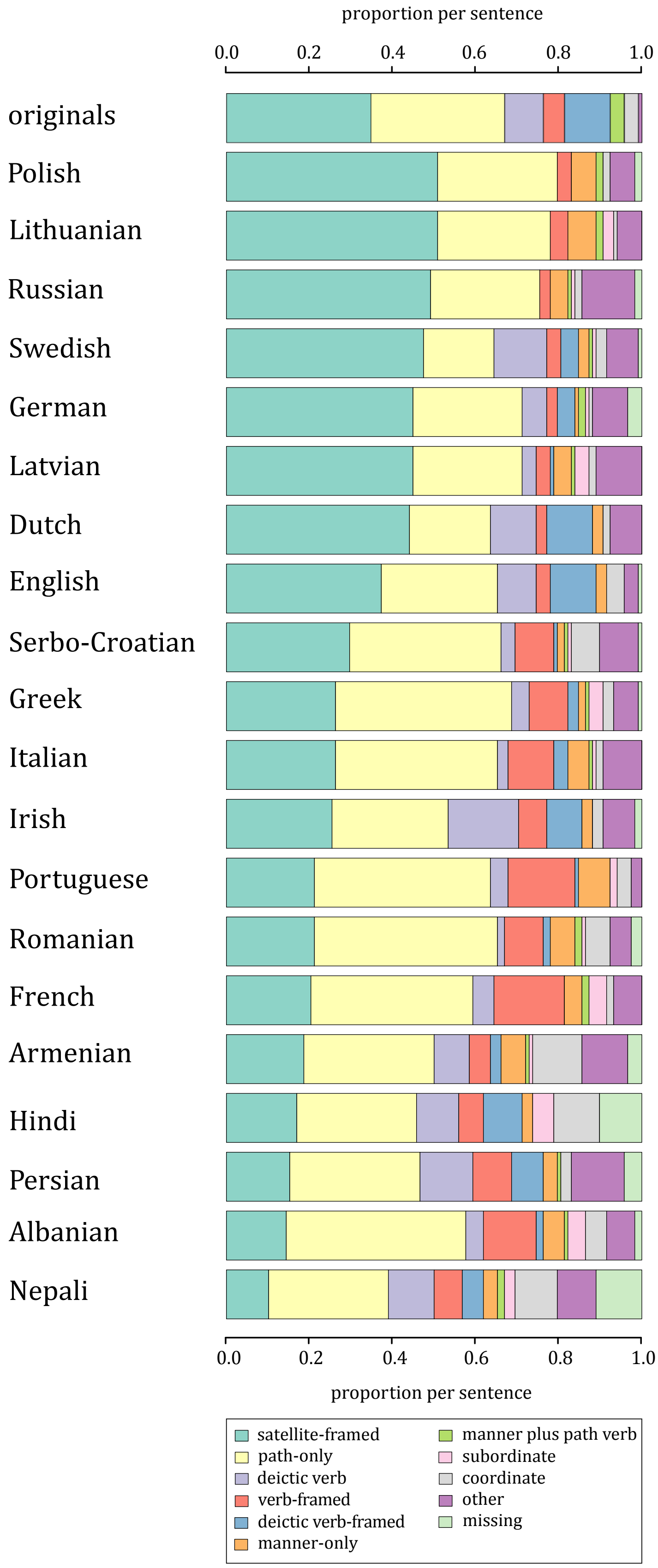


***Figure 4.1**: The frequency of the usage of nine different motion encoding strategies in twenty Indo-European languages*

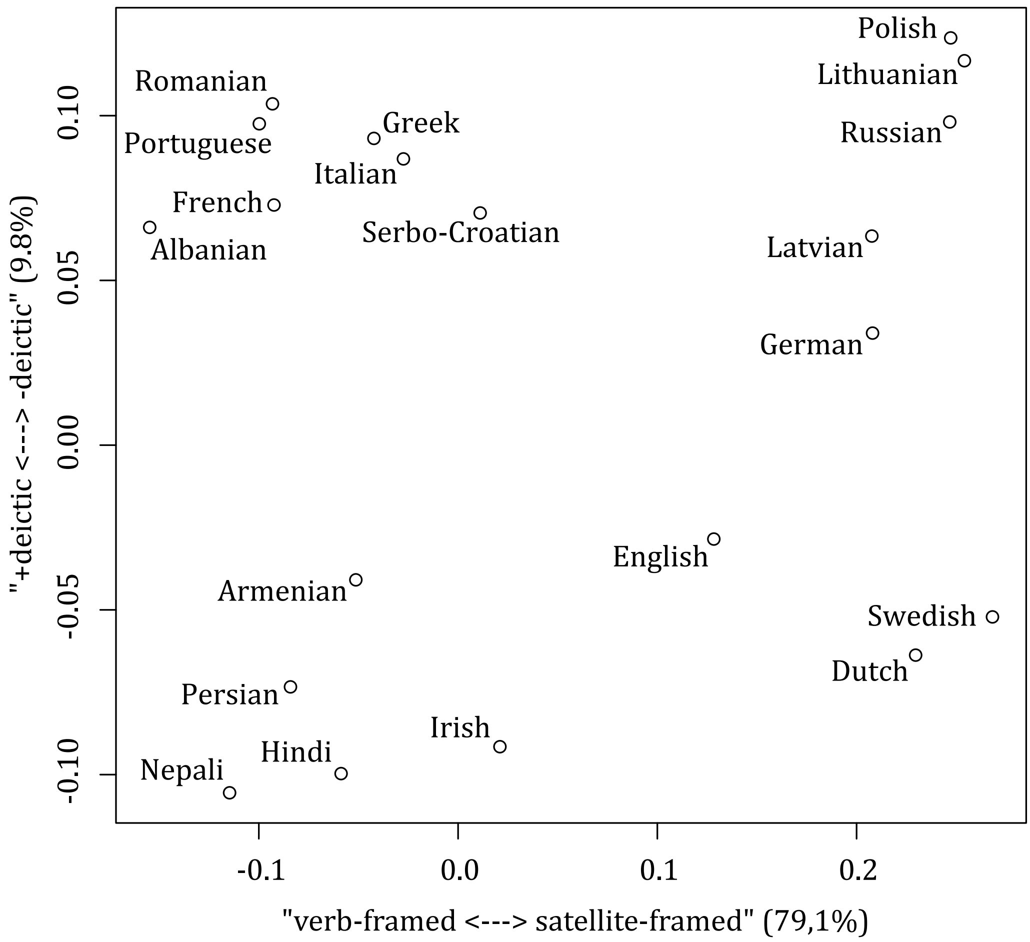


Figure 4.2*: A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 118-sentence sample for 20 Indo-European languages*

A phylogenetic measure of relatedness in the form of a set of phylogenetic trees was used in order to reconstruct Proto-Indo-European motion event encoding and identify typological changes in motion event encoding. This tree set was taken from Bouckaert et al. (2012) as introduced in section 1.3.2.2. The maximum clade credibility tree that summarizes this tree sample was presented in Figure 1.6 and is repeated below as Figure 4.3 for convenience.

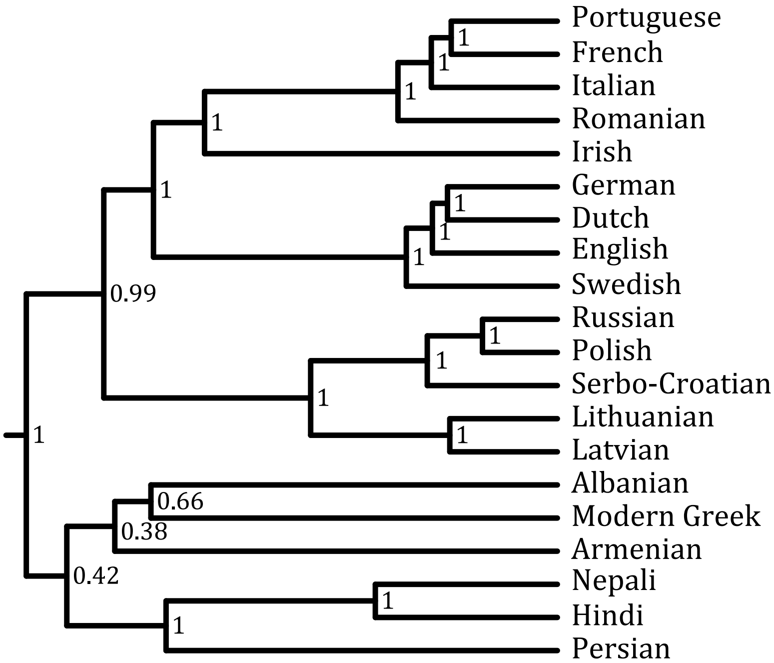


Figure 4.3*: The maximum clade credibility tree of 1000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 20 languages featured in this dissertation.*

First, the presence of phylogenetic signal was investigated as explained in section 1.3.3.1. When phylogenetic signal is present, this implies that languages behave similarly due to shared genealogy. The presence of phylogenetic signal was tested using the parameter λ (Pagel 1999a; Freckleton et al. 2002). λ is a parameter that estimates to what extent the data is evolving exactly as the phylogeny would predict. A perfect match between data and phylogeny would mean that language A would be exactly as similar to language B as would be predicted by the distance between language A and B on the phylogeny. A maximum value of λ (which is typically 1 but may be higher due to features of the phylogenetic tree) indicates that a trait is evolving exactly as predicted by the phylogeny under a random walk model of evolution, while a value of $\lambda = 0$ would indicate that the evolution of the trait is taking place entirely independent from the phylogeny, i.e. completely random with respect to history (Freckleton et al. 2002). For the current dataset and the phylogenetic trees described above, λ was estimated using the function phylosig, part of the R (R Development Core Team, 2011) package phytools (Revell 2012). The estimations of the likelihood of a model in which λ was set to 1 and 0 were conducted using the function fitContinuous, part of the R package GEIGER (Harmon et al. 2008). Results are presented in section 4.4.

Ancestral state estimation (ASE) was also performed using a maximum likelihood approach, using the random walk model of evolution as introduced in section 1.3.3.2 (Schluter et al. 1997; Webster 2002). The estimation of ancestral states is a two-step problem. First, ancestral states for each internal node in the tree are computed as weighted averages of the values of the languages that are connected by that node. The weights are inversely proportional to the branch length that connects the internal node and the descendent language or node, so that descendant languages or nodes that are connected to the internal node with short branches have a larger weight as descendant languages or nodes that are connected to the internal node with longer branches. Second, an algorithm moves from the root towards the languages at the tips of the tree, adjusting the ancestral states of the internal nodes in such a way that the complete phylogeny is considered in the estimation of the most likely ancestral states. This algorithm minimizes the sum of the square of the weighted differences between ancestral nodes and descendent languages and nodes over the whole tree. ASE was conducted using the function getAncStates, part of the R package geiger (Harmon et al. 2008). Supporting functions in the R package ape (Paradis et al. 2004) were used as well. Results are reported in section 4.4.

To measure the degree of dependency between languages due to closeness in geographic space a partial Mantel test, which is a measure of autocorrelation, was conducted. The partial Mantel test was used to calculate the spatial correlation between the motion encoding system and geographical distance while taking into account the phylogenetic distance. The score on the PC1 was used as a characterization of the motion encoding system. The latitudes and longitudes of the capitals of the nations in which the languages in the sample are spoken were used as a measure of geographical distance. The length of the branches between each set of languages on the maximum clade credibility tree presented in Figure 4.3 was used as a measure of phylogenetic distance.

First, the partial Mantel analysis tests for the correlation between the three different distance matrices. Second, it tests whether this correlation is statistically significant by permuting the matrices and estimating the correlation repeatedly to compare the original test statistic to the distribution of test statistics from the permutations. In this way, a simulated p-value is generated. The Mantel test was conducted using the function mantel.partial, part of the R package vegan (Oksanen et al. 2012). Results are presented in section 4.4.

4.4 Results

The test to determine whether phylogenetic signal was present was conducted on the 1000-tree sample. The score on the PC1 was used as a holistic measure of motion event encoding for the 20 languages of the sample. Because a tree sample is used rather than a single tree, we do not have a single estimate of λ, but a range of 1000 estimates of λ. λ was estimated to have a range of 0.95-1.28, with a median of 1.16. An overview of the lambda estimates is included in Figure 4.4.

The maximum possible λ values, given the phylogenetic trees in the current 118-sentence-sample, ranged from 1.09 to 1.28 (see section 1.3.3.2), so the optimized lambda values are quite high. The estimated lambdas were significantly different from 0 for all 1000 trees (median $p = 4.4 \times 10^{-6}$, range p $4.4 \times 10^{-7} - 0.0005$). λ was estimated not to be significantly different from 1 in a large portion of the trees (the p-value ranged between 0.001-0.05 for 914 trees, and ranged between 0.06-1 for the other 86 trees).[12] This indicates that there is a clear phylogenetic signal in these data.

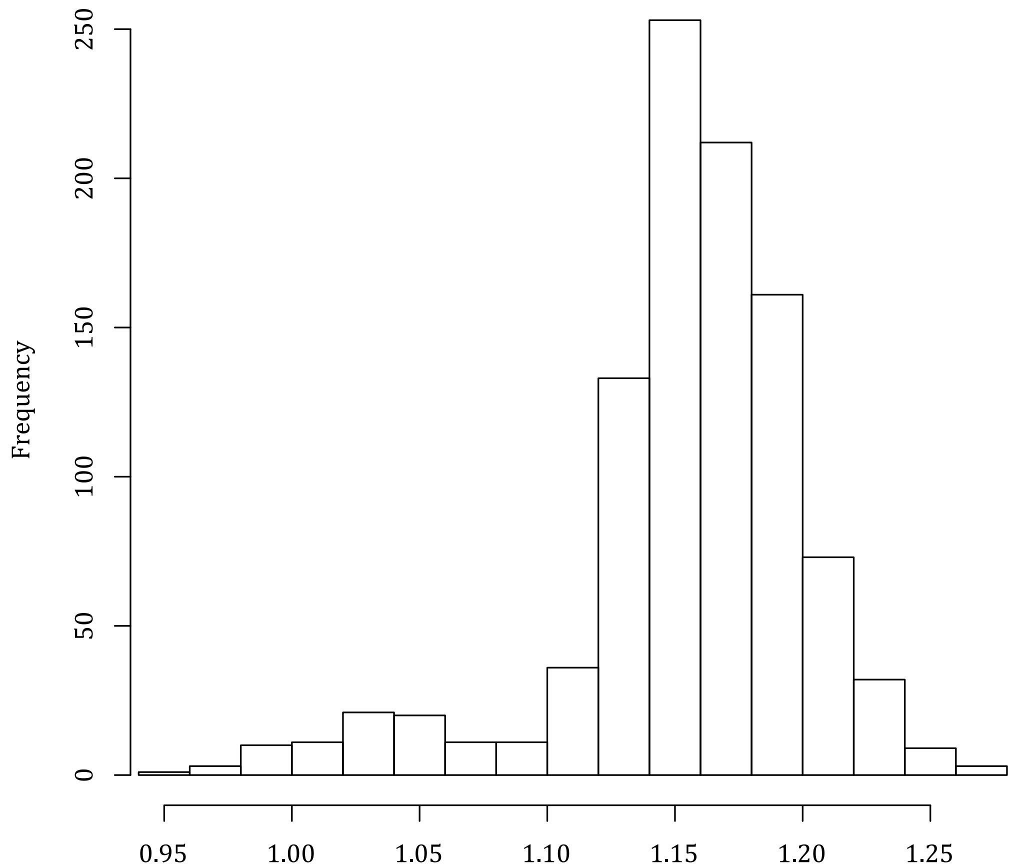


***Figure 4.4**: Estimated λ values (x-axis) for the language scores as captured by the PC1 for the 1000 tree sample*

The results of the ancestral-state estimation analyses for Proto-Indo-European are presented in Table 4.1. Ancestral-state estimates were made for a range of different measures, including the score on the PC1, the score on the PC2, and the proportion of use for the five most frequently used motion encoding constructions (the satellite-framed construction, the path-only construction, the

[12] 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 λ =1. However, the true maximum values could not be used as maximum possible λ values, as they are not accepted by the corPagel function from the R package ape (Paradis et al. 2004). This function only accepts fixed values for λ between 0 and 1.

deictic verb construction, the verb-framed construction, and the deictic verb-framed construction). Since the analyses were conducted over a range of 1000 trees, the estimates again are ranges rather than single numbers. Bear in mind in what follows that the scores on the PC1 range from -0.11 ('verb-framed') to 0.27 ('satellite-framed'), with a mean of 0.05, and the scores on the PC2 range from -0.11 ('+deictic') to 0.12 (-deictic), with a mean of 0.02 (see Table 2.7). In Table 4.1, the median of the ancestral-state estimation analyses over all 1000 trees in the sample is provided first, followed by the range of ancestral-state estimates.

Table 4.1*: Estimated ancestral states for Proto-Indo-European principal component scores and proportions of construction usage*

Measure	**Median**	**Range**
score on PC1 (verb-framed vs. satellite-framed)	-0.001	-0.03–0.02
score on the PC2 (-deictic vs. +deictic)	0.004	-0.01–0.01
proportion of use of the satellite-framed construction	0.26	0.23–0.28
proportion of use of the path-only construction	0.33	0.32–0.34
proportion of use of the deictic verb construction	0.07	0.07–0.08
proportion of use of the verb-framed construction	0.07	0.07–0.08
proportion of use of the deictic verb-framed construction	0.04	0.04–0.05

The results in Table 4.1 indicate that Proto-Indo-European is estimated to be a type of language that is intermediate on the Talmian scale: the ancestral states inferred for the proportion of use of the different motion event encoding constructions are in all cases quite close to the mean values presented in Table 2.7. There seems to be a slight tendency towards the verb-framed side of the scale, as the use of the satellite-framed construction (0.26) is estimated to be slightly lower than the mean value (0.31), and the use of the path-only construction is estimated to be slightly higher (0.33 compared with the mean value 0.32). The score on the first principal component (-0.001 compared with the mean value 0.05) is also directed in favor of the verb-framed side of the scale. The language that seems to come as close as possible to the estimated values is Modern Greek. Modern Greek is verb-framed, but is also very close to the middle of the scale of satellite-framed and verb-framed languages that was presented in section 3.2.

Note that the estimates for the deictic verb construction and the deictic verb-framed construction cannot be interpreted in a meaningful way, as most ancient Indo-European languages and Proto-Indo-European do not have (reconstructed) deictic verbs. The reconstructed Proto-Indo-European verb **$h_1$ei*- 'to go' (Rix 2001: 232-233), for instance, is not reconstructed to be a deictic verb, but is rather a general motion verb that means 'to walk, move, go'. The deictic opposition between verbs meaning 'come' and 'go' in the modern Indo-European languages is likely to have arisen from a difference in aspect between the root *h_1*ei*- and *g^u*em*- 'to go, to come' (Rix 2001: 209-210). Whereas *h_1*ei*- was an present stem, *g^u*em*- was an aorist stem, and this aspectual distinction may later have changed into a deictic distinction:

> It is probable that this situation fairly reflects the difference in value between the two important IE roots *ei- and *g^wem-, as essentially one of aspect. The former expressed the action of going *per se*, while the latter involved the implication of a definite goal (Buck 1949: 693).

The lack of deictic verbs in the ancient Indo-European languages indicates that the results of an ancestral-state estimation analysis can never be interpreted without the linguistic knowledge that we have about these languages, such as the presence of (reconstructed) deictic verbs that are crucial for these two constructions.

Separate ancestral-state estimation analyses for the major Indo-European subgroups Romance, Germanic, Balto-Slavic, and Indo-Iranian were conducted as well. The results are presented in Table 4.2. In Table 4.2, the median of the ancestral-state estimation analyses over all 1000 trees in the sample is provided first, followed by the range of ancestral state estimates underneath the median.

Table 4.2*: Estimated ancestral states for principal component scores and proportion of construction usage of the four most frequently used constructions for the roots of four subgroups: Romance, Germanic, Balto-Slavic, and Indo-Iranian.*

	PC Scores		**Proportion of use measures**			
P[a]	**Score on PC1 (v-framed vs. s-framed)**	**Score on PC2 (-deictic vs. +deictic)**	**satellite-framed construction**	**path-only construction**	**deictic verb construction**	**verb-framed construction**
R	-0.06	0.07	0.24	0.40	0.04	0.12
	-0.04–-0.07	0.07–0.08	0.23–0.25	0.39–0.40	0.04–0.05	0.11–0.12
G	0.19	-0.03	0.41	0.24	0.10	0.04
	0.17–0.19	-0.03–0.02	0.41–0.42	0.23–0.24	0.09–0.10	0.03–0.04
BS	0.14	0.07	0.40	0.30	0.03	0.06
	0.11–0.16	0.06–0.08	0.37–0.42	0.29–0.31	0.02–0.04	0.05–0.06
II	-0.04	-0.03	0.21	0.32	0.09	0.08
	-0.06–-0.02	-0.06–0.02	0.18–0.23	0.31–0.33	0.08–0.10	0.07–0.08

[a]P. = Proto-language of each subgroup; R = Romance; G = Germanic; BS = Balto-Slavic; II = Indo-Iranian.

The scores on the PC1 and the median of the ancestral values estimated for the score on the PC1 for Germanic, Romance, Balto-Slavic, Indo-Iranian, and Proto-Indo-European are placed upon the maximum clade credibility tree in Figure 4.5.

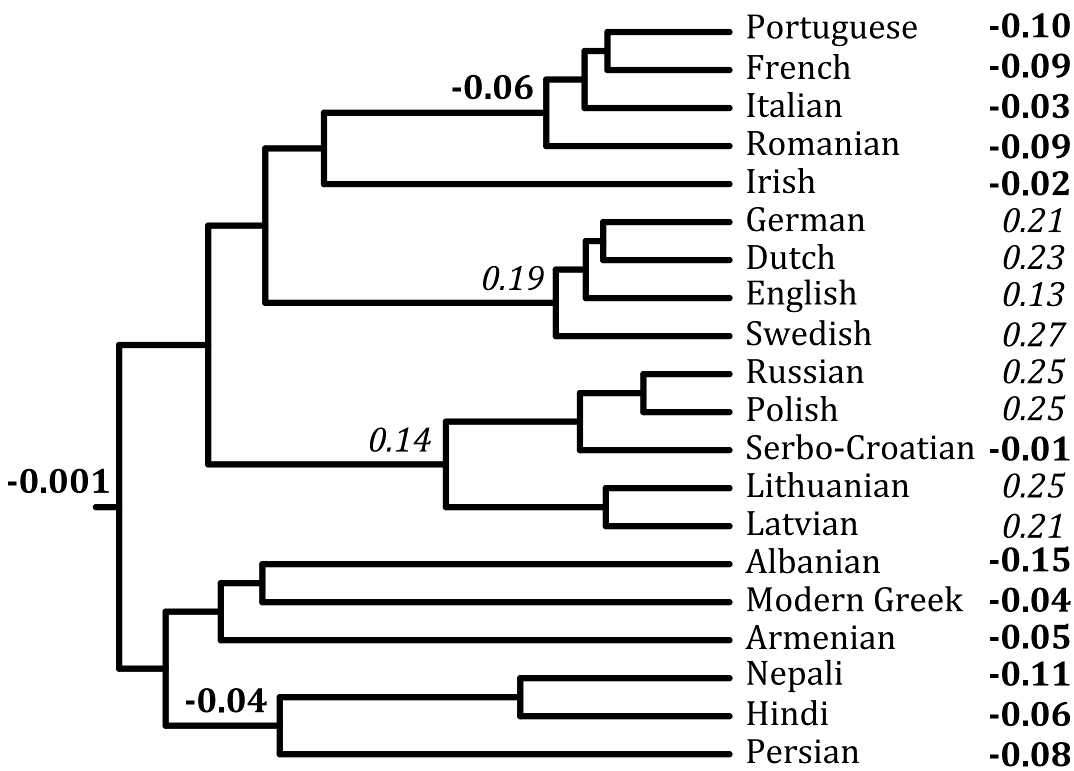


***Figure 4.5**: Scores on the PC1 placed on the maximum clade credibility tree sampled from the posterior sample of trees in Bouckaert et al. (2012). Ancestral-state estimates of the score on the PC1 are added for Proto-Indo-European and the Germanic, Romance, Balto-Slavic, and Indo-Iranian subgroups. Negative PC1 scores ('satellite-framed') are in bold, positive PC1 scores ('verb-framed') are in italics.*

To test how robust these results are, additional ancestral-state estimation analyses were conducted. The first robustness test was performed to see whether the estimated ancestral state for Proto-Indo-European would shift when information on ancient languages was incorporated. To this end, the 1000 phylogenetic tree sample from Bouckaert et al. (2012) was pruned again, this time to include four ancient languages: Classic Armenian, Ancient Greek, Vedic Sanskrit, and Latin (see the maximum clade credibility tree in Appendix 4). Since there are no translations of the two novels that form the parallel corpus on which this study is based in these four ancient languages, it is only possible here to provide an approximation of the behavior of these ancient languages on the measures used in this chapter, which is based on the literature described in section 4.2. This was done by assigning each of these ancient languages the scores of one of the modern languages that it is claimed to resemble. This procedure is of course somewhat arbitrary, as it can only provide a very tentative characterization of motion event encoding in these four ancient Indo-European languages. However, it serves as an illustration of the way in which contemporary and ancient data (when available) may be integrated in future work.

Classical Armenian has been described as verb-framed (Wälchli 2009: 215) and therefore it seemed appropriate to assign Classical Armenian the same values as Modern Armenian. In section 4.2, the satellite-framed nature of Homeric Greek and Vedic Sanskrit, which featured freely moving preverbs, and Latin, which possessed a fully productive set of path prefixes, has been discussed to some extent. These three ancient languages were assigned the values of German, which is clearly satellite-framed, but not as radically as some of the Balto-Slavic languages.

The results of the ancestral-state estimation analyses on the tree sample that incorporated these four ancient Indo-European languages are presented in Table 4.3.

Table 4.3*: Estimated ancestral states for Proto-Indo-European incorporating information on four ancient Indo-European languages*

Measure	**Median**	**Range**
score on PC1 (verb-framed vs. satellite-framed)	0.06	0.03–0.08
score on the PC2 (-deictic vs. +deictic)	0.02	0.003–0.03
proportion of use of the satellite-framed construction	0.31	0.29–0.33
proportion of use of the path-only construction	0.31	0.30–0.33
proportion of use of the deictic verb construction	0.07	0.06–0.08
proportion of use of the verb-framed construction	0.06	0.06–0.07
proportion of use of the deictic verb-framed construction	0.04	0.04–0.05

Table 4.3 indicates that the estimates for Proto-Indo-European, compared with those in Table 4.1, become oriented slightly towards the satellite-framed end of the Talmian scale. The satellite-framed construction is used slightly more often (0.31 compared with 0.26 in Table 4.1) and the score on the PC1 is slightly higher (0.06 compared with -0.001 in Table 4.1). However, compared with the mean values in Table 2.7, the estimated value is still intermediate: the score on the PC1 (0.06 compared with 0.05) is still in the middle of the Talmian scale, although it is now on the satellite-framed side of the scale. The same applies for the values for the satellite-framed construction and the path-only construction.

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

The following four trees were used for this robustness test. Three of the trees were maximum clade credibility trees, constructed using subsets of the tree sample used by Dunn et al. (2011). This tree sample is based on a slightly older version of the cognate data used by Bouckaert et al. (2012), which was used here because this tree sample includes more varied tree topologies: Whereas the Bouckaert et al. (2012) sample is completely resolved with regard to the position of the Germanic, Balto-Slavic and Romance subgroup, the Dunn et al. (2011) sample is not. This allows for the construction of several different maximum clade credibility trees that were used for additional ancestral-state estimation analyses. These included: 1. the full tree set (built from the full set of 1000 trees, similar to the Bouckaert et al. 2012 tree), 2. the sample of trees in which Balto-Slavic and Germanic form a clade (built from the 21 trees in which this subgroup existed), and 3. the sample of trees in which Balto-Slavic and Romance form a clade (built from the 32 trees in which this subgroup existed). The fourth tree is the best phylogenetic tree of the Indo-European language family presented in Nakhleh et al. (2005a): their 'tree A'. These four phylogenetic trees are visually presented in Appendix 4.

The results of these analyses are presented in Table 4.4. Table 4.4 gives an overview of the estimated ancestral states for Proto-Indo-European, using the scores on the PC1 and the PC2 and the proportion of the most frequently used motion encoding constructions as measures. A comparison between the estimates in Table 4.4 to the ranges presented in Table 4.1 indicates that the use of these different phylogenetic tree topologies also results in an intermediate/satellite-framed estimate for Proto-Indo-European. The score on the PC1 is still intermediate, but is now oriented towards the satellite-framed end of the scale (0.07 - 0.10, compared with -0.001 in Table 4.1). The scores on the satellite-framed construction are also higher (0.33 - 0.34, compared with 0.26 in Table 4.1) and the scores on the path-only construction are lower (0.28 - 0.31, compared with 0.33 in Table 4.1). However, this is by no means a very radical shift: these values suggest that Proto-Indo-European is positioned somewhere between English, the least radical satellite-framed language in the current sample, and Modern Greek, the least radical verb-framed language in the current sample (see also Figure 4.1). This indicates that the results presented in Table 4.1 are quite robust: different higher-order subgroupings of Indo-European do not have a large impact on the intermediate ancestral-state estimate that was found for Proto-Indo-European.

***Table 4.4**: Estimated ancestral states for Proto-Indo-European for different phylogenetic trees*

Tree sample[a]	PC Scores: Score on PC1 (v-framed vs. s-framed)	PC Scores: Score on PC2 (-deictic vs. +deictic)	Proportion of use measures: satellite-framed construction	path-only construction	deictic verb construction	verb-framed construction	deictic verb-framed construction
1.	0.08	0.002	0.33	0.29	0.07	0.06	0.06
2.	0.09	0.0005	0.34	0.29	0.07	0.06	0.06
3.	0.10	-0.006	0.34	0.28	0.08	0.05	0.06
4.	0.07	0.03	0.33	0.31	0.06	0.07	0.03

[a]Tree 1. Dunn et al.'s (2011) full tree set; Tree 2. Dunn et al.'s (2011) Balto-Slavic-Germanic tree; Tree 3. Dunn et al.'s (2011) Balto-Slavic-Romance tree; Tree 4. Nakhleh et al.'s (2005a) 'tree A'.

To test for areal patterns in the data, a partial Mantel test was carried out. The results of this test suggest that there is no positive correlation between the score on the PC1 and the distance between the nation capitals ($R = -0.08$, $p = 0.90$). These results suggest that languages that are spoken more closely to each other are not more similar because of their spatial proximity. This reinforces the view that these patterns are largely phylogenetically inherited rather than horizontally transmitted through borrowing.

4.5 Discussion

4.5.1 Interpreting the ancestral-state estimation analysis

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

The link between the scores and inferences presented in Tables 4.1, 4.2, 4.3 and 4.4 plus Figure 4.5 and the description of changes in the Proto-Indo-European preverb system as presented in section 4.2 is clear. Those subgroups in which the Proto-Indo-European preverb system has been retained in the form of a productive set of path prefixes or particles, i.e. Germanic and Balto-Slavic, are satellite-framed.[13] In contrast, the subgroups for which we have some evidence that this system was lost due to increasing unproductivity of the system and the univerbation of prefixes with verb roots, i.e. Romance, Indo-Iranian, Greek, and Albanian, are verb-framed.

In section 4.2, two hypotheses with regard to motion event encoding in Proto-Indo-European were put forward. On the one hand, Talmy (2007) and Acedo Matellán and Mateu (2008) have proposed that Proto-Indo-European was satellite-framed. On the other hand, the review of the historical evidence from ancient Indo-European languages indicated that although all ancient Indo-European languages could use the satellite-framed construction, this strategy might not have been the construction that was used most often (Brucale 2011; Brucale et al. 2011; Brosch 2014). The review of the historical comparative data thus suggests that Proto-Indo-European was typologically mixed. Since the ancestral-state estimation analyses estimate that Proto-Indo-European was mixed, but placed on the satellite-framed side of the Talmian scale, they provide more support for the second hypothesis. Given these results and the evidence from the historical record of ancient Indo-European languages, it seems clear that Proto-Indo-European was not satellite-framed on the same level as the Germanic and most of the Balto-Slavic languages, rather it had a mixed typological nature.

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

> A maximum-likelihood estimate of an ancestor state for a continuous trait is simply a weighted average of the dimensions of the extant species at the tips of the tree (eq. [4]). Reconstructed ancestor states will tend to be intermediate for this reason. […] This analysis suggests that ancestor reconstructions for continuous traits are often too variable to be of much

[13] Note however that the ancestral state estimate for Proto-Balto-Slavic presented in Table 4.2 and Figure 4.5 is 0.14, which seems to indicate a less satellite-framed character as Proto-Germanic. This is due to the mixed or even verb-framed nature of Serbo-Croatian, which will be discussed further below.

> use, except to place ancestor sizes within broad limits. (Schluter et al. 1997: 1706).

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

However, the intermediate estimate for Proto-Indo-European is in line with the comparative historical information reviewed in section 4.2. Given the current results and the comparative historical record, Proto-Indo-European could in fact well be a language that once was highly satellite-framed but that was already losing the productivity of the preverb system. Some indications that this might be true have already been given in section 4.2. More evidence can be found in the idiomatic meanings of preverb-verb combinations. Whitney (1879: 352) notes that preverb-verb combinations in Vedic Sanskrit may undergo a shift in meaning, although it is never so radical that the change cannot be related to the basic meanings of the preverb and the verb (see also Ringe 2006: 58-59). Danesi (2013) gives several examples of preverb-verb idioms, including the preverb-verb combination *prá vac* 'to announce, proclaim', from *prá* 'forward, fore' and *vac* 'to speak' (Danesi 2013: 62). It seems clear that even though the preverb system was highly productive in the Rigveda, a process of lexicalization or univerbation in which the meaning of the preverb-verb combination was no longer transparent was already taking place.

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

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

4.5.2 A possible explanation and further theoretical implications

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

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

The existence of a satellite-framed area in North and Central Europe could also potentially explain Irish and Serbo-Croatian motion event encoding. Irish has lost its verbal path prefixes, unlike the majority of the Germanic and Balto-Slavic languages, but is still satellite-framed. Its geographical location close to English might have prevented Irish from becoming more verb-framed. To investigate this further we would need to compare Irish with the other Celtic languages.[14] Serbo-Croatian is currently undergoing the univerbation process: Serbo-Croatian's verbal prefixes have been merging with the deictic verb *ići* 'to go' to create a range of path verbs. It therefore seems to be currently in the process of becoming a verb-framed language, using the satellite-framed construction less often and the path-only construction and verb-framed construction more often. The reason that Serbo-Croatian is going through this process might be due to its southern location. However, Serbo-Croatian is the only Southern Slavic language in my sample; a comparison with Macedonian, Bulgarian and Slovenian should be made in order to see whether the same process is currently affecting these languages as well.

We have seen that change in the motion event encoding system in Indo-European can be explained at least partly by the first of Croft et al. (2010)'s grammaticalization pathways (see (37a)). Verbs and satellites have merged in all Indo-European subgroups. In addition, coordination is more common in verb-framed and non-satellite-framed languages than it is in satellite-framed languages, suggesting that coordination is a proper mechanism to code manner and path information in one sentence if the satellite-framed construction is not preferred. However, it is also clear that Croft et al. (2010)'s grammaticalization pathways do not describe all the changes that might occur. The two grammaticalization pathways proposed by them both start with coordination, suggesting that all types of satellites should arise from verbs. This is not true, as path satellites do not necessarily arise from verbs (Lehmann 1985; Stevens 1992), although in some cases they do (Crapo 1970). A full description of all the different grammaticalization pathways that are possible would be needed to characterize change in motion event encoding completely.

4.6 Conclusion

The ancestral-state estimation analysis conducted in this chapter indicates that Proto-Indo-European motion event encoding can be placed in the middle of the Talmian scale, with a slight tendency towards the satellite-framed end of the scale. This result supports the picture of Proto-Indo-European motion event encoding that has emerged from studies of Classical Latin (Brucale 2011; Brucale

[14] The parallel corpus also includes a Breton translation of *Alice's Adventures in Wonderland* that was not included in the current dataset. A comparison of the Breton translation with the other 20 translations of *Alice's Adventures in Wonderland* indicates that Breton behaves in a very similar way as Irish. This is unexpected, as one would expect Breton to be influenced by French to a much larger extent than is suggested by this comparison. Further investigations may be able to shed more light on potential contact influences on Celtic motion event encoding.

et al. 2011) and Hittite (Brosch 2014), idiomatic meanings of preverb-verb combinations in ancient Indo-European languages, etymologies of Romance path verbs, and other comparative historical evidence.

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

An analysis that incorporates this additional information may potentially still generate a classification of Proto-Indo-European that is more in line with the claims by Acedo Matellán and Mateu (2008) and Talmy (2007). However, the claim put forward in this chapter is that the productivity of the preverb system used for satellite-framed motion event encoding constructions was already declining in Proto-Indo-European, i.e. that Proto-Indo-European had an intermediate or mixed system of motion event coding. This seems valid both in light of the ancestral-state estimation analysis conducted in this chapter and in light of several other types of evidence discussed in section 4.2 and 4.5. In this scenario, Germanic and Balto-Slavic might have been able to become more satellite-framed over time due to contact with surrounding satellite-framed languages, while Romance and Indo-Iranian did not have this contact and shifted towards the verb-framed pattern.

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

Chapter 5: The correlation between motion event encoding and manner verb and path verb lexicon size

This chapter is a revised version of:
Verkerk, Annemarie. (2013). Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. Language Dynamics and Change, *3(2): 169-217. Special issue, "Phylogeny and beyond: Quantitative diachronic approaches to language diversity", edited by Michael Dunn.*
and:
Verkerk, A. (2014b). The correlation between motion event encoding and path verb lexicon size in Indo-European. Folia Linguistica Historica*.*

In recent decades, much has been discovered about the different ways in which people can talk about motion (Talmy 1985, 1991; Slobin 1996b, 1997, 2004). Specific claims regarding the relation between motion event encoding and the motion verb lexicon have been brought forward. 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. Özçalışkan (2004) suggests that satellite-framed languages and verb-framed languages have approximately the same number of path of motion verbs (such as *exit*, *descend*, and *follow*), as the path domain provides less semantic space for an expanded lexicon. In the current chapter, the correlation between the motion event encoding system and the size of the manner verb lexicon and the path verb lexicon is tested. The data come from parallel texts in 20 Indo-European languages. The methodology applied is a range of phylogenetic comparative methods adopted from biology, which allow for an investigation of these dependencies while taking into account the shared history between these 20 languages. This chapter shows that there is some evidence that Slobin's hypothesis is correct while Özçalışkan's hypothesis is incorrect, as there seems to be a relationship between the motion event encoding system and both the size of the manner verb lexicon and the size of the path verb lexicon.

5.1 Introduction

5.1.1 Motion event encoding and the lexicon

This chapter is concerned with manner of motion verbs (manner verbs) such as *walk*, *swim* and *dash* and path of motion verbs (path verbs) such as *enter*, *ascend*, and *pass*: how is information on the manner and the path of motion lexicalized in verbs, and is the creation of these verbs influenced by the syntactic constructions in which they are used? Recent years have seen an increase in studies in the semantic and lexical typological traditions that address these sorts 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 which words are typically basic and which words are typically derived or peripheral in the languages of the world (Goddard 2001). Manner verbs and path verbs have not been investigated in this tradition until now.

The largest amount of information on manner verbs and path verbs has been gathered in the context of motion event encoding, which has been studied extensively over the last few decades. Talmy (1985, 1991) 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 path outside of the verb, on so-called satellites (satellite-framed construction). The use of these two motion event encoding constructions has been described for a range of different languages during the past decades. Examples from my own corpus of parallel texts are provided in (45) and (46).

45) Dutch

En	*ze*	*liep*	*zo*	*hard*	*als*	*ze*
and	3SG.F.SBJ	run.PST.SG	as	fast	as	3SG.F.SBJ

kon	*naar*	*het*	*deur-tje;*
can.PST.SG	to	DEF.ART	door-DIM

'And she ran as fast as she could to the small door'

46) French

Elle	*repart-i-t*	*à*	*tout-e*	*allure*	*vers*
3SG.F.SBJ	return-PRET-3SG	with	all-F	speed.F	to

la	*petit-e*	*porte*
ART.DEF.F	small-F	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 environments (Talmy 1985). The path of movement in (45) and (46) is the protagonists' 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). The manner of movement in (45) and (46) is running or moving very fast.

The Dutch sentence in (45) encodes the path of motion on the preposition *naar* 'to', and the manner of motion on the manner verb *lopen* 'run', as well as in the adverbial phrase *zo hard als ze kon* 'as fast as she could'. The French sentence in (46) encodes the path of motion on the path 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 (verb-framed construction), while the Dutch sentence encodes this information on a locative particle or satellite (satellite-framed construction).

However, it is not true that French uses only the verb-framed construction, and Dutch uses only the satellite-framed construction. French makes use of manner verbs in certain contexts. Aske (1989) was the first to point out that verb-framed languages such as French were 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. An example of such an atelic motion event in which the satellite-framed construction is used is given in (47). In this atelic motion event, motion does take place, but the motion simply takes place within a single location, as there is no endstate postulated in which the person finds itself as the event is finished. This means that in (47), the reading that the protagonists are still swimming is possible as well as the reading that they have reached the shore. Dutch can make use of path verbs as well as manner verbs, as exemplified in (48). Here, *oversteken* 'to cross' is a path verb with a separable path prefix *over* 'across'. In independent clauses, the object of the verb *oversteken* appears between the main root of the verb and the prefix: *ze stak het beekje over* 'she crossed the small brook'. This verb is a historic remnant of the Proto-Indo-European preverb system, as discussed in section 4.2.

47) 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 towards the shore.'

48) Dutch

… terwijl	*ze*	*na*	*de*	*Koningin*
while	3SG.F.SBJ	after	DEF.ART	queen

het	*beek-je*	*overstak.*
DEF.ART	brook-DIM	cross.PST.SG

'… while she crossed the small brook after the Queen.'

It is clear that most languages have a set of manner verbs, because they may use them in situations like (47), while most languages also have a set of path verbs, as they can use them in contexts where the path of motion is the most relevant dimension of the motion event, as in (48). In addition, languages typically make use of more than one motion event encoding construction, as has been shown by Beavers et al. (2010), Croft et al. (2010), Slobin (2004), and others (see chapter 3). The question that is asked in this chapter is whether there exists a correlation between the use of motion event encoding constructions, and the size of the manner verb lexicon and the path verb lexicon. In order to do that, a review of the hypotheses regarding these correlations is presented in the next two sections.

5.1.2 Manner of motion verbs

Slobin (1997, 2005b) hypothesized that there are differences in the structure and the size of the manner verb lexicon between satellite-framed and verb-framed languages. Slobin (1997) proposed that there are two different classes 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 pay a role in the

> considerable smaller second tiers in V-languages [verb-framed languages, AV]. (Slobin 1997: 459).

This difference in the size of second tier manner verbs has consequences for the overall size of the manner verb lexicon: "S-languages [satellite-framed languages, AV] will have a larger and more diverse lexicon of manner verbs, in comparison with V-languages [verb-framed languages, AV]" Slobin (1997: 458). Manner verb lexicons are thus predicted to have different sizes depending on the motion encoding construction that is typically used by the language. 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 kind of verb.

In later work, Slobin (2003: 165ff, 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).

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 a satellite-framed construction, the manner verb slot in a serial verb construction, and the manner ideophone (amongst others). From Slobin's (2004) hypothesis 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) proposed a similar relationship between lexicalization patterns and the lexicon.

What evidence is there so far that satellite-framed languages have larger manner verb lexicons? Koptjevskaja-Tamm et al. (2010: 339ff) pointed out that there are differences in the size of separate manner subdomains that do not align with the typical syntactic patterns of satellite-framed and verb-framed languages. They show that Slavic languages are far less elaborate in their encoding of the subdomain of aquamotion than Germanic and Romance languages. Iacobini (2009) made a similar point when he discussed 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, Slobin's (2004) hypothesis seems to hold up for several individual languages. Slobin (2005b) 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). Jovanovic and Martinovic-Zic (2004) show that Serbo-Croatian has a larger manner verb lexicon than English. 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) demonstrated 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.

5.1.3 Path of motion verbs

If Slobin's (1997, 2003, 2004, 2006) hypothesis is true, and satellite-framed languages typically have a larger manner verb lexicon as compared with verb-framed languages, it could be the case that verb-framed languages have a larger lexicon of path of motion verbs (path verbs) such as *exit*, *ascend*, and *pass* as compared with satellite-framed languages. On the other hand, however, Özçalışkan (2004) has proposed that both verb-framed and satellite-framed languages should approximately have the same number of path verbs because the path verb lexicon "does not provide many options for elaboration". From one perspective this seems to be correct, because the range of possible purely directional paths of motion seems quite small as compared with for instance the range of possible manners of motion: they can relate to an abstract endpoint (*enter*), sourcepoint (*exit*), mid-point (*pass*, *cross*) or to the basic directions up (*ascend*), down (*descend*), forward (*advance*), back (*return*), around (*circle*), behind, or in front.

Despite this limited set of abstract directions, there are two reasons why the set of path verbs can be large in some languages. The first is that many languages have path verbs that refer to a far more varied set of more or less abstract reference points for motion in the environment. Where *leave* refers to an abstract sourcepoint, Jahai, a Mon-Khmer language spoken in the Malay Peninsula, features verbs that refer to the flow of the water in rivers and their tributaries, distinguishing verbs such as *rkruk* 'to move along the main river (in both upstream and downstream direction)', *piris* 'to move across the flow of water', *dey* 'to move upstream on a tributary' and *həc* 'to move downstream on a tributary' (Levinson and Burenhult 2009). A similar system exists for other geographical features such as mountain ridges, which "distinguish 'motion lengthwise on mountain ridge' vs. 'motion across mountain ridge', as well as 'motion lengthwise on mountain side' vs. 'motion up on mountain side'/'motion down on mountain side'." (Levinson and Burenhult 2009: 161). In addition, languages with sizable path verb lexicons may also feature a range of verbs that have highly similar semantics. Cifuentes Férez (2010) lists 14 Spanish path verbs that relate movement away from a sourcepoint, out of which seven are highly similar (original glosses given): *apartarse*, *distanciarse*, *ladearse* 'to move away from'; *lagarse*, *marcharse*, *partir* 'to leave'; *pirarse* 'to go away' (informal)'. It is clear from this that languages can have big path verb lexicons even if they encode only the most basic set of directional paths.

A review of the literature suggests that there is support for the hypothesis that verb-framed languages have larger path verb lexicons than satellite-framed languages. An overview of path verb lexicon sizes reported in the literature has been provided in Tables 5.1 and 5.2.[15] Table 5.1 provides an overview of different types of studies (experimental, corpus-based, or dictionary-based) that provide information on path verb lexicon size on a satellite-framed language in comparison with different verb-framed languages. Table 5.2 presents path verb lexicon size data from a set of different languages. Note that these numbers are not comparable across languages, as they have not used the same methodology.

***Table 5.1**: Comparisons of path verb lexicons in the motion event encoding literature*

Satellite-framed language	**No. of path verbs**	**Verb-framed language**	**No. of path verbs**	**Source**
English	14	Basque	37	Ibarretxe-Antunãno (2004)
English	24	Hindi	14	Narasimhan (2003)
English	11	Serbo-Croatian	12	Jovanovic & Martinovic-Zic (2004)
English	4	Spanish	8	Naigles et al. (1998)
English	44	Spanish	65	Cifuentes Férez (2010)
English	7	Turkish	13	Özçalışkan (2009)
English	20	Turkish	24	Özçalışkan (2004)
German	36	French	37	Berthele (2006)

Table 5.1 gives an overview of the path verb lexicon sizes in studies that compared two languages, in this case mostly English, being a satellite-framed language, with a verb-framed language. It indicates that although English has a large class of path verbs based on dictionary consultation, as done by Narasimhan (2003) and Cifuentes Férez (2010), in other types of studies English has a smaller class of path verbs than the verb-framed languages to which it is compared.

[15] The inventories listed in Table 5.1 and 5.2 may also include deictic verbs.

***Table 5.2**: Reported sizes of path verb lexicons for individual languages in the motion event encoding literature*

Language	Number of path verbs	Source
Chinese (contested)	13	Chen and Guo (2009)
German (satellite-framed)	'virtually none'	Wienold (1995)
Indonesian (verb-framed)	15	Wienold (1995)
Japanese (verb-framed)	33	Matsumoto (1997) as cited in Matsumoto (2003)
Japanese (verb-framed)	36	Wienold (1995)
Korean (verb-framed)	35	Wienold (1995)
Sidaama (verb-framed)	13	Kawachi (2011)
Thai (equipollently-framed)	19	Wienold (1995)

Table 5.2 gives an overview of some reported path verb lexicons for several individual languages. In Table 5.2 German, the only satellite-framed language for which data on path verb lexicon size could be found, was reported to have the smallest path verb class by Wienold (1995), although Berthele (2006) in Table 5.1 has a very different opinion. The differences between satellite-framed and verb-framed languages seems striking at first, but Matsumoto (2003) writes that the size of the path verb and manner verb lexicon is at least to some extent independent from the strategy that is typically used to encode motion events. For instance, Japanese has around 33-36 path verbs, while Sidaama has 13, even though both are verb-framed (see Table 5.2). It is therefore likely that there are many different factors that determine how many path verbs a language has.

5.1.4 Hypotheses

As we have seen in the previous two sections, several claims with regard to the relation between the use of motion event encoding constructions on the one hand and the manner verb lexicon and path verb lexicon on the other hand have been put forward. Indeed, there exists a large field of studies that investigate the relationship between spatial language and cognition, i.e. the factors that influence the structure and size of spatial semantic domains (Levinson 2003; Majid et al. 2004). Ember (1978) looked at the impact of cultural and biological factors on the growth of basic color term systems. However, a study that investigates Slobin's (2004) hypothesis with regard to manner verbs and Özçalışkan's (2004) hypothesis with regard to path verbs on a language family-wide scale has not been carried out. This chapter aims to make one of the first contributions to answering these questions.

The current chapter investigates the hypothesis that there exists a correlation between the motion event encoding system and the size of the manner verb lexicon and the path verb lexicon. Specifically, the following two questions are asked:

1. Is there a correlation between the more frequent use of the satellite-framed construction and a higher PC1 score (see section 5.2.1) and a larger manner verb lexicon?
2. Is there a correlation between the more frequent use of the verb-framed construction and the path-only construction and a lower PC1 score (see section 5.2.1) and a larger path verb lexicon?

Since the languages in this dataset are all genealogically related, it is necessary to make use of phylogenetic comparative methods, as was explained in chapter 1. Mace and Pagel (1994) and Levinson and Gray (2012), 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) has been inherited from a common ancestor, 2) has been borrowed from one language to the others due to their close proximity, or 3) has 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 motion event encoding system and the size of the manner verb and the path verb lexicon, while taking into account the phylogenetic relationships that cause these types of interdependencies between languages.

This chapter investigates whether correlations exists between the motion event encoding system and the lexicons of manner verbs and path verbs. A dataset based on a sample of 20 Indo-European languages was used. This dataset is presented in section 5.2. To investigate the correlations, phylogenetic comparative methods adopted from biology are used. These methods are discussed in section 5.3. The results are presented in section 5.4. A discussion of results and general conclusion are provided in section 5.5.

5.2 Dataset

5.2.1 Samples and motion event encoding measures

In this chapter, all four samples introduced in section 2.1 are used:

1. the 118-sentence sample to measure the motion event encoding system for all 20 languages;
2. the 132-sentence sample to measure manner verb lexicon size and path verb lexicon size for all 20 languages;
3. the 192-sentence sample to measure the motion event encoding system for 16 languages (excluding Albanian, Persian, Hindi and Nepali);
4. the 215-sentence sample to measure manner verb lexicon size and path verb lexicon size for 16 languages (excluding Albanian, Persian, Hindi and Nepali).

In this section, the measures with regard to the motion event encoding system are introduced briefly. The next two sections (section 5.2.2 and 5.2.3) provide information on the coding of manner verbs and path verbs.

To measure the behavior of the 20 languages with regard to the motion event encoding system, six different measures were used. For analyses on both manner verb lexicon size and path verb lexicon size, the scores on the first principal component (PC1) were used. Results of the principal components analyses were presented in section 2.3. Figures 2.2 and 2.3, which present the results of these analyses, have been reproduced below as Figure 5.1 and 5.2 for convenience. The first two measures are:

1. The PC1 score for the 118-sentence sample for all 20 languages;
2. The PC1 score for the 192-sentence sample for 16 languages (excluding Albanian, Persian, Hindi and Nepali).

Additionally, the proportion of use of the satellite-framed construction was used as a measure of the motion event encoding system in order to investigate manner verb lexicon size. This measure was used because it allows for a direct test of Slobin's (2004) hypothesis. However, note that it is a continuous measure instead of the discrete measure (the dichotomy between satellite-framed and verb-framed languages) originally proposed by Slobin (2004). This reflects the more recent consensus that the typology is clinal rather than absolute (see chapter 3).

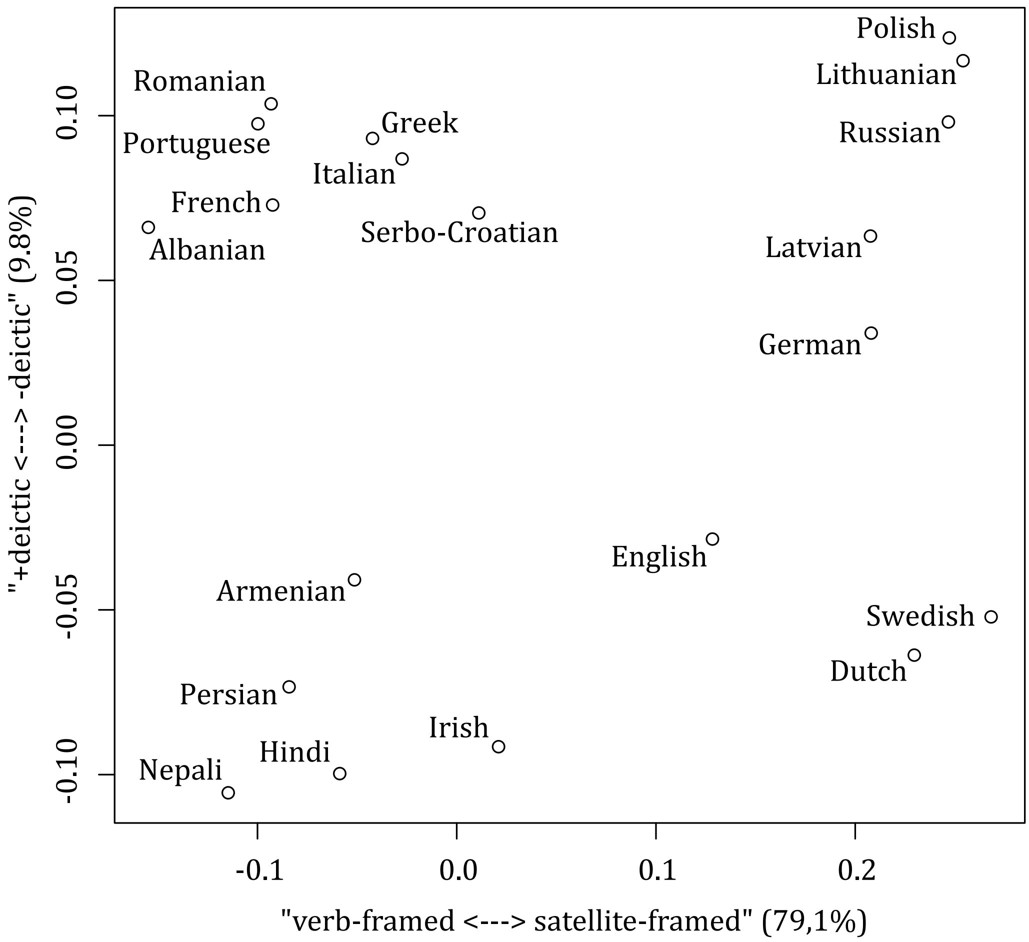


***Figure 5.1**: A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 118-sentence sample for 20 Indo-European languages*

The second two measures therefore are:

3. The proportion of use of the satellite-framed construction in the 118-sentence sample for all 20 languages;
4. The proportion of use of the satellite-framed construction in the 192-sentence sample for 16 languages (excluding Albanian, Persian, Hindi and Nepali).

For the analyses investigating path verb lexicon size, the proportion of use of the path-only construction plus the proportion of use of the verb-framed construction was used as a measure of the motion event encoding system. This measure is a continuous scale that indicates how often a construction with a path verb as the main verb of the sentence is used by each of the languages in the sample. The last two measures therefore are:

5. The proportion of use of the path-only plus the verb-framed construction in the 118-sentence sample for all 20 languages;
6. The proportion of use of the path-only plus the verb-framed construction in the 192-sentence sample for 16 languages (excluding Albanian, Persian, Hindi and Nepali).

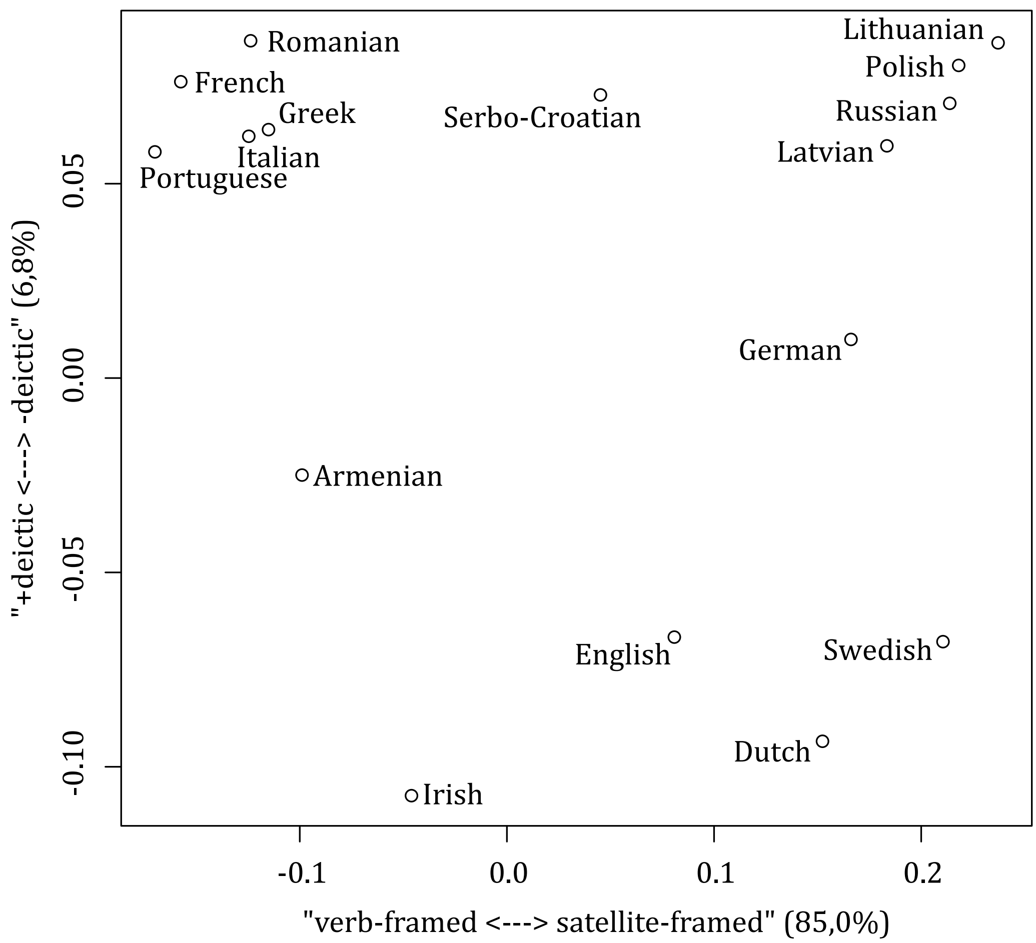


***Figure 5.2**: A phylogenetic principal components analysis conducted on the percentage of usage of each motion encoding construction in the 192-sentence sample for 16 Indo-European languages*

5.2.2 Coding of manner of motion verbs

Although the terms 'manner' and 'path' have been central to studies of motion since Talmy (1985), a cross-linguistic classification of verbs with respect to these features is not without difficulties. Many verbs seem to indicate something about both manner and path. English *climb* and Dutch *klimmen* are good examples, because they can be used for all kinds of paths, including up, down, into, and out of a reference object. Without further specification of direction, however, they indicate movement upwards. Another example is the Hindi verb *bhāgnā*, which may signify both 'run' and 'run away'. In other words, many manner verbs seem to have a path preference. In coding manner verbs for the purpose of this dissertation, 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 sorts of path - English *climb* and Dutch *klimmen* are therefore classified as manner verbs. If a verb indicates both a clear manner and a path, 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 (2005b) and Hsiao (2009), among others, are exceptions. For the current study, the manner verbs that are used emerge from the restricted set of sample sentences that have been 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 list used here is not a complete list of manner verbs in these languages. However, the most important ones relating to various kinds of manner are included. The benefit of this approach is that the manner verbs that are most commonly used are included, while obsolete or less commonly used verbs (as one might find in dictionaries) are not included. 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. An overview of the manner verbs studied in this chapter is presented in Appendix 5.

There are two categories of verbs that are not included as manner verbs in this study that deserve some attention before the actual manner verbs are discussed. These are complex manner verbs and non-motion verbs used as manner verbs.

Complex manner verbs are not considered as part of the quite narrow definition of manner verb that is employed in this study. The reason for this coding decision was that complex 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 complex verbs that were attested in the samples are listed in Appendix 6 in order to provide an overview. From these tables, it seems clear that complex verbs are most common in Germanic and Romance, while they are not common in any of the other languages.

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 2006; 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 for two cases in which sound-emission verbs are used as manner verbs from the current datasets are provided in (49) and (50).

49) Serbo-Croatian

Jao	*eno*	*od-e*	*njegov*
Oh	there	go.PFV-PRS.3SG	3SG.M.POSS.M.NOM.SG

zlat-n-i	*nos-ić! —*	*Golem*
gold-ADJ-M.NOM.SG	nose-DIM.M.NOM.SG	huge.M.NOM.SG

tiganj	*bi-o*	*je*	*odista*
saucepan.M.NOM.SG	be-PST.ACT.PTCP.M.SG	be.PRS.3SG	really

pro-zvižda-o	*mimo*	*dijete*
PRFX-whistle.IPFV-PST.ACT.PTCP.M.SG	next.to	child.N.ACC.SG

i	*gotovo*	*mu*	*odvali-o*
and	almost	3SG.M.DAT	bear.off.PFV-PST.ACT.PTCP.M.SG

nos.
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.'

50) Swedish

Under	*tid-en*	*roa-de*	*sig*	*mygga-n*
under	time-SG.DEF.UT	amuse-PST	REFL.3SG.N	gnat-SG.DEF.UT

med	*att*	***surra***	*runt*	*huvud-et*	*på*	*henne.*
with	to	hum.INF	round	head-SG.DEF.N	on	3SG.F.OBJ

'While the gnat amused itself by humming around her head.'

Narasimhan (2003: 130ff) points out that while the constructions presented in (49) and (50) are common in satellite-framed languages, they are not possible in verb-framed Hindi. The current data support this claim, since translating this sort of sentence seems troublesome in other verb-framed languages as well. The verb-framed languages in the current sample typically translate the sentence in (49) with verbs meaning ‘pass’, ‘fly’, or ‘brush’.

In the current analysis, only verbs that signify 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, 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. This strict definition is needed because verbs like *whistle* and *skim* cannot be said to co-evolve with motion event encoding construction usage, as they belong to different verb lexicons.

However, since there might be interesting cross-linguistic differences in the use of these verbs, the ones that were encountered in the sample have been listed in Appendix 7. These verbs belong to one of the following classes: 1) ‘self-contained motion verbs’ (*flutter*, *wriggle*), 2) sound emission verbs (*rattle*, *hum*, *splash*), 3) verbs that are used in the subordinate construction (*struggle*, *have trouble to*), 4) verbs that are used in a metaphorical sense (*skim*), and 5) verbs that signify increasing or diminishing speed. In line with Narasimhan’s (2003) claims on the use of this class of verbs 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 types of manner verbs encountered in the samples is presented in Appendix 5. Table 5.3 lists the numbers of types of manner verbs attested in the 132-sentence and the 215-sentence sample for each language. Figure 5.3 presents these numbers visually. Table 5.3 and Figure 5.3 are ordered per subgroup. Appendix 8 lists the numbers of types of manner verbs for each book individually.

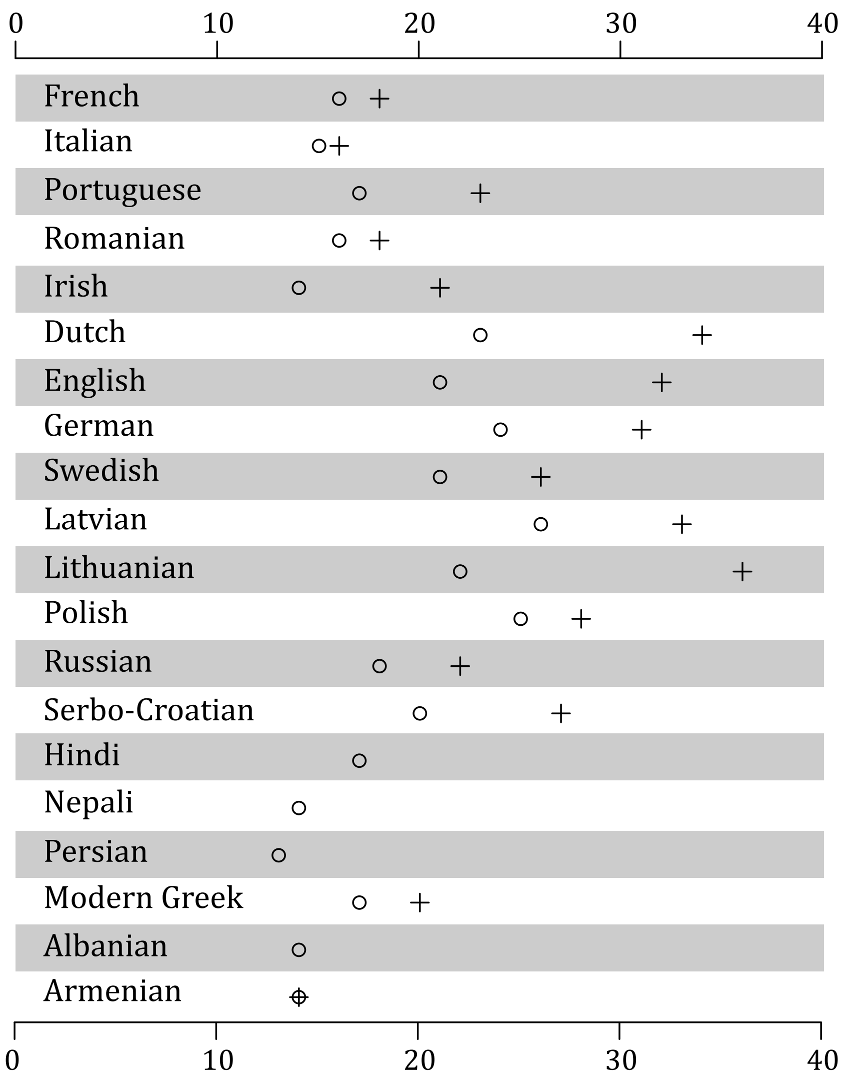


***Figure 5.3**: Number of types of manner verbs in the 132-sentence and 215-sentence samples (○ = 132-sentence sample, + = 215-sentence sample)*

From Table 5.3, Figure 5.3, Appendix 5 and Appendix 8 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. Languages that are more closely related resemble each other more than languages that are less closely related.

Appendix 5 suggests that every language in the 132-sentence sample has a manner verb for the most prototypical kinds of manner, i.e. a word for RUN, FLY, SWIM, and WALK. RUN and WALK are also the meanings for which both satellite-framed and verb-framed languages typically have more than one verb. Most languages have at least one verb that denotes STROLL alongside the general verb WALK. Other kinds of manner that are lexicalized in all languages include ROLL (except Albanian), JUMP (both some satellite-framed and some verb-framed languages have more than one verb for jumping), DASH (lexicalized by three to six different verbs in both some satellite-framed and some verb-framed languages), and WANDER/ROAM (except Persian). STEP, TREAD SLIDE/SLIP are lexicalized by most satellite-framed and most non-satellite-framed languages.

Table 5.3: *Number of types of manner verbs in 20 Indo-European languages*

Language	132-sentence sample	215-sentence sample
French	16	18
Italian	15	16
Portuguese	17	23
Romanian	16	18
Irish	14	21
Dutch	23	34
English	21	32
German	24	31
Swedish	21	26
Latvian	26	33
Lithuanian	22	36
Polish	25	28
Russian	18	22
Serbo-Croatian	20	27
Hindi	17	-
Nepali	14	-
Persian	13	-
Modern Greek	17	20
Albanian	14	-
Armenian	14	14

Domains in which satellite-framed languages have more verbs than non-satellite-framed languages include CRAWL (only some satellite-framed languages and Modern Greek feature more than one word for CRAWL), SCRAMBLE (only some satellite-framed languages feature a word for SCRAMBLE), GALLOP and TROT (more satellite-framed languages have verbs for GALLOP and TROT than non-satellite-framed languages), 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), MARCH, SNEAK, and RIDE. The manner meaning CLIMB is lexicalized only sporadically and mostly by satellite-framed languages, in the other languages climbing manner is typically lexicalized as a verb meaning ‘climb up’, having both path and manner, which would be classified as a manner plus path verb by the coding rules employed in this dissertation (see section 2.2.2).

Slobin (1997) claimed that manner verbs can be divided into a first tier of more general or neutral verbs and a second tier of more specific and expressive verbs. Satellite-framed languages have extensive second tier manner verb lexicons, while the first tier 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 5.3: there exists a clear set of meanings that are lexicalized into verbs in each of these languages, including WALK, STROLL, RUN, FLY, SWIM, ROLL, JUMP, RUSH/HURRY, and WANDER/ROAM. The other semantic subdomains are lexicalized into verbs in some languages and not in others, and usually more of these semantic fields are lexicalized by satellite-framed languages and not by verb-framed languages. In addition, satellite-framed languages typically have more verbs in a given semantic subdomain than verb-framed languages. See for instance the use of six different verbs for the meaning DASH in Lithuanian and Serbo-Croatian and the use of four different verbs for JUMP in English.

This section has shown that there is considerable variation in the size of the manner verb lexicon in the current sample of 20 Indo-European languages. The data suggest that a core set of manner verb types is shared by all languages, and extension of the manner verb lexicon is present 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 creation of these 20 manner verb lexicons, section 5.3 and 5.4 will focus on the influence of the use of the satellite-framed construction.

5.2.3 Coding of path of motion verbs

The classification of a verb as a path verb was done as follows. As noted in chapter 2, the semantics of each motion verb were discussed with a native speaker. In most cases, if a verb encoded the path of motion, this was often clear from the start of the discussion as it was key to the meaning of the verb. If a verb encoded both path and manner it was coded as a manner plus path verb (see chapter 2). Examples of manner plus path verbs are Modern Greek *koytroyvalo* 'tumble down', Lithuanian *kopti* 'climb up', and Persian *goriḵtan* 'run away'. In the motion event encoding literature, these verbs are sometimes included in the manner verb class, and sometimes in the path verb class. Because this class of verbs is semantically different from both the class of path verbs and the class of manner verbs, as they encode both manner and path, and because the origins and use of this verb class are interesting topics of investigation in their own right, they are not included in the path verb class in the current study. The most important criterion for coding path verbs, aside from their encoding of a path of motion, was whether the verb could be used in various different manner contexts. If it could be used with different manners of motion, it was coded as a path verb. For instance, the English path verb *enter* can be used both in combination with the adverb *speedily* as well as with the manner verb participle *crawling*, and is therefore coded as a path verb.

Several path verbs in the current dataset have originated from merged path prefixes + verbs. This is due to the ancient Indo-European system of spatial verbal prefixes or 'preverbs' that has become unproductive in certain branches of Indo-European, such as Romance and Indo-Aryan, while still being productive in others, such as Balto-Slavic (Watkins 1964; Kuryłowic 1964; see also chapter 4). This kind of path verb may be inherited from a older stage of the language, such as French *descendre* 'descend' from Latin *descendere* 'descend', which is a combination of the productive Latin prefix *de-* + verb *scandere*. Or they may be more recent combinations, such as French *retourner* 'return' from French *re-* + *tourner*. In some cases, the original prefix and verb may not have merged in the same way as French *descendre* and *retourner*, note for instance the separable verbs in Dutch (such as *terugkeren* 'return') and German (*zurückkehren* 'return'). Especially for the Slavic languages, of which many still have a productive system of spatial prefixes on verbs, it was sometimes difficult to decide when a prefix + verb combination should be coded as a path verb or not. The criterion used here was that if the verb was a motion verb without the prefix, this was a productive use of that motion verb with an added path satellite in the form of a prefix. However, if the verb without the prefix was not a motion verb, i.e. when the meaning of the path verb could no longer be derived from the meaning of the prefix plus the meaning of the verb, the prefix + verb combination as a whole was classified as a path verb. An example of such a path verb is Polish *przybyć* 'arrive', a derivation with a spatial prefix *przy-* with the verb *być* 'to be'.

A full overview of the path verbs encountered in the samples is presented in Appendix 9. Table 5.4 lists the numbers of types of path verbs attested in the 132-sentence sample and the 215-sentence sample for each language. Figure 5.4 presents these numbers visually.

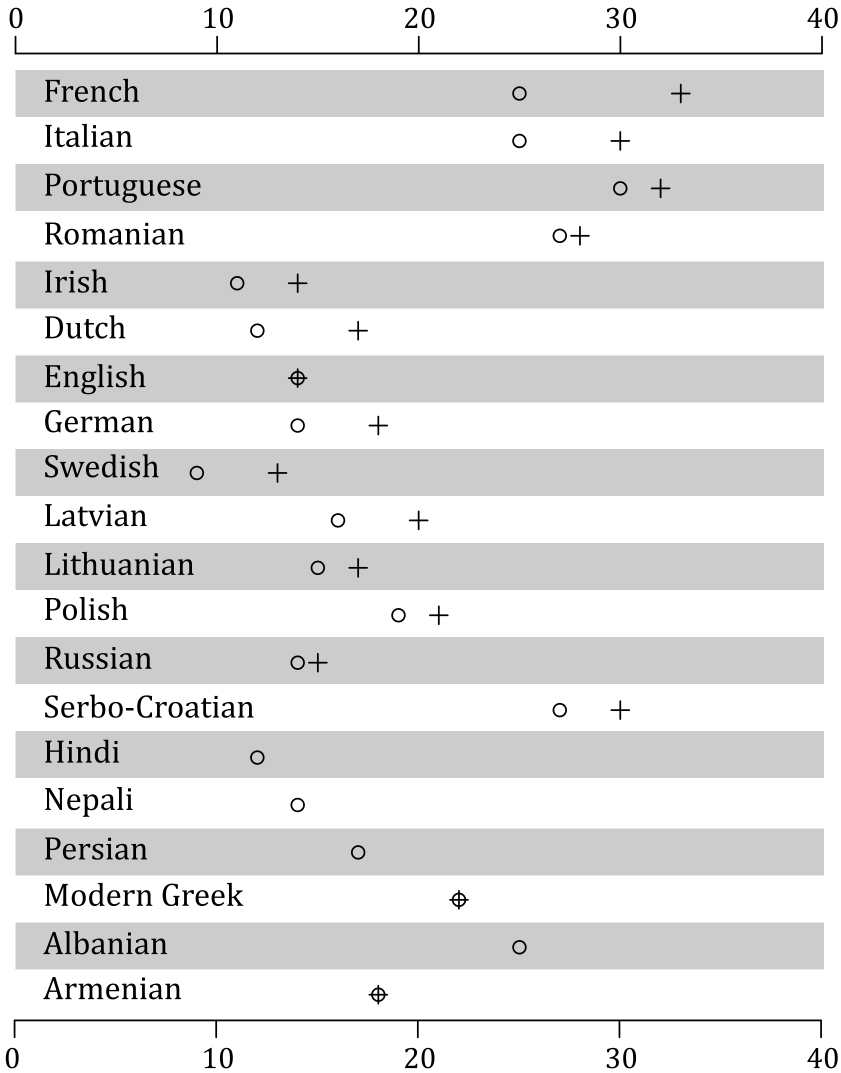


Figure 5.4: *Number of types of path verbs in the 132-sentence and 215-sentence samples (○ = 132-sentence sample, + = 215-sentence sample)*

From Appendix 9, as well as from Table 5.4 and Figure 5.4 it becomes clear that there are differences between the number of path verbs from language to language: for the 192-sentence sample, Swedish has the lowest number of path verbs (13), while French has the largest number (33). Some verb-framed languages (French, Portuguese, Italian, Romanian, Serbo-Croatian, Albanian) have quite large path verb lexicons, while satellite-framed languages typically have small path verb lexicons. In addition, languages that are closely related, such as French, Portuguese, Italian and Romanian, have a similar number of path verbs: between 28 and 33 for the 192-sentence sample.

Table 5.4: *Number of path verbs per sample*

Language	132-sentence sample	215-sentence sample
French	25	33
Italian	25	30
Portuguese	30	32
Romanian	27	28
Irish	11	14
Dutch	12	17
English	14	14
German	14	18
Swedish	9	13
Latvian	16	20
Lithuanian	15	17
Polish	19	21
Russian	14	15
Serbo-Croatian	27	30
Hindi	12	-
Nepali	14	-
Persian	17	-
Modern Greek	22	22
Albanian	25	-
Armenian	18	18

Appendix 9 suggests that the verb-framed languages in the sample, most importantly French, Portuguese, Italian, Romanian, Serbo-Croatian, and Albanian, have the largest number of path verbs. These languages often have more than one verb for a single English verb, see for instance the six French verbs for RETURN. Some kinds of paths seem to be encoded by a verb in almost all or all 20 languages, such as movement away from a location (LEAVE), movement arriving at a location (ARRIVE), unsupported movement downwards (FALL), and movement back to a location previously abandoned (RETURN). Other kinds of paths are encoded by some but not by others. Some kinds of paths are only encoded by a verb in verb-framed languages, which includes movement into an enclosure (ENTER), movement out of an enclosure (EXIT) and general movement downwards (DESCEND). This division into meanings encoded by all languages and meanings encoded just by verb-framed languages seems very similar to the two-tiered lexicon for manner verbs introduced by Slobin (1997) that was discussed in section 5.2.3. For path of motion verbs, LEAVE, ARRIVE, FALL and RETURN form the first tier, while ENTER, EXIT and DESCEND, and others form the second tier.

5.2.4 Discussion and summary

This chapter investigates the relationship between the use of specific motion event encoding constructions and the number of manner verbs and path verbs attested in the two sentence samples. Since the corpus is restricted in size, it may be useful to employ analyses that give an estimate of the total number of manner verbs and path verbs that the languages might have, or that provide some weighting to the verb type-token ratio so that the frequency of individual verbs is taken into account. In Verkerk (2013), one of the papers on which the current chapter is based, a measure called 'the Chao index' is used in addition to the counts of unique manner verbs reported on in this chapter. The Chao index gives an estimate of the total number of manner verbs based on the number of verbs that appear only once or twice, and serves to provide an estimate or extrapolation of true manner verb lexicon size. However, the Chao index seemed only to be of limited use, as for some languages, estimates were made that were quite divergent from the simple count scores. In addition, as this method is developed for accurately counting species in biological ecology, it is unclear how well it can be applied to estimating true verb class sizes in linguistics. It would be very useful, however, if methods that may extrapolate from limited samples of counts such as these to more apprehensive characterizations of (verb) class size could be developed in linguistics.

Tables 5.5 and 5.6 present a short overview of the measures used for the phylogenetic comparative analyses discussed in section 5.3 and 5.4. The abbreviated names with which each measure is designated are given between brackets.

***Table 5.5**: Samples and measures used for manner verb lexicon size correlations*

Sample	Motion event encoding construction measure	Manner verb lexicon size measure
118/132-sentence	PC1 score (PC1)	unique manner verb count (MV count)
118/132-sentence	proportion satellite-framed (% SAT)	unique manner verb count (MV count)
192/215-sentence	PC1 score (PC1)	unique manner verb count (MV count)
192/215-sentence	proportion satellite-framed (% SAT)	unique manner verb count (MV count)

***Table 5.6**: Samples and measures used for path verb lexicon size correlations*

Sample	Motion event encoding construction measure	Path verb lexicon size measure
118/132-sentence	PC1 score (PC1)	unique path verb count (PV count)
118/132-sentence	proportion path-only + verb-framed (% PO+VF)	unique path verb count (PV count)
192/215-sentence	PC1 score (PC1)	unique path verb count (PV count)
192/215-sentence	proportion path-only + verb-framed (% PO+VF)	unique path verb count (PV count)

5.3 Methodology

This chapter investigates the correlation between the motion event encoding system and the size of the manner verb lexicon and the path verb lexicon. Specifically it tests 1) whether languages that use the satellite-framed motion event encoding constructions more commonly or have a higher PC1 score also have a larger manner verb lexicon and 2) whether languages that use the path-only and verb-framed motion event encoding constructions more commonly or have a lower PC1 score also have a larger path verb lexicon. These hypotheses have to be investigated using phylogenetic comparative methods, because these methods take into account the genealogical relationships between the languages in the sample. In order to conduct these analyses, a measure of genealogical relations in the form of a set of phylogenetic trees is needed. The phylogenetic tree sets that are used in this chapter are those taken from Bouckaert et al. (2012) that were introduced in section 1.3.2.2. The maximum clade credibility trees that summarize these two tree samples have been repeated below as Figure 5.5 and Figure 5.6 for convenience.

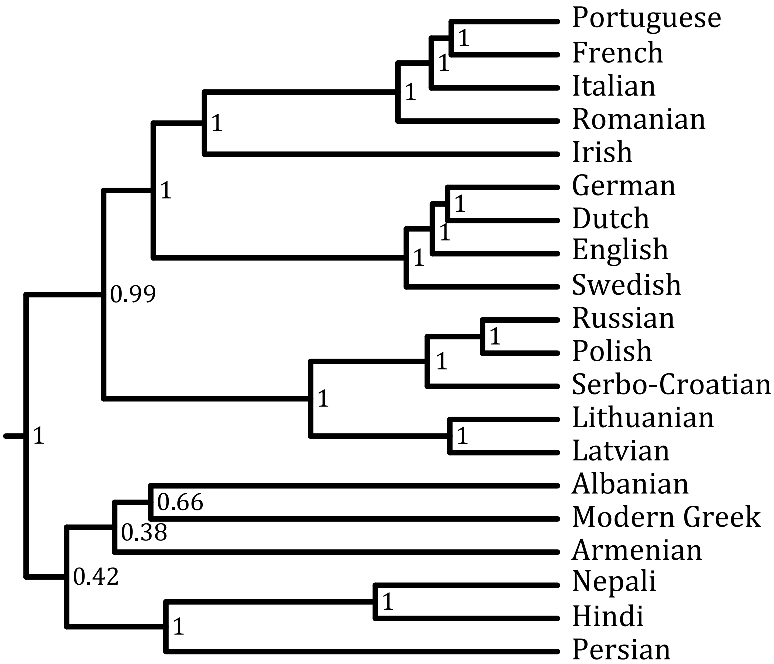


Figure 5.5*: The maximum clade credibility tree of 1000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 20 languages featured in this dissertation.*

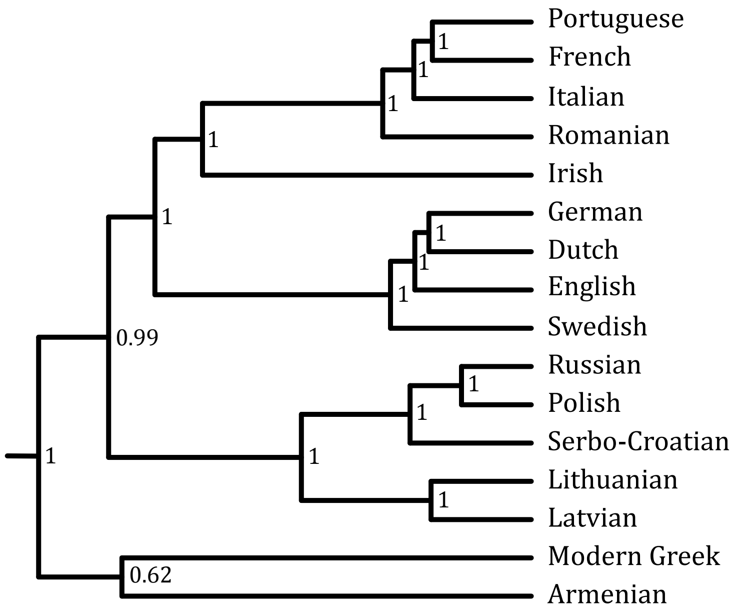


Figure 5.6*: The maximum clade credibility tree of 1000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 16 languages featured in some analyses presented in this chapter.*

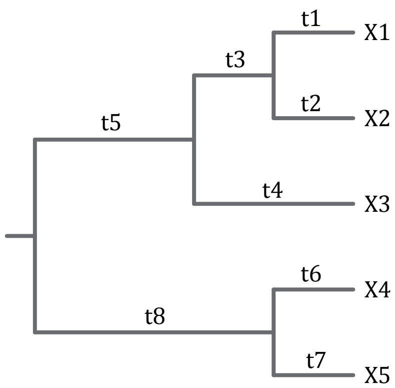
The first phylogenetic comparative test that was employed, was a test for phylogenetic signal. For a dataset that includes several (closely and less closely) related languages from a single language family, it is necessary to test whether phylogenetic signal is present. If phylogenetic signal is present within a linguistic feature, this means that languages that are closely related behave similarly with regard to that feature, because they have inherited this feature from a common ancestor. When this is the case, it becomes necessary to use phylogenetic comparative methods to analyze the feature, because these methods can take into account the genealogical relationships between the languages.

In the current study, the estimation of the branch scaling parameter λ was used to test whether phylogenetic signal was present (Pagel 1999a; Freckleton et al. 2002). This test has been introduced in section 1.3.3.2. The estimation of λ determines the presence or absence of phylogenetic signal by optimizing the length of the shared branches of the phylogenetic tree. A high λ value (1 or similar) indicates the presence of phylogenetic signal, and a low value (0 or similar) indicates the absence of phylogenetic signal. The estimation of λ and the likelihood ratio tests to determine its statistical significance were conducted using the function physig, part of the R (R Development Core Team, 2011) package phytools (Revell 2012) and the function fitContinuous, part of the R package GEIGER (Harmon et al. 2008). The results of the tests for phylogenetic signal will be discussed in section 5.4.1.

The second set of phylogenetic comparative tests that was employed were Phylogenetic Generalized Least Squares (PGLS) correlation analyses (Pagel 1997; Freckleton et al. 2002; Rohlf 2006). Although this test has been introduced and explained to some extent in section 1.3.3.4, it will be recapitulated here for convenience. PGLS correlation analyses were carried out in order to actually test whether languages that use the satellite-framed construction more often or have a higher PC1 score have larger manner verb lexicons and whether languages that use the path-only and verb-framed constructions more often or have a lower PC1 score have larger path verb lexicons. A PGLS correlation is a multivariate analysis of correlations among two or more dependent variables that takes into account phylogenetic information provided by a sample of phylogenetic trees (Pagel 1997: 337ff; Rohlf 2006: 1509-1510). The correlation model is given in (51) (as repeated from (11) in section 1.3.3.4):

51) $Z = 1_n\mu + e$

In a regular correlation analysis, Z is an $n \times p$ matrix of n observations of p dependent variables, 1_n is an unit vector of length n, μ is an $1 \times p$ vector of means, and e is an $n \times p$ matrix of multivariate normally distributed errors. If the measures on motion event encoding and manner verb lexicon size were independent, the correlations between the dependent variables in Z could be estimated using conventional techniques. However, the language data is not independent due to the shared history of the languages. The values in the matrix Z are interdependent, with closely related languages having more similar values as compared to less closely related languages because they have inherited features from a shared ancestor. PGLS correlation analysis takes this shared history into account by using the shared branch lengths of the phylogenetic tree. See Figure 5.7 (repeated from Figure 1.8) for illustration.



	X1	X2	X3	X4	X5
X1	t5+t3+t1	t5+t3	t5	-	-
X2	t5+t3	t5+t3+t2	t5	-	-
X3	t5	t5	t5+t4	-	-
X4	-	-	-	t8+t6	t8
X5	-	-	-	t8	t8+t7

Figure 5.7*: An example of a tree with a corresponding variance-covariance matrix*

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. The shared history of each language with each other language is formally assessed by the length of the branches that are shared. In Figure 5.7, X1 and X2 are closely related, which can be stated formally in terms of the shared branch length between X2 and X3, t5+t3. In contrast, X1 and X3 are less closely related, which can be stated formally in terms of the shared branch length between X2 and X3, t5. A PGLS correlation analysis first constructs a variance-covariance matrix that embodies all this information on unique and shared branch length (see Figure 5.7). It then uses this variance-covariance matrix to adjust the error terms e in the formula in (51). Using these phylogenetically adjusted error terms, the PGLS correlation can then describe the relationship between the motion event encoding system and the size of the manner verb lexicon and path verb lexicon while taking into account the shared history between the languages.

In section 1.3.3.4, two types of statistical analyses for the investigation of co-evolution were introduced: PGLS regression and PGLS correlation. Correlation analyses of the type described in (11) in section 1.3.3.4 and (51) investigate how much two or more variables change together and how strong the relationship is between them. Regression analysis as discussed in (10) in section 1.3.3.4 rather focuses on the relationship between a dependent variable and one or more independent variables. Correlation analysis makes no *a priori* assumption as to whether one variable is dependent on any of the other(s) and is not concerned with the direction of the relationship between variables, instead it gives an estimate as to the strength of the association between the dependent variables. In this chapter, the results from a set of PGLS correlation analyses is described. There are two reasons for focusing on correlation analyses. The first is that the hypotheses under investigation, i.e. whether having a certain kind of motion encoding system is related to having a certain kind of motion lexicon, are suited better for investigation with PGLS correlation than with PGLS regression. Rather than focusing on a specific directional relationship between the variables, it is most relevant to test the presence and strength of a potential association between these variables. The second reason is that PGLS regressions have been conducted for the same manner verb dataset (Verkerk 2013), which turned out to be problematic due to very low phylogenetic signal in the error term of the PGLS regression. This seemed to be a quirk in the manner verb dataset, as the data on path verbs (as reported on in Verkerk 2014b) do not have the same problem. In order to present a unified set of analyses in this chapter, the choice was made to conduct PGLS correlation analyses.

PGLS correlation analysis can describe the relationship between the motion event encoding system and the size of the motion verb lexicon while taking into account shared history as operationalized by the variance-covariance matrix. If the variance-covariance matrix taken from a phylogenetic tree is taken as is, the feature is modeled to be 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 to the data. Therefore, the parameter λ, which measures the degree to which the phylogeny predicts the covariance between the languages, is estimated in parallel by the PGLS correlation analysis. The most likely value of λ is estimated to modify the variance-covariance matrix of the tree in such a way that it best reflects the amount of phylogenetic dependence between the variables (see section 1.3.3.2). 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 correlation analysis, without correction for phylogeny). If λ is between 1 and 0, the tree topology and branch lengths are modified accordingly as illustrated in Figure 1.9 in section 1.3.3.2.

Using the scaled branch lengths as a measure of genealogical relatedness, PGLS correlations can provide an answer to the question of whether two features are correlated. PGLS correlations analyses were conducted using the phylogenetic comparative method *Continuous* (Random Walk Model A) implemented in the BayesTraits package (Pagel 1997) that can be found at http://www.evolution.rdg.ac.uk/BayesTraits. The PGLS correlation analysis was conducted on the set of 1000 phylogenetic trees in order to take into account the uncertainty present in the phylogenetic estimation. *Continuous* was used to conduct maximum likelihood PGLS correlations for all relevant combinations of motion event encoding measures and motion verb lexicon size measures.

5.4 Results

5.4.1 Non-phylogenetically corrected correlations

The measures used for the motion event encoding system as well as the size of the manner verb lexicon and the path verb lexicon are highly correlated when non-phylogenetic correlation analyses (Pearson correlation coefficient) are carried out. This is true for both measures across samples (Table 5.7) and motion event encoding and motion verb lexicon size measures (Table 5.8). Phylogenetically corrected correlations follow in section 5.4.3.

Table 5.7*: Non-phylogenetically corrected correlation analyses for the motion event encoding and motion verb lexicon size measures across samples*

Measure 1	**Measure 2**	**Pearson**
215-sentence MV count	132-sentence MV count	0.88
192-sentence % SAT	118-sentence % SAT	0.98
215-sentence PV count	132-sentence PV count	0.95
192-sentence % PO+VF	118-sentence % PO+VF	0.99
192-sentence PC1	118-sentence PC1	0.98

Table 5.8*: Non-phylogenetically corrected correlation analyses for the motion event encoding and motion verb lexicon size measures within samples*

Measure 1	**Measure 2**	**Pearson**
192-sentence PC1	215-sentence MV count	0.80
118-sentence PC1	132-sentence MV count	0.85
192-sentence % sat	215-sentence MV count	0.81
118-sentence % sat	132-sentence MV count	0.88
192-sentence PC1	215-sentence PV count	-0.67
118-sentence PC1	132-sentence PV count	-0.58
192-sentence % PO+VF	215-sentence PV count	0.60
118-sentence % PO+VF	132-sentence PV count	0.65

5.4.2 Phylogenetic signal

The results of the tests for the presence of phylogenetic signal are presented in Table 5.9. The median λ for each individual measure is given in the third column, while the range found within the 1000 tree sample is given in the fourth column. As explained briefly in section 1.3.3.2, the λ that is estimated can be slightly higher than 1 due to characteristics of the data and the phylogenetic tree. For the current two tree samples, the maximum λ is 1.28. Table 5.9 also includes the results of the likelihood ratio tests. The fifth and sixth column give the number of trees for which λ was estimated to be significantly different from λ = 1 and λ = 0 respectively on the $p < 0.05$ significance level.[16]

Table 5.9: *Tests for phylogenetic signal for the motion event encoding and motion verb lexicon size measures*

Sample	Data measure	Median λ	Range λ	λ 1[a]	λ 0[b]
192-sentence	PC1	1.16	1.09–1.28	999	1000
118-sentence	PC1	1.16	0.95–1.28	914	1000
192-sentence	% SAT	1.16	1.08–1.27	180	1000
118-sentence	% SAT	1.15	0.97–1.27	4	1000
192-sentence	% PO+VF	1.01	0.83–1.22	0	1000
118-sentence	% PO+VF	1.01	0.87–1.14	0	1000
215-sentence	MV count	1.02	0.95–1.11	0	1000
132-sentence	MV count	0.84	0.77–0.93	9	1000
215-sentence	PV count	0.95	0.83–1.08	0	1000
132-sentence	PV count	0.93	0.82–1.03	0	1000

[a]λ 1 lists the number of trees for which λ was estimated to be significantly different from 1.
[b]λ 0 lists the number of trees for which λ was estimated to be significantly different from 0.

The estimated λ values for the PC1 scores, the proportion of use of the satellite-framed construction, the proportion of use of the path-only and verb-framed constructions, the manner verb count, and the path verb count are all quite high (medians 0.84–1.16). This was the case for all sentence samples. For these ten measures, the estimated λ was significantly different from 0 for all phylogenetic

[16] 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 λ =1. The maximum possible λ values ranged from 1.090 - 1.279 for the tree sample used for the analysis with the 192-sentence sample and the 215-sentence sample, and 1.090 - 1.277 for the tree sample used for the analysis with the 118-sentence sample and the 132-sentence sample. However, these values could not be used as maximum possible λ values, as they are not accepted by the corPagel function from the R package ape (Paradis et al. 2004). This function only accepts fixed values for λ between 0 and 1.

trees. They were also mostly not significantly different from a model in which λ was set to 1 (except for the PC1 score in both samples and the % SAT for the 192-sentence sample, see footnote 16). Note that the analyses using the samples including data from 16 languages (the 192-sentence sample and the 215-sentence sample) are on the limit of a having a sample size that is too small to properly detect phylogenetic signal. Freckleton et al. (2002) demonstrate through simulations that for trees with 10 tips, in 75% of their analyses, the absence of phylogenetic signal was correctly identified, while the presence of phylogenetic signal was correctly identified in about 48% of their analyses. These numbers rise to about 97% and 85%, respectively, for trees with 20 tips. However, since the analyses using data on 16 languages are very much in line with the results for the analyses that use data on all 20 languages, the analyses using data from 16 languages seem to have correctly identified the presence of phylogenetic signal. The results of the analyses indicate that there exists a clear phylogenetic signal, and the behavior of the languages with regard to these data measures is dependent on phylogenetic history.

5.4.3 Phylogenetic Generalized Least Squares correlations

Second, the results from the PGLS correlation analyses are discussed. To investigate the hypothesis that motion event encoding and the motion verb lexicon size are correlated, two models were tested: the null model that states that these two variables are not correlated but rather evolved independently (M_i) and the alternative model that states that they have evolved together and are correlated (M_c). For each dataset, BayesTraits was run twice, once while estimating the strength of the association (M_c) and one while not allowing for a correlation (M_i). The alternative model of correlated evolution can be said to be an improvement on the null model of independent evolution if its likelihood is significantly larger. Statistical significance can be assessed by comparing the mean likelihoods of the two models using the likelihood ratio test (Pagel 1997: 342).

The results of the PGLS correlation analyses for manner verb lexicon size are presented in Table 5.10. Table 5.10 indicates that the PGLS correlation analysis produced λ values that ranged between 0.69 and 0.99, depending on the measures and the sample. This suggests that shared history explained a large 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 correlation analysis that adjusts for phylogenetic relatedness is validated. In all analyses, the alternative model has a higher likelihood than the null model, which is evidence that there is a correlation between motion event encoding and manner verb lexicon size. In the two analyses with the 118/132-sentence sample (that includes data from all 20 languages) this correlation is statistically significant

(PC1 motion event encoding measure: p = 0.005; %SAT motion event encoding measure: p = 0.001). The unsignificant results for the two analyses with the 192/215-sentence sample are likely due to small sample size – as we will see below, the correlations for the 118/132-sentence sample have lower p-values than the 192/215-sentence sample for path verbs as well.

***Table 5.10**: PGLS correlation results for the 192/215-sentence and the 118/132-sentence samples for manner verb lexicon size*

Sample	**Measure**	**Model**	**ln Lh**[a]	**λ**	**R**	**LR**[b]	**p**[b]
192/215-sentence	PC1	M_c	31.02	0.99	0.33	0.93	0.11
		M_i	30.08	0.99	0		
192/215-sentence	%SAT	M_c	33.91	0.99	0.39	1.27	0.07
		M_i	32.64	0.99	0		
118/132-sentence	PC1	M_c	43.42	0.71	0.60	3.39	0.005
		M_i	40.04	0.93	0		
118/132-sentence	%SAT	M_c	47.06	0.69	0.67	4.53	0.001
		M_i	42.53	0.94	0		

[a]The ln Lh is the mean of the ln likelihood over all 1,000 phylogenetic trees in the sample for each analysis.
[b]The LR is the Likelihood Ratio 2(ln Lh M_c – ln Lh M_i). This test statistic is assumed to approach a chi-square distribution with the degrees of freedom equal to the difference in the number of parameters (this analysis has 1 degree of freedom). The p-value indicates whether the LR is statistically significant on the chi-square distribution.

This means that this dataset provides some evidence for the hypothesis that when a language has a higher PC1 score as depicted in Figure 5.1 and 5.2 or when it uses the satellite-framed construction more often, it will also have a more sizable manner verb lexicon. Because a PGLS correlation analysis was used, we can be sure that this correlation was not caused by similar behavior in closely related languages due to shared history.

The results of the PGLS correlation analyses for path verb lexicon size are presented in Table 5.11. Table 5.11 indicates that the PGLS correlation analyses produced high λ values that ranged between 0.91 and 0.99, depending on the measures and the sample. This suggests that shared history explained a large part of the covariance between the measures of motion event encoding and the measures of size of the path verb lexicon, and the use of a correlation analysis that adjusts for phylogenetic relatedness is validated. In all analyses, the alternative model has a significantly higher likelihood than the null model, which is evidence that there is a correlation between motion event encoding and path verb lexicon size (see last column of table 5.11). The p-values for the analyses with the 118/132-sentence sample are lower than those for the analysis with the

192/215-sentence sample, suggesting that a larger language sample yields more significant correlations.

Table 5.11: *PGLS correlation results for the 192/215-sentence and the 118/132-sentence samples for path verb lexicon size*

Sample	**Measure**	**Model**	**ln Lh**[a]	**λ**	**R**	**LR**[b]	**p**[b]
192/215-sentence	PC1	M_c	29.03	0.99	-0.55	2.80	0.01
		M_i	26.22	0.99	0		
192/215-sentence	%VERB	M_c	36.77	0.93	0.49	2.21	0.02
		M_i	34.55	0.91	0		
118/132-sentence	PC1	M_c	35.54	0.99	-0.67	5.95	0.0003
		M_i	29.59	0.99	0		
118/132-sentence	%VERB	M_c	48.56	0.99	0.68	6.05	0.0003
		M_i	42.51	0.95	0		

[a]The ln Lh is the mean of the ln likelihood over all 1,000 phylogenetic trees in the sample for each analysis.
[b]The LR is the Likelihood Ratio 2(ln Lh M_c – ln Lh M_i). This test statistic is assumed to approach a chi-square distribution with the degrees of freedom equal to the difference in the number of parameters (this analysis has 1 degree of freedom). The p-value indicates whether the LR is statistically significant on the chi-square distribution.

These results indicate that there exists a negative relationship between PC1 score and path verb lexicon size: a higher score on the PC1 (i.e. a more satellite-framed nature) implies a smaller path verb lexicon (see the negative R's in Table 5.11). There exists a positive relationship between the proportion of use of the path-only and verb-framed constructions and path verb lexicon size: a higher usage rate of the path-only and verb-framed constructions implies a larger path verb lexicon. These correlations is not due to shared history between the languages in the sample, because PGLS correlation analyses were used to remove covariance due to shared descent.

5.5 Discussion and conclusion

The results from the PGLS correlation analyses indicate that there exists a relationship between the motion encoding system, measured in various ways, and the size of the manner verb lexicon and path verb lexicon. These correlations hold, even though the genealogical relationships between the languages in the dataset explain part of the covariance. They support the findings reported in the literature on motion event encoding as discussed in section 5.1, which also generally seem to suggest that satellite-framed languages typically have larger manner verb lexicons, while verb-framed languages typically have larger path verb lexicons.

This study is the first language family-wide positive confirmation of the earlier suggestions (Wienold 1995: 322; Slobin 2004: 252) and later demonstration for English manner verbs (Fanego 2012) that there exists a relationship between motion event encoding patterns and the size of the manner verb lexicon and the path verb lexicon. One of the most interesting aspects of this relationship are the evolutionary mechanisms on a population-wide scale that are involved in the emergence and continuation of the salience of manner and path.

Given the correlations presented in this chapter, perhaps the most straightforward hypothesis on the nature of the relationship between specific constructions and the parts of the lexicon that feature prominently in these constructions is as follows: Once a language starts to use a particular motion event encoding construction more often, more verbs that can be placed in these constructions arise through various processes, such as semantic shift or borrowing. However, the results presented in section 5.4 are simply evidence for the existence of a correlation between the motion event encoding system and the lexicon while taking into account the phylogenetic dependencies among the languages. To test directional hypotheses using phylogenetic comparative methods, a much larger dataset including more languages would be needed. Lindenfors et al. (2004), for instance, test population size in primates and find that female group size changes first and male group size afterwards. In other words, the correlation between female group size and male group size can be explained by changes in females' sociality. Given a sample with more languages, similar analyses can be carried out to investigate the correlations between motion event encoding and manner verb and path verb lexicon size.

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, then this can be related to the evidence presented in chapter 4 that over time, Germanic and Balto-Slavic languages have used the satellite-framed strategy more and more often. Is this emphasis on manner in these languages 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.

With regard to path-salience, a more definite claim can be made. As was explained in chapter 4 and section 5.2.3, many Indo-European languages show remnants of a once productive system of path prefixes, most importantly

Romance and Indo-Aryan (Vincent 1999; Kopecka 2006; Bloch 1965). While the system of path prefixes became unproductive, many prefix-verb combinations became monomorphemic path verbs. It is highly likely that the emergence of these new path verbs caused the motion event encoding constructions that employ path verbs, namely the path-only and verb-framed constructions, to become more common. A feedback cycle could then have originated: since path-only and verb-framed constructions became more common, more verbs that could feature in these constructions emerged, etc.

The relationship between syntactic constructions and the lexical classes that appear in these constructions definitely seems to be one of the mechanisms that is involved in the expansion or contraction of lexical classes, but it is certainly not the only one. A language that stands out in this dataset is Irish, which has a rather small manner verb lexicon (14 verbs in the 118/132-sentence sample, 21 in the 192/215-sentence sample) for a language that uses the satellite-framed construction more often than any of the verb-framed languages. Irish also has a small path verb lexicon (only 11 path verbs in the 118/132-sentence sample, 14 in the 192/215-sentence sample). Irish has a rather small verb lexicon, not only in the motion domain, but throughout the lexicon, which is characterized by high degrees of polysemy. Hindi, Nepali, and Persian also do not have large classes of manner verbs nor large classes of path verbs. Clearly, there are other factors at play that have shaped these lexicons, 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. However, this is not true for all satellite-framed languages, as is illustrated by the borrowing of path verbs of Romance origins into 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 or path 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, including correlations between motion event encoding and other motion lexicons, such as manner ideophones (Wienold 1995), manner adverbs, adpositions and case systems. It is unknown how much variation language families typically exhibit with regard to motion event encoding and the size of the manner verb and path verb lexicons. The Indo-European language family might be quite unusual in having languages that are as dissimilar as Latvian and Nepali; however, it might also not be. Wälchli's (2009: 215) results suggest that the genealogical stability of motion event encoding is overrated, indicating that we expect a fair bit of diversity in language families and also some in subgroups. For large language families, big datasets may be constructed that would also allow for the investigation of directional diachronic hypotheses, as has been done for biological features by Deaner and Nunn (1999) and Lindenfors et al. (2010). A further step would be to investigate implicational hierarchies of manner and concepts, i.e. whether there exists a universal order in which manner and path concepts are added to manner verb, ideophone, adverb, and adposition lexicons.

An alternative to gathering data on a sample of (closely) related languages and analyzing them using phylogenetic comparative methods to answer questions about correlations between features would be to collect data from a number of unrelated languages and analyze them with conventional correlation or regression methods (see also section 1.3.4). However, there are a number of disadvantages to such an approach. First of all, although it would be computationally less intensive to calculate such correlations, the results would be far less informative as compared to the results presented in the current chapter. The reason for this is that although one might learn that the two features are correlated, one does not know through what mechanism this correlation has emerged. By studying closely related languages it is possible to observe enough variance for the inference of historical processes. The historical processes that have given rise to the differences between satellite-framed and verb-framed languages feature in chapter 4 and 6. Secondly, using only languages from different language families would only push back the potentially confounding genealogical and contact relationships further in time, as all languages are related on some level (but see Freedman & Wang 1996 and others for a different opinion). Incorporating what is known about genealogical relatedness is a more valid approach than ignoring it as a potential but unaccounted for source of error.

Understanding the dynamics of the relationship between the sentence structures we use and the words we put in our sentences will continue to be an important aspect in motion event encoding research. 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 indicates that it has great potential in uncovering cross-linguistic patterns that will allow for a better understanding of the structure of meaning in human language. This chapter has made a contribution to this topic by showing that there is a relation between constructions and motion verbs and that it is possible that the dynamics between them influence both syntax and the lexicon.

Chapter 6: The evolution of motion verbs

This chapter is a slightly revised version of:
Verkerk, A. (submitted). Where do all the motion verbs come from? The speed of development of manner verbs and path verbs in Indo-European. Diachronica*.*

The study of the linguistic encoding of motion has been intensively examined in the last three decades (Talmy 1985, 1991; Slobin 1996b). From the start, this domain of study includes studies of the lexicon (Slobin 1997, 2004, 2005b). Comparisons between satellite-framed and verb-framed languages suggest that satellite-framed languages typically have a larger manner of motion verb lexicon (*walk*, *swim*, *dash*), while verb-framed languages typically have a larger path of motion verb lexicon (*descend*, *enter*, *cross*) (chapter 5; Slobin 2004). This chapter will investigate through what evolutionary processes these differences between the motion verb lexicons of satellite-framed and verb-framed languages arise. To investigate these questions, phylogenetic comparative methods adopted from biology and an etymological study as is commonly conducted in philology are used to investigate manner verbs and path verbs in the Indo-European dataset. It is shown that manner verbs and path verbs typically have different types of etymological origins. It is also demonstrated that manner verbs evolve faster in satellite-framed subgroups, while path verbs evolve faster in verb-framed subgroups.

6.1 Introduction

Motion event encoding has been a popular area of investigation in cognitive linguistics and linguistic typology over the last three decades. Increased interest in this domain started with the work of Leonard Talmy (1985, 1991), who framed the opposition between verb-framed and satellite-framed languages: verb-framed languages typically use the verb-framed construction, in which information on the path of motion is encoded on the verb, while satellite-framed languages typically use the satellite-framed construction, in which information on path is encoded outside the verb on a so-called satellite. Examples are provided in the parallel translations (52) and (53):

52) Italian (verb-framed)

Alice	*si*	*inoltr-ò*	*corre-ndo*	*nel*
Alice	REFL	penetrate-PST.3SG	run-PRS.PTCP	in.DEF.ART.M.SG

bosco	*e*	*si*	*ferm-ò*	*sotto*	*un*	*albero*	*frondoso*
wood.M.SG	and	REFL	stop-PST.3SG	under	one.M	tree.M.SG	leafy.M.SG

'Alice entered the wood running and stopped under a leafy tree'

53) Swedish (satellite-framed)

Alice	*sprang*	*in*	*i*	*skog-en*	*och*	*ställde*
Alice	run.PST	into	in	forest-SG.DEF.UT	and	stand.PST

sig	*under*	*ett*	*stort*	*träd*
REFL.3SG.N	under	INDF.ART.N	large	tree.SG

'Alice ran into the forest and placed herself under a large tree'

In the Italian example in (52), the path of the motion (from the outside of the forest to the inside of the forest) is encoded on the verb *inoltrarsi* 'penetrate', while the manner of the motion (running) is encoded on the participle *correndo* 'running'. In the Swedish example in (53), the manner of motion is encoded on the verb *springa* 'run', while the path of motion is encoded on the adverb *in* 'into' and the preposition *i* 'in' (see Sinha & Kuteva 1995: 190-191 on this type of double marking in Danish).

During the last three decades, these and other syntactic constructions used to encode motion have been described for many languages spoken around the world. Of special interest in this chapter is the relationship between these syntactic constructions and the motion verb lexicon. In recent work, a relationship between the motion event encoding construction that is commonly used and the size of the lexicon of certain types of motion verbs has been proposed. Languages that commonly make use of the satellite-framed construction, such as Swedish, are claimed to have a larger lexicon of manner verbs such as *walk*, *fly*, and *limp* (Slobin 2004). On the other hand, languages that commonly make use of the verb-framed construction, such as Italian, are claimed to have a larger lexicon of path verbs such as *ascend*, *exit*, and *cross* (Cifuentes Férez 2010). Chapter 5 has provided some evidence that both these claims are true for a sample of 20 Indo-European languages.

If such correlations truly exist, that means that the history of the motion verb lexicons of the Indo-European languages should reflect these differences: manner verbs should emerge faster in branches of the Indo-European tree that lead to satellite-framed languages in order to generate the larger manner verb classes that are found in contemporary satellite-framed Indo-European languages. Likewise, path verbs should emerge faster in branches of the Indo-

European tree that lead to verb-framed languages, as there otherwise would be no explanation why verb-framed languages have larger path verb lexicons. This chapter aims to investigate where Indo-European manner verbs and path verbs are coming from etymologically and whether there is evidence for faster evolution of these verb classes in various branches of the Indo-European tree.

The Indo-European languages are an interesting case study because they are quite diverse with respect to the motion event encoding constructions that are typically used (chapter 2 and 3; Slobin 2004, 2005b). The Germanic and Balto-Slavic languages are mainly satellite-framed, while the Romance languages are mainly verb-framed. Albanian and Modern Greek are verb-framed as well. The Indo-Iranian and Celtic languages in the sample, as well as Armenian, can be considered to be verb-framed with respect to their frequency of use of the satellite-framed construction, but frequently use deictic verbs instead of path verbs. Given this diversity, the following questions can be investigated:

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

For the current dataset, the hypothesis is that the answer to the two questions above is ‘yes’. This hypothesis will be investigated by studying the etymological origins of the manner verbs and path verbs from a sample of 20 different Indo-European languages. Phylogenetic comparative methods are also used to estimate the rate of change or rate of evolution (see for an example from biology Figure 6.1). The rate of change, which refers to the speed with which a feature such as lexicon size is evolving, is an important variable if we want to model the evolution of linguistic features. During the history of the Indo-European languages, manner verb lexicons and path verb lexicons might have been expanding or shrinking at faster or slower rates given their sizes in different languages. By studying differential rates of change in different parts of the tree, we might find further support of the hypothesis that the growth and decline of the manner verb and the path verb class is correlated with syntactic patterns of motion event encoding.

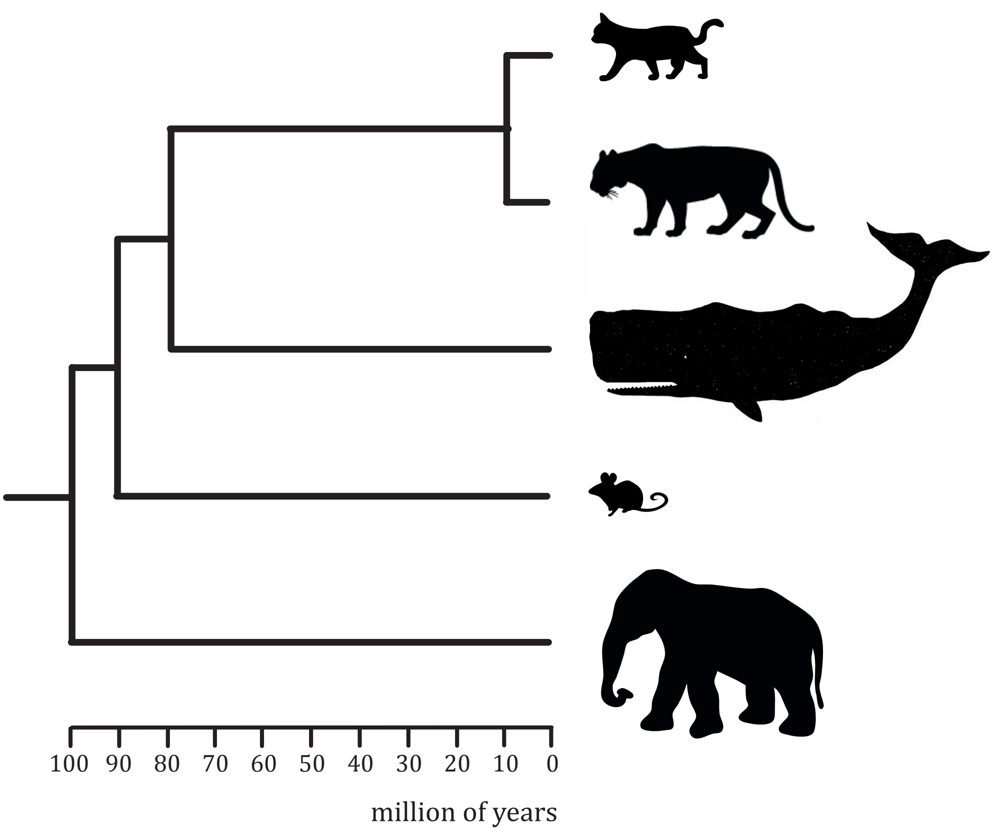


Figure 6.1*: Body size evolution in mammals. Body size is known to be highly diverse in the class of mammals. Studies suggest that change in body size during some periods has accelerated and during other periods has decelerated, depending on a host of variables (Venditti et al. 2011).*

The current study provides a contribution to lexical typology by studying diachronic change in two semantic subfields. Aside from many studies on isolated processes of lexical change such as borrowing, compounding, affixation, and semantic shift, there are remarkably few studies of lexical change that take into account all word formation processes in a single language, or that take into account word formation in a specific semantic domain. Studies that focus on general word formation are Algeo (1980), who studied the origins of a randomly chosen 1000 English words, and Cannon (1978), who studied the origins of Merriam-Webster's 6000 Words, a collection of new words in American English published in 1976. Studies that look at word formation in specific semantic domains include Witkowski et al. (1981), who study words for 'tree', and Brown (1983), who studies the etymological origins of cardinal direction terms. In this chapter, the results of a cross-linguistic study into the evolutionary processes that act on the manner verb lexicon and the path verb lexicon will be presented.

6.2 Manner verbs and path verbs

In this chapter, the manner verbs and path verbs are taken from the 132-sentences sample introduced in section 2.1.2. The coding of path verbs and manner verbs is explained in section 5.2.2 and section 5.2.3, respectively. As a brief recap, the procedure for classifying each kind of verb was as follows. If a manner verb encoded a transparent manner of motion and could be used with different paths, such as English *run*, *roll*, or *crawl*, it was coded as a manner verb. If a path verb encoded a transparent path of motion and could be used in different manner of motion contexts, such as English *descend*, *enter*, and *pass*, it was coded as a path verb. In this way, verbs that encode both path and manner at the same time, such as Modern Greek *skarfalono* 'to climb up' and Persian *gorik̲tan* 'to run away', are kept in a separate category, that of manner plus path verbs (see chapter 2). Complex manner verbs, such as *give a jump* and *make a rush*, and non-motion verbs used as manner verbs such as *wiggle* and *crash*, were not included as manner verbs in this study (see section 5.2.2). For path verbs consisting of a lexicalized combination of a path prefix and a verb, strict criteria regarding the semantic transparency of the verb complex were maintained (see section 5.2.3). Lists of all the unique types of manner verbs and path verbs included in this study are presented in Appendix 5 and 9. Figure 6.2 and Table 6.1 summarize these appendices by providing the size of the manner verb lexicon and the path verb lexicon for each language of the sample.

From Figure 6.2 and Table 6.1, it is clear that languages may vary quite widely in their manner verb and path verb lexicons. Some languages have a large manner verb lexicon and a small path verb lexicon, such as Swedish, other languages have a small manner verb lexicon and a large path verb lexicon, such as Portuguese, and still other languages have a small manner verb lexicon and a small path verb lexicon, such as Irish.

As originally proposed by Slobin (1997), the manner verb lexicon can be divided into two parts: a first tier of more general or neutral verbs and a second tier of more specific and expressive verbs. Satellite-framed languages have extensive second tier manner verb lexicons, while the first tier manner verbs should be more or less the same set in both non-satellite-framed and satellite-framed languages. This distinction seems to be a valid qualitative description of the current dataset. In section 5.2.2, it was noted that there exists a clear set of manner of motion concepts that are lexicalized into verbs in each of the languages in the current sample. This set includes 'walk', 'stroll', 'run', 'fly', 'swim', 'roll', 'jump', 'rush, hurry', and 'wander'. A similar division into first tier and second tier path verbs in the current dataset was introduced in section 5.2.3, i.e. certain kinds of paths seem to be encoded by a verb in almost all or all 20 languages included in the current chapter. These included 'leave', 'arrive', 'fall', and 'return'. Other kinds of paths are lexicalized less often. Given the hypotheses on the evolution of manner verbs and path verbs, a feasible hypothesis is to expect that first tier manner verbs and path verbs emerge early and are stable as they are the most frequently used motion verbs, while specific or expressive manner verbs and path verbs emerge later.

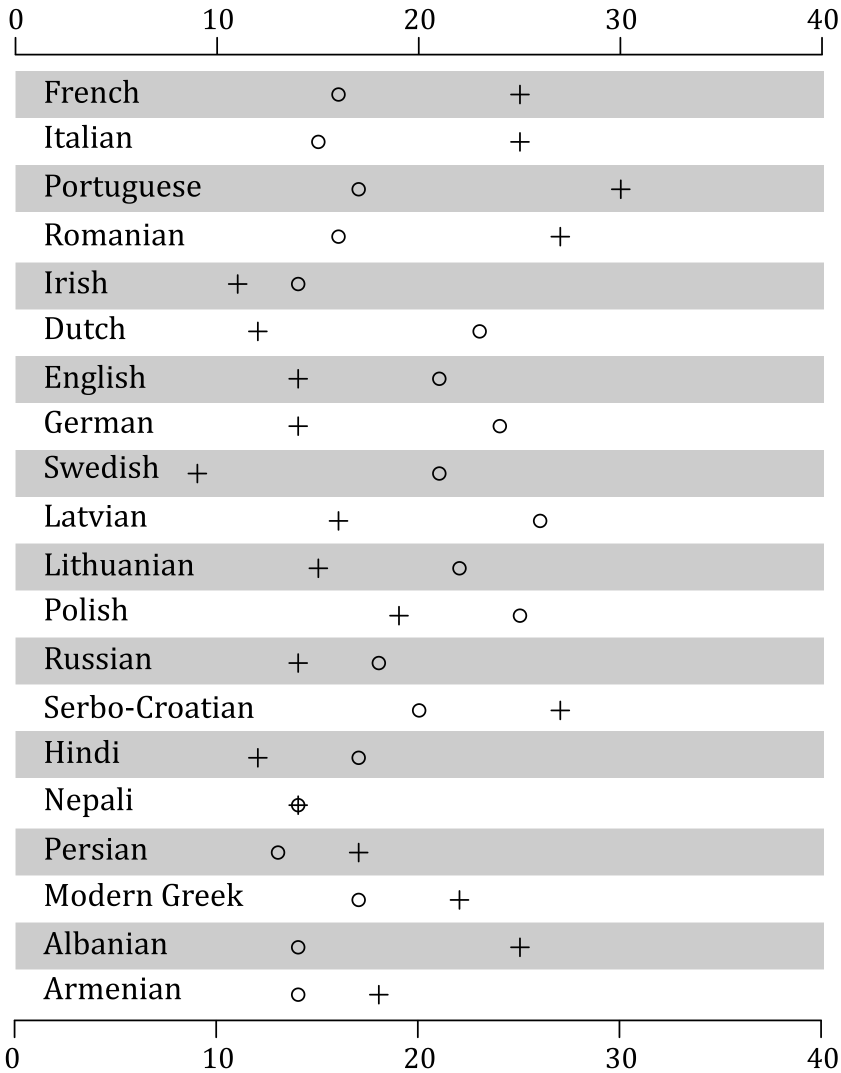


Figure 6.2*: Number of types of manner verbs and path verbs in the 118-sentence sample (○ = manner verbs, + = path verbs)*

The investigation into the etymological origins and rates of change of manner verbs and path verbs starts off with an etymological analysis. The results of this analysis, including a discussion on whether the evolution of individual first tier and second tier manner verbs and path verbs is different, will be discussed in the next section.

Table 6.1: *Number of types of manner verbs and path verbs in the 132-sentence sample*

Language	Manner verbs	Path verbs
French	16	25
Italian	15	25
Portuguese	17	30
Romanian	16	27
Irish	14	11
Dutch	23	12
English	21	14
German	24	14
Swedish	21	9
Latvian	26	16
Lithuanian	22	15
Polish	25	19
Russian	18	14
Serbo-Croatian	20	27
Hindi	17	12
Nepali	14	14
Persian	13	17
Modern Greek	17	22
Albanian	14	25
Armenian	14	18

6.3 An etymological analysis of manner verbs and path verbs

6.3.1 Etymological classifications

The first sources of information on where manner verbs and path verbs come from are of course the etymologies of the verbs in question. For this reason, an etymological investigation of all attested manner verbs and path verbs was conducted. The etymology of these verbs was found by consulting etymological dictionaries, other written source materials, and in certain cases, consultation with experts.[17] The source materials used for each language have been listed in Appendix 10.

The etymology of each verb was classified into one of the following categories. Note that these categories are necessarily a simplification, and do not always do justice to the intricacies of the histories of words. However,

[17] Thanks go to Per Ambrosiani, Johnny Cheung, Hrach Martirosyan, and Ranko Matasovic for answering questions about Russian, Persian, Armenian and Irish etymologies.

distinguishing these classes was required if any cross-linguistic comparison was to be made:

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

The first category, language-specific development, is divided into three subcategories: 1a. unspecified / semantic shift; 1b. derived from a non-verbal element; 1c. borrowed. All three designate developments 'within' the language, i.e. developments that have taken place in the period since the language has split away from its closest sister language. An example of the first subcategory, unspecified internal development / semantic shift, is the Albanian manner verb *nxitoj* 'to hurry', which is cognate with the Proto-Indo-European verb *g^{w}ei̯-3 'to live' (Orel 1998: 303). The manner of motion sense 'hurry' is not attested in any of the Indo-European cognates of *g^{w}ei̯-3 except for *nxitoj*, and therefore it seemed to have been an independent semantic shift that took place only in Albanian. An example of the second subcategory, derivation from a non-verbal element, is the Italian manner verb *camminare* 'to walk', derived from the noun *cammino* 'road' (Dizionari sapere.it 2012). An example of the third subcategory, borrowing, is the Dutch manner verb *spurten* 'to skurry', which was borrowed from the English verb *spurt* (de Vries 1971).

The second category, subgroup-specific development, is used for verbs that have developed within one of the big Indo-European subgroups: Germanic, Romance, Balto-Slavic, and Indo-Iranian. For a verb to be classified in this class, there must exist at least one cognate within the subgroup that is also a manner verb or a path verb, without there being a relevant Indo-European history outside of the subgroup. Examples are the Hindi path verb *girnā* 'to fall' and the Nepali path verb *girnu* 'to fall'. These two verbs are related to Sanskrit *girati*, which means 'devours, swallows' rather than 'falls', and ultimately to Proto-Indo-European *$g^{w}o/erh_3$-o/h_2- 'devouring, devoured' (Turner 1962-1985; Uhlenbeck 1899: 80; de Vaan 2008: 690). None of the cognates outside of Indo-Iranian are therefore path verbs, suggesting that the change in meaning from 'devour' to 'fall' has happened only within the Indo-Iranian subgroup. If the cognate verbs are related to a non-verbal element in a related ancient language, such as the French path verb *monter* 'to ascend' and the Catalan path verb *montar* 'to ascend', which are ultimately related to Latin *mons* 'mountain' (von Wartburg 1922-2003 vol. 6.3: 106), the verbs are classified as subgroup specific developments.

The third category, Indo-European-wide development, is used when a verb has cognates that are also manner or path verbs in at least one other Indo-European branch. An example is the Polish manner verb *pływać* 'to swim', which is cognate with the Proto-Indo-European root **pleh$_3$(u)* 'to flee, run, flow, swim', with Sanskrit *plávate* 'to swim, float', and with English *float* (Derksen 2008: 407; Uhlenbeck 1899: 181).

The fourth category contains those verbs that are (ancient or modern) lexicalizations of preverbs and verbs. As was explained in section 4.2, Proto-Indo-European and the ancient Indo-European languages had a system of path denoting particles that could be positioned anywhere in the sentence (Delbrück 1888, 1893; Hewson and Bubenik 2006; Hofmann and Szantyr 1965; Lehmann 1974). This system was still in place in Homeric Greek, and is illustrated in (54) and (55) (example (54) is repeated from example (43) for convenience). In (54), the path particle *epì* is functioning as a preposition, and in (55), *epì* is functioning as a preverb (although it is separated from its verb, a process called 'tmesis').

54) Homeric Greek

epì	*hoî*	*kalésas…*	*subóten*
to	him.DAT	call.AOR	swineherd.ACC

'having called the swineherd to him' Hewson and Bubenik (2006: 4)

55) Homeric Greek

kai	*epì*	*knéphas*	*hierón*	*élthēy*
and	on	darkness	sacred	come.3SG

'and the sacred darkness closes in' Hewson and Bubenik (2006: 6)

At later stages, namely in Classical Sanskrit, Classical Latin and Classical Greek, these particles became more closely associated with nouns and verbs. The path particles that were closely associated with nouns formed the preposition and postposition systems of contemporary Indo-European languages. The path particles that belonged with verbs formed the preverb systems that later became spatial prefixes. In the transition from Latin to the modern Romance languages, as well as from Sanskrit to the modern Indo-Aryan languages, these spatial prefixes merged with the verb completely. In Balto-Slavic, the system of spatial prefixes is still in place. As a result, lexicalized preverb-verb combinations are an important group of path verbs in many Indo-European languages. Given that the different Indo-European subgroups were all supplied with the same tools to combine preverbs with verb, there are calques of preverb-verb combinations, both within subgroups and between subgroups. Within subgroups an example could be Dutch *oversteken* and German *überqueren*, both mean 'to cross' and are lexicalized combinations of a cognate adverb meaning 'across' and a verb. Latin has calqued several preverb-verb combinations from Greek, which is in a different Indo-European subgroup (see for examples Nicolas 2005: 131 and also Adams 2003: 459ff). The extend to which these lexicalized preverb-verb combinations are independent inventions or calques can be hard to assess, especially as most etymological dictionaries think of these as derivations and often do not pay much attention to highly similar derivations in closely related languages.

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

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

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

6.3.2 Manner verbs and path verbs have different etymologies

The results of the etymological investigation of manner verbs and path verbs are depicted in Figure 6.3 and 6.4 and Appendices 11–16. Figures 6.3 and 6.4 give an overview of the etymologies of manner and path verbs for the 20 Indo-European languages. Appendices 11 and 12 present the etymological classifications of the attested manner verbs and path verbs. Appendices 13 and 14 give summaries of etymology types across the big Indo-European subgroups (Romance, Germanic, Balto-Slavic, and Indo-Iranian). Appendix 15 gives a summary of etymology types per tier for both the manner verbs and the path verbs. Appendix 16 presents the results of paired sample t-tests that indicate whether the distributions of etymological categories across languages of manner verbs and path verbs are significantly different or not.

Looking at Figures 6.3 and 6.4, the most notable differences between manner verbs and path verbs are as follows. Manner verbs often originate at the Proto-Indo-European level and at the subgroup level (see Figure 6.3). On average, 29% of the manner verbs can be traced back to Proto-Indo-European, compared to only 18% of the path verbs (see Appendices 13 and 14). This difference is statistically significant (see Appendix 16, $t = 4.37$, $p \leq 0.001$). In contrast, 25% of the manner verbs can be traced back to the subgroups, compared to 21% for the path verbs (see Appendix 13 and 14). This difference is not statistically significant (see Appendix 16, $t = 1.33$, $p = 0.20$). It also seems that manner verbs have approximately the same distribution of origin types across languages, while path verbs have much more varied origins from language to language. This suggests that in Indo-European, manner verbs are more stable than path verbs: they are typically older and have similar origins across languages.

There are also differences between the etymological origins of manner verbs across the different languages (see Figure 6.3) and subgroups (Appendix 13). Germanic has the largest proportion of manner verbs that originate at the subgroup level (40%, compared with 25% overall, see Appendix 13). This can be seen as evidence for a faster rate of change in the creation of manner verbs at the Proto-Germanic level. Proto-Germanic is likely to have been satellite-framed (see section 4.4), which could have been a pressure to develop more manner verbs before it split up into the different Germanic language groups. Germanic is the only clade where the derivation of manner verbs from non-verbal elements is not attested at all. This is a fairly common process in verb-framed languages such as Portuguese, Romanian, and Modern Greek, and it is also common in Irish. Borrowed manner verbs are most common in Germanic, French and Albanian. Lexicalized preverb-verb combinations are not very common origins of manner verbs, although they are attested in Romanian, Portuguese, Nepali, and Assamese. The large number of Lithuanian manner verbs that have an Indo-European origin is probably due to the fact that Lithuanian is very conservative in many respects, also for a long time Lithuanian etymologies were the best-studied etymologies of all Balto-Slavic languages.

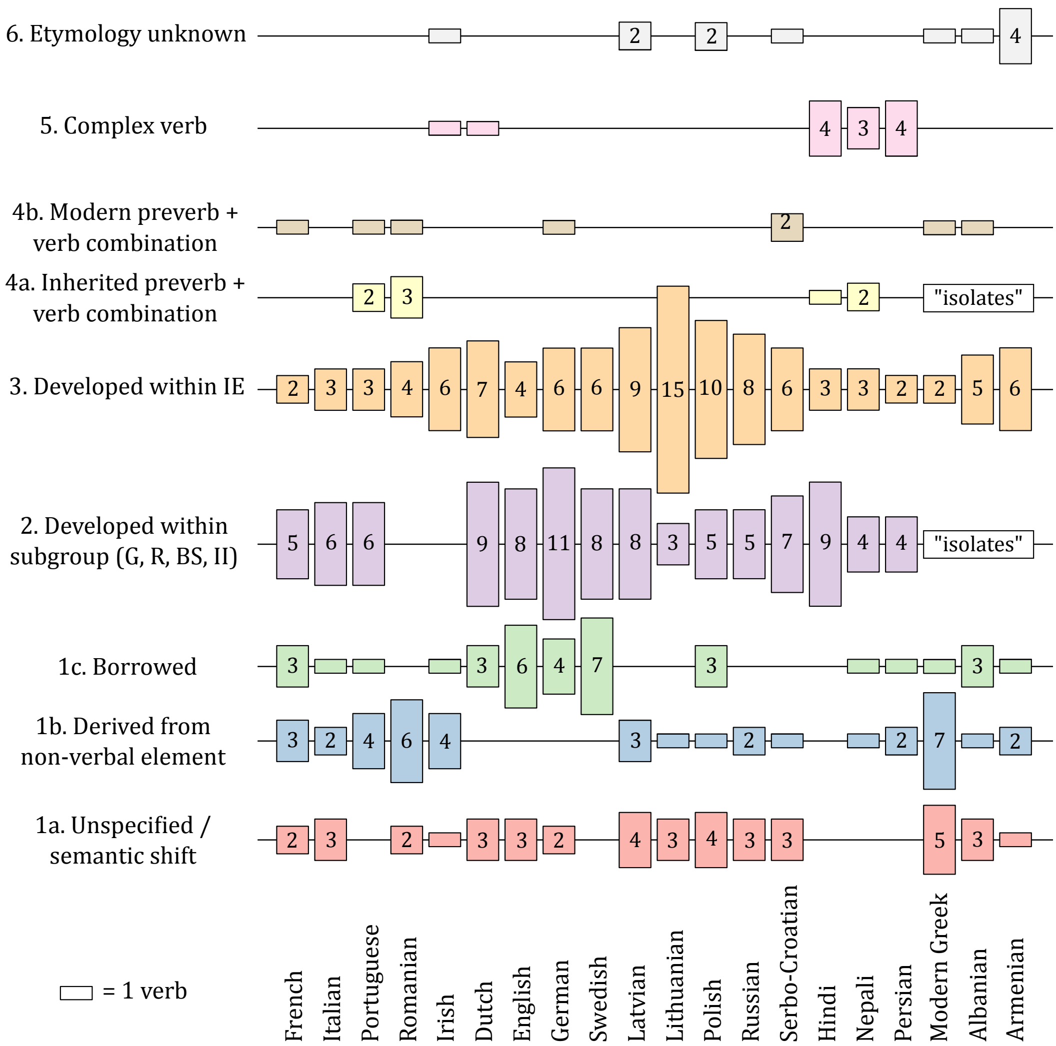


Figure 6.3: *The etymologies of the manner verbs in the 132-sentence sample. See the beginning of section 6.3.1 for a description of the etymological categories.*

As for path verbs, their etymologies are generally far more diverse if we compare them across languages (see Figure 6.4). The formation of path verbs through lexicalized preverb-verb combinations is a very important process of path verb origins. Compared to manner verbs, significantly more path verbs are formed through inherited, older preverb-verb combinations (see Appendix 16, t = -2.78, p ≤ 0.05) and modern preverb-verb combinations (t = -4.02, p ≤ 0.001). The formation of path verbs through inherited preverb-verb combinations is attested in all Romance and in all Indo-Iranian languages. The formation of path verbs through modern preverb-verb combinations is attested in French, Portuguese, Dutch, German, Swedish, all Balto-Slavic languages, Persian, Albanian, Modern Greek and Irish. This indicates that in these languages, some form of the ancient preverb system is still productive or has been productive until recently.

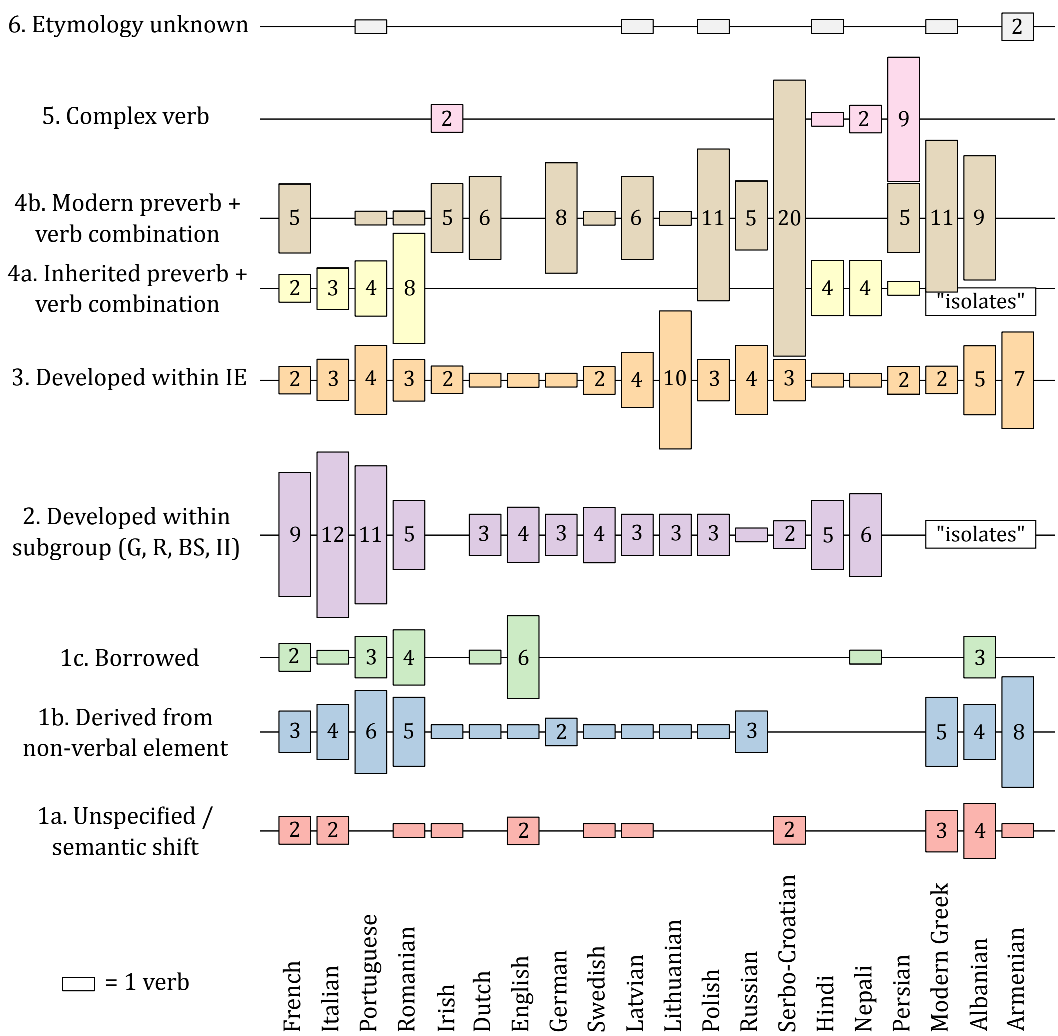


***Figure 6.4**: The etymologies of the path verbs in the 132-sentence sample. See the beginning of section 6.3.1 for a description of the etymological categories.*

Both the Romance and the Indo-Iranian languages have many path verbs with an origin at the subgroup level: 35% of the Romance path verbs and 28% of the

Indo-Iranian path verbs originate at the subgroup level, compared with 21% overall (see Appendix 14). This can be regarded as evidence for an increase of the number of path verbs in the Romance and Indo-Iranian subgroups. Germanic also has an increase of the number of path verbs at the subgroup level: 30% of Germanic path verbs originate at the subgroup level, compared with 21% on average. English has borrowed almost half of its path verbs (Aske 1989). The large number of Lithuanian path verbs that have an Indo-European origin is again likely to be due to Lithuanian (etymologies) being conservative and well studied. The lack of preverb-verb combinations in Armenian can be explained by the fact that even in Classical Armenian, preverb-verb combinations were rarely used (Schmitt 1981; see section 4.2).

Given the division between first tier (more basic) and second tier (more specific) manner verbs and path verbs that was briefly introduced in section 5.1.2, 5.2.2, 5.1.3, and 5.2.3, it is interesting to see whether second tier verbs are newer and whether first tier verbs are older. This might lead to generalizations about the sequential order in which manner verbs and path verbs may be added to the lexicon, such as those that exist for other semantic domains, most famously color words as in Berlin and Kay (1969).

The current analysis only provides a tentative answer to the question of whether second tier verbs are typically newer and whether first tier verbs are typically older. Since satellite-framed languages have few path verbs and verb-framed languages have few manner verbs, the two groups are difficult to compare as the total number of verbs is not similar. Even so, a summary of etymologies separated by tier is presented in Appendix 15. Across the whole of the Indo-European language family, 32% of the first tier manner verbs emerge at the Indo-European level, compared to 26% of the second tier manner verbs. Second tier manner verbs are slightly more often derived from a non-verbal element than first tier verbs (2nd tier: 14%; 1st tier: 9%); the same applies to borrowings (2nd tier: 14%; 1rd tier 7%). Comparing the big subgroups, it is certainly not true that second tier manner verbs emerge on the subgroup level and that first tier manner verbs emerge at the Indo-European level in Germanic and Balto-Slavic languages. Germanic and Balto-Slavic second tier manner verbs are just as likely to originate at the Indo-European level as they are to originate at the subgroup level. However, all Romance manner verbs that emerge at the Indo-European level (which are Italian *saltare*, Portuguese *saltar*, French *sauter*, and Romanian *sări* 'to jump', Portuguese *nadar*, Italian *nuotare*, and Romanian *înota* 'to swim', as well as Romanian *umbla* 'to wander'), except for Romanian *fugi* 'to flee', are first tier manner verbs.

For the Indo-European path verb lexicons, no evidence can be found for the early emergence of first tier path verbs and the later emergence of second tier path verbs. Appendix 15 indicates that slightly more first tier path verbs have an Indo-European origin (18%) as compared with second tier path verbs

(15%). But this difference is very small. And in contrast, 33% of first tier verbs are modern preverb-verb lexicalizations, as opposed to 18% of second tier verbs. The only evidence pointing in the direction of the late emergence of second tier verbs is that they are more often derived from a non-verbal element (19% of second tier as opposed to 8% of first tier path verbs). Note, however, that this division of first and second tier path verbs has been made only on the basis of path verbs attested in the current study, and it might not be a good representation of internal path verb class divisions. Lexicalization of preverb-verb combinations has been so common that it has affected all Indo-European languages and has given rise to widespread parallel evolution and calquing of path verbs. The range of Indo-European verbs meaning 'to return', for instance, are sometimes similar sounding but actually partly parallel developments and partly calques: French *retourner*, Portuguese *retornar*, Dutch *terugkeren*, German *zurückkehren*, Swedish *återvända*, Modern Greek *epistrefo*, Persian *bāzgaštan*, and Russian *vozvrašat'sja* are all based on the combination of preverbal element with a verb meaning TURN. In some languages, the preverbal element is still productive, in others it is not.

The results presented in Figures 6.3 and 6.4 and Appendices 11–16 suggest that manner verbs and path verbs typically have quite different etymological histories. The question that will be addressed in the next section is whether inferring the rates of evolutionary change can provide more information on the evolutionary processes that underlie the etymological verb histories.

6.4 Rates of evolutionary change for manner verb and path verb lexicon size

Phylogenetic comparative methods were used to study the evolution of manner verbs and path verbs further. Specifically, the hypotheses that manner verb lexicon size evolves faster in satellite-framed languages and path verb lexicon size evolves faster in verb-framed languages are investigated. These hypotheses were already partly confirmed by the large amount of manner verbs that arise at the subgroup level for Germanic languages, and the large amount of path verbs that arise at the subgroup level for Romance languages, as presented in section 6.3.2. In this section, it is shown that similar patterns emerge when the rates of evolutionary change are investigated. In order to do this, a set of phylogenetic trees is transformed in various ways, in order to see which model of evolution provides the best fit to the data. The sample of phylogenetic trees that was used for this task has been taken from Bouckaert et al. (2012) and introduced in section 1.3.2.2. Figure 1.6, which gives the maximum clade credibility tree of the tree sample used for the 118-sentence sample, is presented below as Figure 6.4 for convenience.

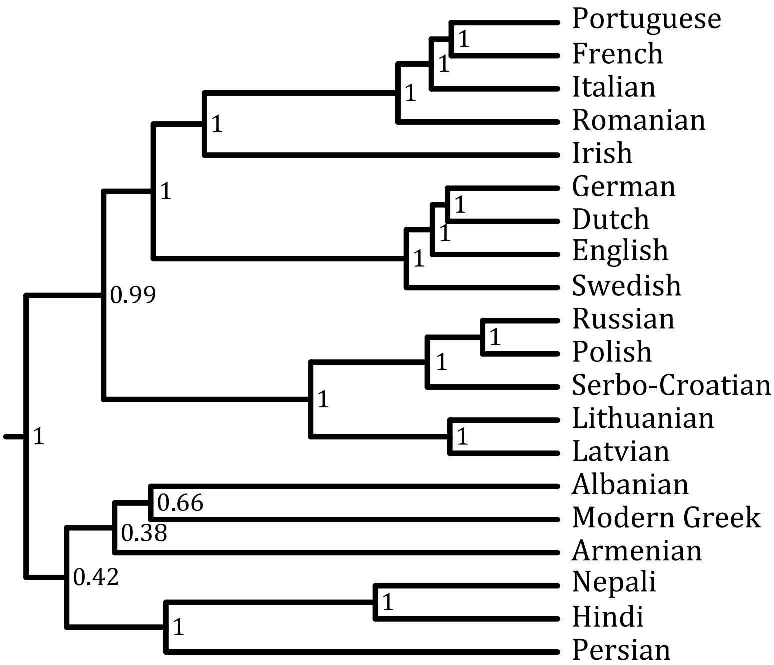


***Figure 6.4**: The maximum clade credibility tree of 1000 phylogenies sampled from the posterior sample of trees in Bouckaert et al. (2012). The MCC tree was pruned to include only the 20 languages featured in this dissertation.*

Testing for differential rates of evolution of manner verb lexicon size and path verb lexicon size was done by testing whether a model of variable rates of change was more likely to generate the data as a model of invariable rates of change. The methodology is explained below. Part of this discussion is also included in section 1.3.3.1, but this section explains the methods in more detail.

The standard model of evolution of continuous features (i.e. features that range on a scale rather than in a number of separate classes, such as lexicon size) is the constant-variance random walk model of evolution (Pagel 1999a), also sometimes called 'Brownian motion'. In this model, the constant rate of change is a measure of how much a feature changes instantaneously at each moment of time. For a given data set and phylogenetic tree, the rate of change is integrated over evolutionary time as represented by branch length. Traditional 'standard' phylogenetic analyses estimate a single constant rate of change, while taking into account all sub-branches and data points (Pagel 1999a). This is simply in accordance with the simplest model of evolution. This paper is concerned with more advanced phylogenetic analysis that allows the rate of change to differ in different parts of the tree. But before introducing this type of analysis, an illustration what the rate of change on a phylogenetic tree signifies is given in Figure 6.5 and below.

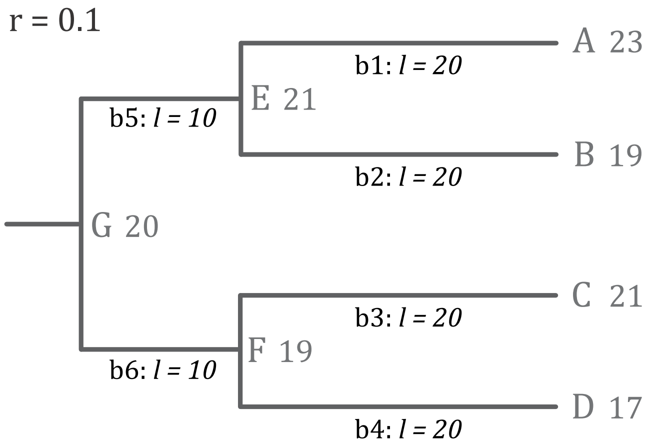


Figure 6.5*: A single rate of change of manner verb lexicon size in an hypothetical data set*

In Figure 6.5, an ultrametric phylogenetic tree is depicted that represents the evolution of manner verb class size in four hypothetical languages called A, B, C, and D. Manner verb class size for each of these languages is listed to the right of the names A, B, C, and D. Language A and B are descendants of language E, while language C and D are descendants of language F, which are marked by ‘E’ and ‘F’ on the respective internal nodes of the tree. All languages descend from language G, which is also marked on the root of the tree. Let us assume that we know the ancestral states of the languages E, F, and G – the numbers to the right of the names E, F, and G give the ancestral manner verb size of each of these ancestral languages. The sub-branches are labeled b1 through b6. Since the branches are a representation of the evolutionary process through which languages A, B, C, and D emerged, branch length is a representation of the evolutionary timeline. Evolutionary time can be measured in different ways, for phylogenetic trees which have not been time-calibrated it is typically a representation of the amount of linguistic change, i.e. the number of cognate replacements. For time-calibrated trees evolutionary time typically is given in years. In Figure 6.5 however the unit of evolutionary time is a generation, and the length of each sub-branch is given under each sub-branch in generations (a generation being approximately 20 years). Given that we know the ancestral states, the contemporary states, and the length of each sub-branch in generations, the rate of change in each sub-branch that is needed to generate the distribution of contemporary manner verb lexicon sizes is 0.1. For example, for language A to have 23 manner verbs, it has gained 0.1 manner verb per generation (or 1 verb per 10 generations) since its ancestor language G, 30 generations ago.

Note that the rate of change does not specify the direction of change: the evolutionary model depicted in Figure 6.5 specifies an increase of 0.1 per generation along branches b5 and b1 to model the contemporary manner verb lexicon size of language A, while the model relates a decrease of 0.1 per generation along branches b6 and b4 to model the contemporary manner verb lexicon size of language D. Change along the branches of a phylogenetic tree can, at any moment in time, both be a value increase as well as a value decrease, as the size of the lexicon can become larger or smaller. The rate of change simply reflects the pace at which such changes are taking place given a unit of evolutionary time.

In this example, a single constant rate of change suffixes to explain the different manner verb lexicons of our hypothetical languages A, B, C, and D. However, many linguistic features do not evolve at a constant rate of change. If the comparative dataset has evolved in a particular directional trend, or if the rates of change vary from time period to time period, the fit between the dataset and the phylogenetic tree will not be optimal when only a single constant rate of change is estimated. The fit between the dataset and the tree will typically improve if the phylogenetic tree is adjusted to take into account information on varying rates of change or directional trends. An example of a dataset that can be explained by employing a set of different rates of change is given in Figure 6.6 and below.

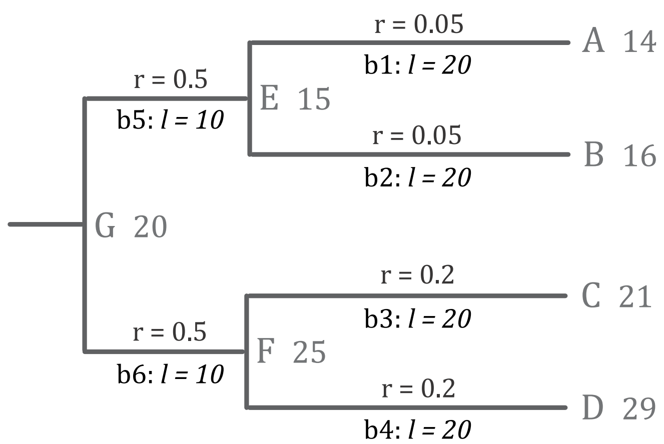


***Figure 6.6**: Different rates of change of manner verb lexicon size in an hypothetical data set*

In Figure 6.6, a different tree and set of languages are depicted, the languages are again named A through G. The manner verb lexicon sizes of these languages vary more widely than those depicted in Figure 6.5, and different rates of change are used to fit this diversity on the tree. In branches b5 and b6, the rate of change is fast, 0.5 per generation (or 1 verb per 2 generations), whereas the rates of change in branches b1 and b2 are slow, 0.05 per generation (or 1 verb per 20 generations). In the history of language A, it has first lost verbs at a fast rate (0.5) on branch b5, and then at a more slower rate (0.05) on branch b1, in total having lost 6 verbs compared to its ancestor language G.

Rates of change of manner verb lexicon size and path verb lexicon size were optimized using the maximum clade credibility tree presented in Figure 6.4. In this way it was tested whether there was any evidence for larger rates of change of manner verb lexicon size in branches leading to satellite-framed languages versus larger rates of change in path verb lexicon size in branches leading to verb-framed languages. This was achieved by modeling the evolution of the two datasets on lexicon size on the maximum clade credibility tree. The modeling was done with the software *BayesTraits*, which uses a Phylogenetic Generalized Least Squares model of trait evolution (Pagel 1997, 1999b) in a Bayesian reversible-jump Markov chain Monte Carlo framework (Green 1995) that can be used to trace the evolutionary history of shifts in the rate of evolution (Venditti et al. 2011). Given a phylogenetic tree and a dataset, the Bayesian MCMC algorithm allows for the optimization of the rate of change parameters to find the best model of evolution of the data on the phylogenetic tree. Because this method finds optimized (individual and clade) branch lengths, it can only accommodate a single (MCC) phylogenetic tree, as comparing optimized branch lengths across different tree topologies is computationally very difficult.

To explain how the Bayesian MCMC algorithm accomplishes this, let us imagine we have a another sample dataset, this time of five related languages A, B, C, D, and E. Imagine that language A has 16 manner verbs, B has 17, C has 20, D has 30, and E has 24. We also know that languages A, B, and C are closely related, and that D and E are closely related. The group of language A, B, and C has smaller manner verb lexicons (16–20 verbs) than the group of language D and E (24-30), and it could be possible that language D and E have an increased rate of change of manner verb lexicon size, or that language A, B, and C have an decreased rate of change of manner verb lexicon size. In order to test this, the Bayesian MCMC algorithm can be used to find the most optimal shared and non-shared branch lengths of the phylogenetic tree of language A, B, C, D, and E. From these optimized branch lengths, the optimized rates of change can be derived: branches that have been scaled to be longer imply increased rates of change, while branches that have been shrunken to be shorter imply decreased rates of change.

In theory, the lengths of the branches of the tree can have an almost infinite number of values. However, some values will be far more likely than others – the aim of the Bayesian MCMC algorithm is to find those values which have the highest likelihood. It searches the so-called parameter space that contains all the possible combinations of values for the different parameters that describe the phylogenetic tree with its branch lengths. The algorithm does this by building a Markov chain, a mathematical device that jumps from state to state. For the current example, a state is the phylogenetic tree of language A, B, C, D, and E, with some parameter values describing the branch lengths of this tree. This process is illustrated in Figure 6.7, where an illustration of an MCMC chain within a parameter space is given for our hypothetical sample dataset of language A, B, C, D, and E. In Figure 6.7, the dark line represents the MCMC chain that is jumping from state to state within the parameter space, beginning at the top and moving downwards. Each numbered tree is a state, each change from one numbered tree to the next on the chain represents a jump. The probability of a jump to a new state is determined by how much of an improvement in likelihood the new state is over the current state. New states that have a lower likelihood than the current state may be adopted as well, depending on how much worse it is compared with the current state.

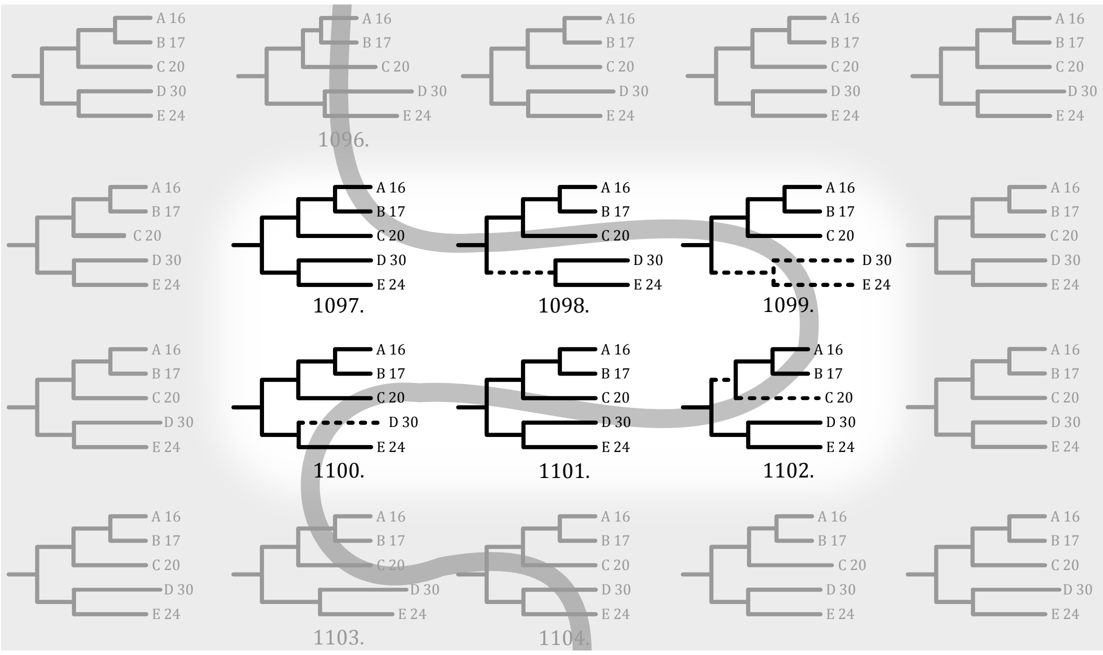


***Figure 6.7**: A MCMC chain that searches the tree space to optimize branch lengths. The dark line represents the MCMC chain, the numbered trees represent the different states on the chain.*

In the parameter space in Figure 6.7, the trees have the same topological structure, only the length of the various sub-branches is different. The MCMC chain samples the parameter space by making chances in branch length and jumping to a new state (a tree with different branch lengths) when the proposed new state is a sufficient improvement. Changes in branch length are marked by the dashed branches in Figure 6.7. Branch length changes are made by optimizing two parameters: one that changes the length of a single branch (single branch modification, as in tree no. 1098, 1100, and 1102 in Figure 6.7), and one that changes the length of a branch and all its descendent branches (clade modification, as in tree no. 1099). Note that the original, unchanged tree is also part of the parameter space, and can be visited by the chain again and again if it has a high likelihood (tree no. 1097 and 1101 have undergone no branch length changes). When the chain is run for an appropriate number of iterations, it becomes stationary. In stationary distribution, the chain no longer moves to different phylogenetic trees with higher and higher likelihoods, but moves around a mean likelihood. It samples parameter values in proportion to their frequency of occurrence in the parameter space. In this way it constructs a sample of phylogenetic trees with optimized branch lengths that constitute the posterior probability distribution. The result is a sample of trees that have the most optimal branch lengths given the data, which in turn relate the most optimal rates of change given the dataset. Such a sample is then further summarized into a single tree by taking the mean branch lengths over the posterior tree sample (see Figures 6.8 and 6.9 below). The resulting tree has optimized branch lengths that reflect the most optimal rates of change given the data: if the rate of change in a specific branch has been lower, that branch will have become shorter; if the rate of change was inferred to be higher, that branch will have become longer.

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

To see whether the alternative model is an improvement on the null model, the harmonic means of the Bayesian posterior density of likelihoods are compared using Bayes factors (Pagel et al. 2004; Kass and Raftery 1995). The results are summarized in Table 6.2. Table 6.2 shows that for both the manner verb lexicon size analysis and the path verb lexicon size analysis, the variable rate models (M_v and P_v) are favored over the fixed rate models (M_f and P_f). However, neither of these results are strong statistical evidence for the existence of such tendencies, as the Bayes factor values are low (between -2–2, "barely worth mentioning").

***Table 6.2**: Comparison of variable rate and fixed rate models of evolution of motion verb lexicon size in the 132-sentence sample*

Dataset	Fixed model	ln Lh[a]	Variable model	ln Lh	BF[b]
Manner verb lexicon size	Mf	19.60	Mv	19.63	-0.06
Path verb lexicon size	Pf	8.78	Pv	9.06	-0.56

[a]The ln Lh denotes the marginal likelihood of each analysis. This is the harmonic mean of the ln likelihoods of the posterior 500 samples taken from the stationary part of the chain.
[b]Bayes factors were calculated by taking BF = 2(ln Lh Mf/Pf – ln Lh Mv/Pv) (Pagel et al. 2004). Bayes factor values indicate which model is statistically favored. If values are positive, the fixed model is better supported. If values are negative, the variable model is better supported. The Bayes factor score is read as follows: 0-2, barely worth mentioning; 2-6, positive; 6-10, strong; >10, very strong (Kass and Raftery 1995).

The statistically insignificant results of the comparison of the variable rate and fixed rate model presented in Table 6.2 are likely to be due to the small number of languages used for the current analysis as well as the limited amount of change in rates that was found (see below). As will be discussed below, the changes in evolutionary rate that were found were quite small – the branches that were stretched the furthest do not even signify a doubling of the rate of change. Methods for identifying rates of change are typically tested and perform well on far greater changes in rate – see for instance Revell et al. (2011: 141), who test their method against 5-fold and 10-fold rate increases and decreases (see also Eastman et al. 2011 for similar analyses). For small rate of change differences like the ones found here to be statistically significant, a much larger number of languages would be needed. However, the exact size of the language sample that would be required is difficult to determine as the detection of significant rate of change differences depends on the topology of the phylogeny, the comparative dataset, and the amount of rate variation found in the feature.

Even though there is no strong evidence that the variable rates model provides a significantly better model of evolution for these two datasets, it does tell us something about rates of change and why using this methodology is interesting. The reason for this is that I obtained distinct results for manner verb lexicon size and path verb lexicon size: the phylogenetic trees with optimized branch lengths and the rates of change that can be derived from them do suggest different rates of change for the two different verb classes. A comparison of the results of the manner verb and path verb analyses is presented in Tables 6.3 and 6.4 and Figures 6.8 and 6.9. Table 6.3 gives the average rates of change within each of the four main subgroups of Indo-European (Romance, Germanic, Balto-Slavic, and Indo-Iranian) compared with the average rates of change for the whole tree. Figures 6.8 and 6.9 present the maximum clade credibility tree presented in Figure 6.4 scaled to reflect the optimal rates of change of manner verb lexicon size and path verb lexicon size as found by the Bayesian MCMC algorithm that used variable rate model. Table 6.4 gives an overview of the percentage of scaling of the branches leading to the individual contemporary languages as presented in Figures 6.8 and 6.9.

Table 6.3*: Optimized rates of change within the main subgroups of Indo-European on the phylogenetic tree by Bouckaert et al. (2012) scaled to reflect manner verb lexicon size and path verb lexicon size evolution*

Subgroup	Manner verbs	Path verbs
Complete tree	0.91	1.07
Romance	0.85	1.26
Germanic	0.90	1.08
Balto-Slavic	1.17	1.17
Indo-Iranian	0.88	1.04

Table 6.3 relates that overall, the rates of change across the whole tree are higher than 1 for the path verb dataset, which indicates that rates of change have been increasing over time. This is in line with expectations, as the adaptions of rates of change are not a zero-sum game, i.e. the total length summed over all the branches does not need to be the same as that of the original tree. Rather, the Brownian motion model of evolution works to make those branches longer for which an increase of rate of change is most likely, and makes those branches shorter for which an decrease of rate of change, or even complete stasis, is most likely. However, apparent slow rate of change or stasis can be explained on long branches as well as on short branches. Imagine for instance that a given language has 16 manner verbs, while we know its ancestor had 15. In such a case, Brownian motion dictates that the language could have lost and gained several verbs, but ultimately was left with one verb more than its ancestor. There is more than enough time for such Brownian change on a long branch. The Brownian motion model in this sense 'prefers' long branches over short branches, as long branches provide more evolutionary time for changes to occur. However, if subgroups behave highly uniformly (i.e. have very similar manner verb lexicon sizes, for instance), this indicates a decrease in rate or stasis in the whole subgroup, and the branches in that clade may become shorter. For

individual branches, a decrease in rate of change will only be found when there is substantial evidence for a slow rate or stasis.

Table 6.3 and Figures 6.8 and 6.9 relate that in order to optimize the variable rates of change on the phylogenetic tree, most branches are scaled to be longer to some extent for the path verb lexicon size analysis, while most branches are scaled to be shorter for the manner verb lexicon size analysis. This difference can be related to the comparison of the null model and the alternative model in Table 6.2: the path verb lexicon seem to evolve at a higher rate than as is predicted by the phylogenetic tree overall and prefers the variable rates model, whereas the manner verb lexicon evolves faster in certain parts and slower in most parts, therefore on the whole the constant rate model is the better match.

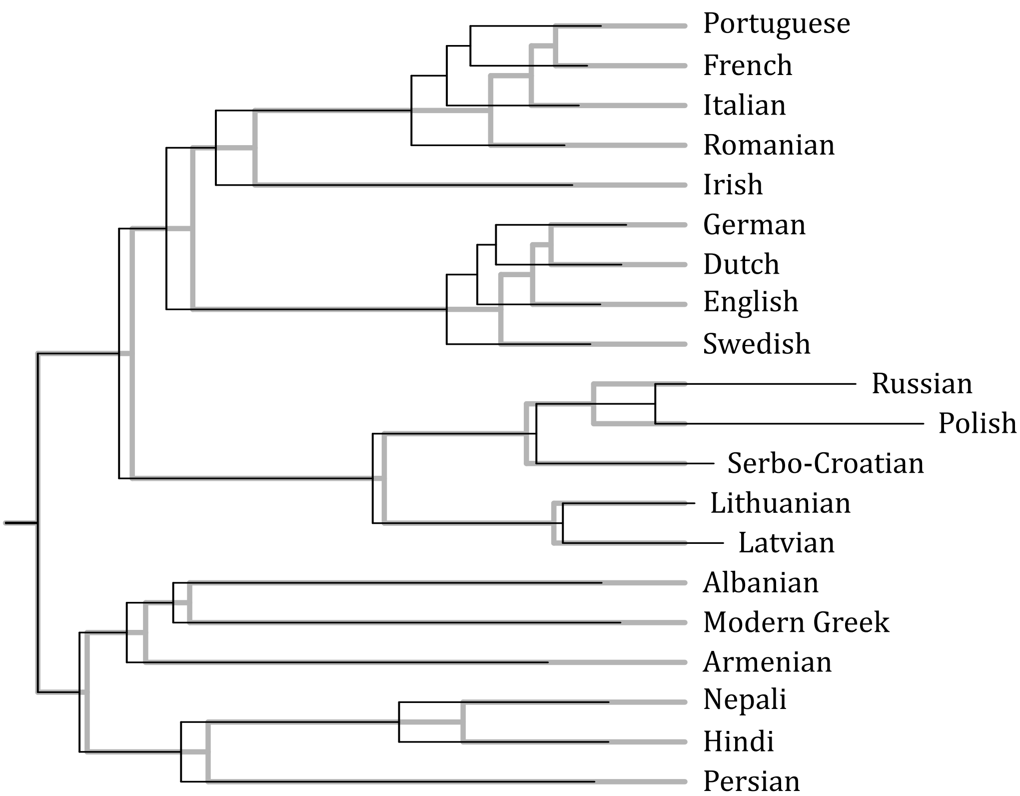


Figure 6.8*: The Indo-European tree by Bouckaert et al. (2012) scaled to reflect manner verb lexicon size. The trees in gray are the original tree, the trees in black are the scaled trees.*

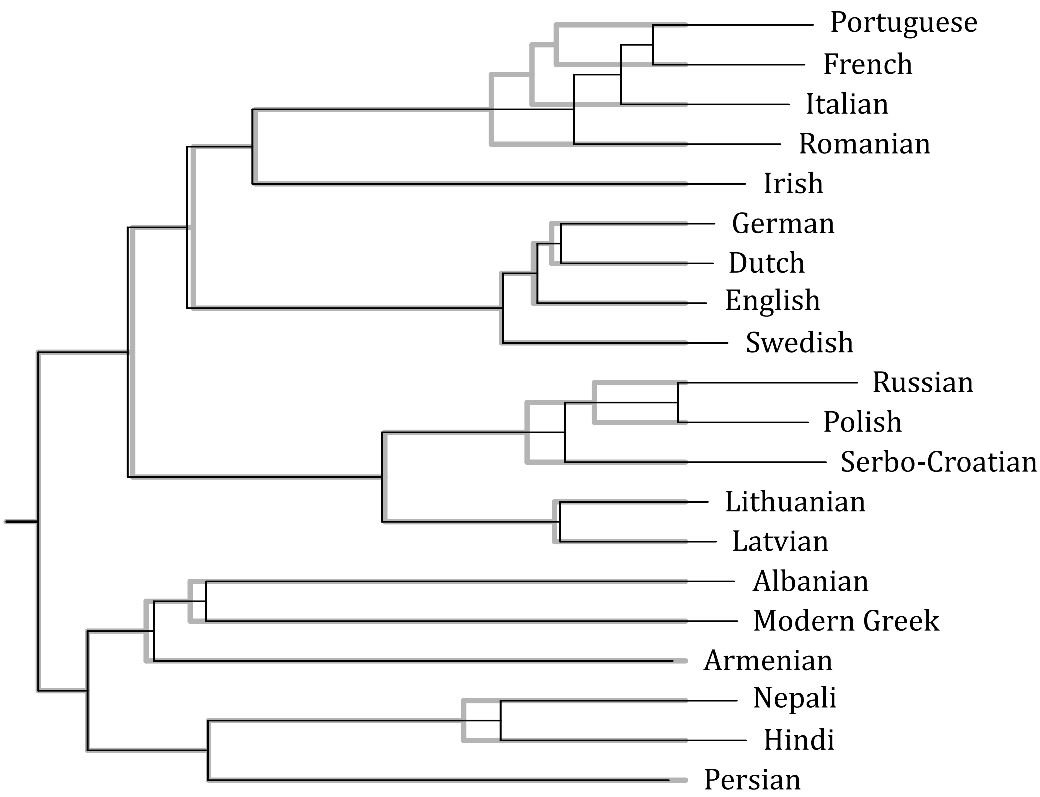


Figure 6.9*: The Indo-European tree by Bouckaert et al. (2012) scaled to reflect path verb lexicon size evolution. The trees in gray are the original tree, the trees in black are the scaled trees.*

If we compare the individual branches, such as those leading to the four big Indo-European subgroups (in bold in Table 6.4) and the individual languages (Figures 6.8 and 6.9 and Table 6.4), we find that some branches are optimized to be longer to account for manner verb lexicon size while others are estimated to be longer to account for path verb lexicon size. Table 6.4 displays the variability in branch length optimization. The branches of Romance languages, including the one leading to Proto-Romance, are scaled to be longer (Table 6.4 and Figures 6.8 and 6.9) and have higher average rates of change (Table 6.3) to account for path verb lexicon size evolution than manner verb lexicon size evolution. The branches of Germanic languages, including the one leading to Proto-Germanic, are scaled to be longer and have higher average rates of change to account for path verb lexicon size evolution than for manner verb lexicon size evolution. The branches of Balto-Slavic languages, including the one leading to Proto-Balto-Slavic, are scaled to be longer to account for manner verb lexicon size evolution than for path verb lexicon size evolution. The branches of Indo-Iranian languages, including the one leading to Proto-Indo-Iranian, are scaled to be longer and have higher average rates of change to account for path verb lexicon size evolution than for manner verb lexicon size evolution. Following the overall pattern within the tree, the branches leading to the remaining four languages, Irish, Albanian, Armenian, and Modern Greek, are scaled to become slightly shorter to account for manner verb lexicon size evolution and slightly longer to account for path verb lexicon size evolution. For these four languages, the optimized branch lengths imply slightly accelerated evolution of path verbs as compared with manner verbs.

These results suggest that the satellite-framed Balto-Slavic languages have higher rates of evolutionary change for manner verbs, while the verb-framed Romance languages and non-satellite-framed Indo-Iranian languages have higher rates of evolutionary change for path verbs. Contra expectations, satellite-framed Germanic languages have higher rates of evolutionary change for path verbs than for manner verbs. The reason for this result seems to be that the manner dataset is as a whole far more homogenous than the path verb dataset: the range of manner verb lexicon size is 13–26, while the range of path verb lexicon size is 9–30. Especially the Romance and Germanic subgroups have very consistent manner verb lexicon sizes: Romance languages have between 15–17 manner verbs, and Germanic languages have between 21–24 manner verbs. Comparing these results with Verkerk (submitted), where the same analyses were conducted on a slightly different manner verb and path verb dataset, it seems that the lack of divergence within subgroups has caused lower rates of change throughout the Indo-European tree for the manner verb lexicon size analysis. Note that we see the opposite in the manner verb lexicon size analysis for Polish (29 manner verbs) and Russian (19 manner verbs), that both have massively stretched branches. More divergence within subgroups or between sister languages causes longer branches in that subgroup, while less divergence within subgroups causes shorter branches in that subgroup. Note also that out of the three subgroups with descreased rates of change for manner verb lexicon size, the Germanic rates are highest.

Table 6.4: *Scaling of branches of the phylogenetic tree by Bouckaert et al. (2012) scaled to reflect manner verb lexicon size and path verb lexicon size evolution*

Branch leading to	Manner verbs	Path verbs
Proto-Romance	**84%**	**137%**
French	86%	119%
Italian	84%	116%
Portuguese	88%	120%
Romanian	82%	115%
Irish	**83%**	**110%**
Proto-Germanic	**92%**	**103%**
Dutch	91%	105%
English	88%	103%
German	92%	105%
Swedish	86%	107%
Proto-Balto-Slavic	**102%**	**101%**
Latvian	107%	105%
Lithuanian	102%	104%
Polish	138%	119%
Russian	128%	127%
Serbo-Croatian	105%	122%
Proto-Indo-Iranian	**85%**	**100%**
Hindi	89%	110%
Nepali	89%	108%
Persian	87%	98%
Modern Greek	**91%**	**108%**
Albanian	**88%**	**108%**
Armenian	**80%**	**98%**

6.5 The evolutionary processes that shape the lexicon

The results of the etymological investigation and the branch length optimization point out several differences between the evolutionary processes that generate manner verbs and those that generate path verbs in Indo-European:

1. Manner verbs typically have longer histories than path verbs, or, in other words, path verbs are typically derived and therefore less old;
2. The process in which preverb-verb combinations merged into path verbs is of immense importance for the class of path verbs, but not for the class of manner verbs;
3. There is evidence for higher rates of change in the manner verb lexicon for the Balto-Slavic subgroups, leading to a larger manner verb lexicon;

4. There is evidence for higher rates of change in the path verb lexicon for the Romance and Indo-Iranian subgroups, leading to larger path verb lexicons.

The findings provide evidence for a partial affirmative answer to the hypotheses put forward in section 1: yes, there is evidence that manner verb evolution proceeds at a quicker pace in the satellite-framed subgroups (Balto-Slavic) and yes, there is evidence that path verb evolution proceeds at a quicker pace in the verb-framed subgroups (Romance, Indo-Iranian). The finding that manner verbs evolve faster in Balto-Slavic and path verbs evolve faster in Romance support the correlations that were found between motion encoding and the lexicon in chapter 5. The fact that we did not find increased rates of change for manner verb lexicon size but rather increased rates of change for path verb lexicon size in Germanic is due to manner verb lexicon size being more similar across languages and being very similar within subgroups. To compare, see Verkerk (submitted) for an analysis that includes the non-motion verbs used as manner verbs in the counts for manner verb lexicon size that does not have this problem.

The etymological origins of manner verbs and path verbs are readily unified with the results of the rate of change optimizations for the big Indo-European subgroups. The higher rates of change for manner verb lexicon size in Balto-Slavic can be related to the high number of manner verbs that have emerged at the level of the Balto-Slavic subgroup (see also Dickey 2010; Greenberg 2010). The higher rate of change for path verb lexicon size in Romance and Indo-Iranian can be related to the high number of path verbs that have emerged at the level of the Romance and Indo-Iranian subgroup and through inherited preverb-verb combinations.

It is more difficult, however, to compare the inferred rates of change for the individual Indo-European languages to the etymological patterns. In some cases, ancestral states explain the results. Swedish, for example, has a very low number of path verbs, lower than any of the other Germanic languages. Swedish also has the largest optimized branch length of all the Germanic languages in the phylogenetic tree scaled to optimize the evolution of path verb lexicon size. The high rate of change suggested by this stretched branch reflects the fast reduction of the size of the Swedish path verb lexicon. However, not all of the optimizations for individual languages can be explained in this way. And it is impossible to compare rates of change directly to etymological sources: English has the highest number of manner verbs that originate at the language level, i.e. after it split up with its closest sister language, but does not have a very long optimized branch length. The reason for this is that the rate of change analysis does not have information on when exactly a verb appears – information that we do have because of the etymological analysis. The rate of change analysis only has information on the size of the two motion verb classes, and has to infer from this in which parts of the tree the rate of change was higher and in which parts of the tree it was lower. In the future, it might become possible to give the phylogenetic algorithm etymological information, so that the two types of analyses are truly unified and rates of change for individual languages reflect etymological history more closely.

What do these results tell us about the processes that drive change in manner verb lexicon evolution and path verb lexicon evolution? Although this dissertation has focused on phylogenetic models throughout, in order to gain a wider perspective on the motion verb lexicon evolution, we need to go back to the areal perspective that was introduced in section 4.5.2. Within the last two decades, a picture of a European north-south divide in motion event encoding patterns has emerged: north and central Europe are satellite-framed (Germanic languages, Balto-Slavic languages (see for these Slobin 2005b), Finno-Ugric languages including Estonian, Finnish and Hungarian, Daghestanian languages including Avar, Lezgian and Udi, and Kartvalian languages including Georgian (see for these Wälchli 2009: 214)) whereas south Europe is verb-framed (Romance languages (Slobin 2005b), Basque (Ibarretxe-Antuñano 2004), Albanian (Verkerk 2014c), and Turkish (Slobin 2005b)), with several languages 'in between' that have mixed encoding (northern Italian dialects (Slobin 2005b), Serbo-Croatian (Filipović 2007; Vidaković 2012), and Modern Greek (Hickmann et al. to appear)). The proximity of satellite-framed languages in the north is likely to have contributed to the maintenance of satellite-framed encoding and large manner verb lexicons in the Germanic and Balto-Slavic languages. Vice versa, the proximity of verb-framed languages in the south is likely to have contributed to the maintenance of verb-framed encoding and larger path verb lexicons in the Romance languages.

This areal perspective allows us to unify the results of the etymological investigation and the optimization of rate of change of manner verb lexicon evolution presented in the current paper with the claims made in the motion literature. Slobin (2004) proposed that speakers of satellite-framed languages pay more attention to the encoding of manner of motion information, which results in the larger and larger manner verb lexicons of these languages. This is shown by the higher rates of change of manner verb lexicon evolution for Balto-Slavic. In addition, there is some evidence for different etymologies for manner verbs from different tiers: first tier manner verbs are slightly more likely to have an Indo-European history, while second tier verbs are more likely to emerge on the level of individual languages (see again Appendix 15). On the whole, however, the proportions of different etymology types for satellite-framed and verb-framed languages are far more similar for manner verbs than they are for path verbs. In some sense, then, satellite-framed languages seem to simply have more manner verbs than verb-framed languages.

The Germanic and especially the Balto-Slavic languages retain far more second tier manner verbs with an Indo-European origin than the Romance languages, implying that manner of motion has been a salient concept throughout the history of these languages. It seems highly likely that their proximity to each other, as well as to other satellite-framed languages such as the Finno-Ugric languages, has helped to maintain satellite-framed encoding patterns and large manner verb lexicons. Evidence for this scenario is provided by Fanego (2012: 45ff), who points out that in Middle English, 51 out of 181 new manner of motion verbs were borrowed from Old Norse, Middle Dutch and Middle Low German, and 37 from French. In my own data, out of 20 Germanic borrowed manner verbs, 6 are from other Germanic languages, 12 are from French, and 2 are from Italian. Clearly, Germanic languages borrowed manner of motion verbs from each other, while the cultural importance of France during the Middle Ages clearly has left its mark on motion verb vocabulary as well. Future work should focus on the satellite-framed Finno-Ugric languages, to find out more about the proposed areality of satellite-framed motion encoding in North Europe, as well as the etymologies and rates of change of Finno-Ugric manner of motion verbs.

The development of the path verb lexicon and the increase of the use of the verb-framed construction in the Romance and Indo-Iranian languages is probably a single, unified development (see also section 5.5). What are now path verbs such as French *descendre* 'to descend' and *revenir* 'to return' where once satellite-framed constructions with preverbs *de-* and *re-* and verbs *scandere* 'to mount' and *venir* 'to come' in Latin. As preverb-verb combinations lexicalized into path verbs, the verb-framed construction became more frequent at the same time. The data presented in this dissertation suggest that as a consequence of this development, path verbs become more frequently used and more salient, leading to an increase in the development of path verbs at the subgroup level. In Romance, this takes place mostly through derivation of non-verbal elements, as in French *entrer*, Italian *entrare*, Portuguese *entrar* and Romanian *intra* 'to enter' from Latin *intrare* 'to enter', which is derived from the Latin preposition *inter* 'among, between'. In Indo-Iranian, this process takes place mostly through semantic shift, as in Hindi *baṛhnā* and Gujarati *vǎ̆dhvũ* 'to advance' from Sanskrit *várdhatē* 'to grow, increase'. The syntactic changes that resulted in the non-productivity of the spatial prefixes in the emergence of the modern Romance and Indo-Iranian languages seem to have triggered a chain reaction in their development of path verbs. It seems that this chain reaction continues in Romance languages, which have far more within language path verb etymologies than the Germanic and Balto-Slavic languages. It is likely that this process has been influenced by southern European verb-framed languages such as Albanian, Basque, Turkish, and Modern Greek, the latter of which is mixed but also has many lexicalized preverb-verb derivations (see section 4.2).

However, Indo-Iranian languages do not seem to have continued down the same path of path verb creation as the Romance languages, as Hindi, Nepali, and Persian hardly have any etymological within language developments. Although an overall correlation between the size of the path verb lexicon and the use of path-only and non-satellite-framed motion event encoding constructions was found for Indo-European (see section 5.4.3), this correlation does not seem to be present for Indo-Iranian. If the number of manner verbs and path verbs in Persian, Hindi, and Nepali (see Table 6.1) is considered, it seems clear that these three languages neither have a large manner verb lexicon nor a large path verb lexicon.

So although the Indo-Iranian languages have undergone the same changes with regard to the productivity of the preverb system as the Romance languages, the three Indo-Iranian languages in the current sample do not have an equally large path verb lexicon as the Romance languages. One reason why Hindi, Nepali, and Persian might not have path verb lexicons similar to Romance in size is that while they use path verbs to encode motion, they use deictic verbs frequently as well (see section 3.2). It might be the case that the frequent use of deictic verbs in these languages has inhibited the emergence of a large class of path verbs. In addition, the existence of many complex verbs path verbs in Persian makes it very difficult to see whether these verbs have emerged at the individual language level or the subgroup level, making comparisons across subgroups difficult. If the variability of motion verb lexicon size in the Indo-Iranian subgroup that was found in this study is a proxy for the diversity of motion verb lexicon sizes in the rest of the Indo-Iranian languages, clearly these languages are more diverse in their motion event encoding patterns as some of the other Indo-European subgroups (see chapter 3). More investigation into the large subgroup of Indo-Iranian languages is clearly needed.

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

Chapter 7: Summary, discussion and conclusions

7.1 Introduction

The introduction of this thesis started with the intriguing opposition between an English and a French headline, repeated below in (56) and (57) for convenience.

56) Daredevil Wallenda becomes first person to walk on tightrope across Niagara Falls

57) Le funambule Nik Wallenda traverse les chutes du Niagara sur un fil

In the introduction, it was explained that the English headline is an example of the satellite-framed construction, while the French headline is an example of what is traditionally called the verb-framed construction. Since English typically makes used of the satellite-framed construction, it has traditionally been classified as ‘satellite-framed’, and since French typically makes use of the verb-framed construction, it has traditionally been classified as ‘verb-framed’. These two constructions and the way in which they are used to classify languages have played a major role in this thesis.

In the chapters between the introduction and this conclusion a start has been made to explain the historical dimension of the difference between the English and the French headlines. Phylogenetic comparative methods were used to study change in motion event encoding in a sample of twenty Indo-European languages that includes English and French. Chapter 3 demonstrated that English and French, as well as the other eighteen languages included in the sample, do not simply make use of just these two motion event encoding constructions. The usage patterns of nine different motion event encoding constructions in twenty Indo-European languages demonstrated that languages could not simply be classified in terms of a dichotomy between a satellite-framed class and a verb-framed class. Chapter 4 presented evidence that Proto-Indo-European, the ancestor of all Indo-European languages, was neither radically satellite-framed nor radically verb-framed. Since the Indo-European languages split up into various subgroups, English has become more satellite-framed, and French has become more verb-framed. Chapter 5 demonstrated that the results of these changes in the usage of motion constructions are correlated with differences in the motion verb lexicon: satellite-framed languages have larger manner verb lexicons (including verbs such as *walk*, *dash*, and *swim*), while verb-framed languages have larger path verb lexicons (including verbs such as *cross*, *descend*, and *pass*). Chapter 6 focused on the etymologies of motion verbs. Here, it was shown that manner verbs evolve faster in branches leading to satellite-framed subgroups such as Germanic. The verb *walk* in (56) actually is an example of this process, as this verb originates in the Germanic subgroup. In addition, path verbs evolve faster in the branch leading to the Romance subgroup. Again, the verb *traverser* in (57) is an example of this development, as this verb and its Romance cognates such as Italian *traversare* are derived from the Latin preposition *trans*.

By focusing on these topics in the diachrony of motion event encoding, this thesis has investigated the evolutionary dynamics of motion event encoding in the Indo-European language family. This investigation is of interest for the broader study of (motion event) typology and historical linguistics for several reasons. First, there are relatively few studies of change in motion event encoding (exceptions are Acedo Matellán and Mateu 2008, 2010; Croft et al. 2010; Dufresne et al. 2003; Iacobini and Masini 2006, 2007; Kopecka 2006, 2009a; Kramer 1981; Masini 2005; Peyraube 2006; Slobin 2006; Talmy 2007; Vincent 1999, but none of these have the same scope as this thesis). This thesis is the first study that investigates motion event encoding in a sample of languages from the same language family. This allowed for the examination of the processes through which languages change syntactic and lexical aspects of motion encoding (chapters 4 and 6) as well as interactions between various aspects of motion encoding (chapter 5). Investigating change in motion event encoding allows for a better perspective on the diversity of contemporary motion construction usage, the dependencies of motion encoding on the linguistic tool-box of individual languages, and a better understanding of the linguistic encoding of motion in general.

Second, the use of phylogenetic comparative methods to analyze linguistic features is a relatively new development. Together with publications by other colleagues (Dunn et al. 2011; Levinson and Gray 2012; Levinson et al. 2011; Jordan 2011), this thesis exemplifies some of the first efforts that use these methods in the study of syntax and the lexicon. Phylogenetic comparative methods have provided immense opportunities for evolutionary biologists to study the evolution of biological species (Felsenstein 1985; Harvey and Pagel 1991; Weins 2000) and are increasingly being used by anthropologists as well (Currie 2013; Mace and Pagel 1994; Nunn 2011). Given the large body of knowledge on language families that linguists have gathered over the course of centuries, it makes sense to use this knowledge to help discover diachronic change in a range of linguistic features, including motion event encoding. This thesis includes some examples of the sorts of questions that can be answered with phylogenetic comparative methods, which hopefully may serve as an impetus for other analyses of this sort. In these next sections, the main findings and conclusions are summarized, and the implications for our understanding of historical linguistics and diachronic typology are considered.

7.2 Summary of the main findings

Chapter 2 introduced the parallel corpus from which the data that was employed in this thesis was taken, as well as the theoretical framework that was used to analyze the data. It also presented information on the aggregation of the datasets.

Chapter 3 presented an overview of the motion event encoding systems in the sample of twenty Indo-European languages (French, Italian, Portuguese, Romanian [Romance], Irish [Celtic], Dutch, English, German, Swedish [Germanic], Latvian, Lithuanian, Polish, Russian, Serbo-Croatian [Balto-Slavic], Hindi, Nepali, Persian [Indo-Iranian], Modern Greek [Hellenic], Albanian, and Armenian). Seven of these languages, namely Albanian, Armenian, Irish, Latvian, Lithuanian, Nepali and Romanian, had not been studied before in the Talmian motion event literature. It was shown that languages employ a range of different motion encoding constructions and that classifying languages in terms of the traditional Talmian dichotomy cannot capture this diversity. The traditional approach disregards the variation attested within the verb-framed class and within the satellite-framed class, and cannot account for languages that do not belong to either of these classes. The Neighbor-Net analysis and the multidimensional scaling (MDS) analysis that were conducted support this assessment of Talmy's (1991) dichotomy.

Chapter 4 explored historical patterns of motion event encoding construction usage by looking at qualitative data from the historical record of Indo-European and by providing quantitative phylogenetic comparative analyses. A maximum likelihood ancestral-state estimation analysis was carried out in order to investigate Proto-Indo-European motion event encoding as well as the changes that had taken place along the branches of the Indo-European language family that lead to the contemporary motion event encoding systems. The ancestral-state estimation analysis indicated that Proto-Indo-European motion event encoding was situated in the middle of the Talmian scale, with a slight tendency towards the satellite-framed end of the scale. This result was supported by qualitative data from studies of Classical Latin and Hittite, as well as information on idiomatic meanings of preverb-verb combinations in ancient Indo-European languages and etymologies of Romance path verbs. Over time, Germanic and Balto-Slavic have become more satellite-framed, whereas Romance and Indo-Iranian have become more verb-framed.

Chapter 5 reported the results of a phylogenetically controlled correlation analysis of the motion event encoding system and manner verb and path verb lexicon size. It gave a full overview of the manner verbs and path verbs that were attested in the sample of twenty Indo-European languages. Using Phylogenetic

Generalized Least Squares (PGLS) correlation analysis, it was shown that a positive relationship exists between the motion encoding system and the size of the manner verb and the path verb lexicon. This implied that languages that are closer to the satellite-framed side of the Talmian scale or that use the satellite-framed construction more often have a larger manner verb lexicon, and that languages that are closer to the verb-framed side of the Talmian scale or that use the path-only and the verb-framed constructions more often have a larger path verb lexicon. It may be speculated that languages that use the satellite-framed construction more often get larger manner verb lexicons over time because the verb-slot is 'free' to encode manner. For path verbs, it seems that the merging of Proto-Indo-European preverb + verb combinations resulted in a larger path verb lexicon as well as increased usage of the path-only construction and the verb-framed construction.

Chapter 6 explored the etymological origins of motion verbs and their evolutionary rates of change. The etymological origins of all manner verbs and path verbs that were attested in the sample were investigated using published materials and consultation with experts. Each etymology was classified in one of the following classes: semantic shift, borrowing, derivation from a non-verbal lexical item, development within subgroup (Romance, Germanic, Balto-Slavic, Indo-Iranian), development within Indo-European, derived modern prefix + verb combination, derived inherited preverb + verb, and complex verb. The rates of change of the size of the manner verb lexicon and the size of the path verb lexicon were investigated by branch length transformations. The results indicated that the merging of Proto-Indo-European preverb + verb combinations into path verbs is the etymological source of many path verbs, especially in the Romance languages. Higher rates of change of manner verb lexicon size were found in the branches leading to the Balto-Slavic subgroup, while higher rates of change of path verb lexicon size were found in the branch leading to the Romance subgroup. These higher rates of change relate to a growth of the manner verb lexicon for the Balto-Slavic subgroup, and a growth of the path verb lexicon for the Romance subgroup.

7.3 The evolutionary dynamics of motion event encoding

7.3.1 Diversity in motion event encoding

This dissertation took as its starting point the seminal work by Talmy (1985, 1991), who first proposed the opposition between satellite-framed and verb-framed languages. In the last two decades it has become clear that the original dichotomy is too restrictive (Naigles et al. 1998; Narasimhan 2003; Noonan 2003; Slobin 2004; Filipović 2007; Wälchli 2009; Beavers et al. 2010; Croft et al.

2010). The descriptions of motion event encoding attested in the twenty Indo-European languages included in chapter 3 of this thesis support this trend.

As it has become clear that a strict dichotomy is not the most suitable way to characterize and measure motion event encoding in a given language, a discussion on more suitable measures has ensued. In this thesis, this problem was tackled by incorporating information on the usage of each attested construction into a continuous measure of motion encoding with principal components analysis (chapter 2). This compiled the behavior of the different languages into a single measure (the first principal component that captured a large amount of variance) that was used in chapters 4 and 5. However, the future of motion event encoding research might benefit from not lumping everything together into one big measure, but splitting the domain of motion encoding into a set of smaller construction types. Narasimhan (2003) and Croft et al. (2010) have proposed such a construction-based outlook.

Narasimhan (2003) studies motion event encoding in Hindi. She shows that Hindi displays verb-framed characteristics, because combinations of manner verbs with path satellites (which are in fact both present in the lexicon) are illicit. Hindi's behavior leads her to conclude that lexical items that are similar cross-linguistically (manner verbs, path satellites) may nevertheless combine in different ways phrasally because of the existence of certain constructions. English has a specific construction, namely [subject - verb -oblique path phrase], that allows any manner verb, and even non-motion verbs such as *rattle* or *melt*, to be used in a satellite-framed construction. Hindi lacks this construction and is therefore severely limited in its use of the satellite-framed construction. The identification of such detailed construction types in different languages will help to identify the exact core of cross-linguistic variation.

Croft et al. (2010) complement this idea by showing that Talmy's (1985, 1991) original typological classification applies only to individual complex event classes within a language, and not to languages as a whole. The term 'individual complex event class' signifies individual combinations of verbs and path satellites, such as 'run out of (the room)' or 'float into (the air)'. Part of their work is based on that of Aske (1989), who showed that Spanish encoded boundary-crossing and non-boundary-crossing events differently. Croft et al.'s (2010) prediction is that complex events that denote more typical or natural combinations of manner and path ('run out of' being more typical than 'dance across') will be encoded by more highly integrated morpho-syntactic constructions (the satellite-framed construction being more integrated than the coordinated construction). The cross-linguistic study of such fine-grained construction types will help identify exactly which complex event types can be encoded by which syntactic motion construction. It may therefore find that Hindi can only use the satellite-framed construction for a very limited amount of highly typical or natural manner + path combinations. Future studies that take into account specific complex event types and constructions will provide a better understanding of the differences between complex event types and their typical encoding.

7.3.2 Diachronic change in the use of motion event encoding constructions

The shift in the study of motion event encoding from a classification of a languages into two or three classes to describing usage patterns of different constructions in a language also has consequences for the study of the diachrony of motion event encoding. Previous studies of change in motion event encoding (Acedo Matellán and Mateu 2008, 2010; Dufresne et al. 2003; Talmy 2007; Vincent 1999) have classified motion encoding in Proto-Indo-European and ancient Indo-European languages on the basis of the existence of certain motion constructions. However, this often does not match up with the usage-based studies that have been conducted since the work of Berman and Slobin (1994).

In chapter 4, motion event encoding was investigated from a usage-based perspective. Contemporary usage data was used to infer the placement of Proto-Indo-European on the continuous scale of motion event encoding systems that was created using principal components analysis. The choice was made to investigate motion event encoding in Proto-Indo-European and ancient Indo-European languages by looking at the usage rates of constructions, not their absence or presence, as was done by the studies cited above. After all, the uniformitarian hypothesis would be that if all attested Indo-European languages include a construction, Proto-Indo-European must as well. It is easy to say that the satellite-framed construction must be attested in Proto-Indo-European, as it is attested in all Indo-European languages. Chapter 4 showed that a usage-based measure is more detailed and informative than a classification on the basis of absence or presence of a construction. Future work on motion event encoding in ancient languages should focus more on quantitative investigations of the usage of different motion event encoding constructions.

The inference of the ancestral state of Proto-Indo-European as well as the changes on the branches leading from Proto-Indo-European to the contemporary languages was really an investigation into typological change of usage patterns of a set of comparable constructions. However, syntactic change played a role as well, as for instance the loss of case and the merging of preverb + verb combinations in most Indo-European languages influenced these typological changes. Even though the study of diachronic change in linguistics is vital to the discipline as a whole, there exists no generally recognized approach to syntactic change (Harris and Campbell 1995; Campbell 2004: 283-311) or to the young field of diachronic typology (Croft 2003: 232-279; Dediu 2011; Dunn et al. 2011; Levinson et al. 2011). Even so, there exist many studies of syntactic and typological change (DeLancey 1985, 1991; Durie 1988; van Gelderen (ed.) 2009;

van Gelderen 2011; Greenberg 1978, 1980; Matisoff 1976; Schwegler 1990), especially on the topic of grammaticalization (Campbell 2001; Heine et al. 1991; Heine and Reh 1984; Hopper and Traugott 1993; Janda 2001; Joseph 2001; Lehmann 1995). These studies have in common that they examine proto-language behavior, changes in syntactic and typological behavior, and the processes that underlie these changes.

However, none of these studies are quantitative or make use of statistical methods of the type that were used in chapter 4. In this respect, historical linguistics and diachronic typology lag behind evolutionary biologists, who have been using statistical methods to infer change on phylogenetic trees since Felsenstein (1985). Only in the last few years has linguistics started to use these types of method (Dunn et al. 2011; Jordan 2011). Phylogenetic comparative methods that model the evolution of linguistic features allow for the investigation of ancestral states and the rates of linguistic change along the branches of the phylogenetic tree. The use of these methods in linguistics enables linguists to generate quantitative statements about how likely it is that Austronesian sibling terminologies gain or lose a relative sex distinction (Jordan 2011) or how likely it is that Proto-Oceanic had serial verb constructions to express manner and result (Verkerk and Frostad 2013). Instead of arguments about whether certain grammaticalizations are uni-directional or not (see Janda 2001: 291-304), it is possible to generate quantitative statements how likely change in either direction is. These statements can be compared across language families, thereby quantifying which changes are universal and which are not (Fortunato and Jordan 2010). If linguists invest in the creation of evolutionary models of linguistic change that capture what decades of research have already discovered about linguistic change in syntax and typology, phylogenetic comparative methods have the prospect of becoming the general accepted approach to change in these fields that has been missing so far.

7.3.3 Correlations between syntactic and lexical features of motion event encoding

After Talmy's (1985, 1991) publications, it was mostly Dan Slobin who developed the field of motion event encoding research into what it is today. In a range of publications, Slobin developed the concept 'rhetorical style' (Slobin 1996a, 1996b, 1997, 2000, 2003, 2004). The rhetorical style of a language is the preferred way in which different languages linguistically encode motion. This does not only include the preferred construction to encode motion, i.e. the satellite-framed construction, or the verb-framed construction, or any other construction, but a set of preferences that relate to other aspects of motion event encoding. These have been summarized in Table 7.1.

***Table 7.1**: Summary of Slobin's (1996b, 1997, etc.) rhetorical style for motion*

Verb-framed languages	**Satellite-framed languages**
less ground elements per verb	more ground elements per verb
less path elements per trajectory of extended motion	more path elements per trajectory of extended motion
more 'bare' manner verbs without path-ground expressions	less 'bare' manner verbs without path-ground expressions
less frequent and less differentiated expression of manner of movement	more frequent and more differentiated expression of manner of movement
more static scene-setting	less static scene-setting
smaller and less diverse manner of motion verb lexicon	bigger and more diverse manner of motion verb lexicon
manner is only expressed when it's exceptional	manner is expressed as default

As is evident from Table 7.1, there is a range of motion encoding properties that can be associated with the preference for the verb-framed construction or the satellite-framed construction, respectively. Some of these refer to syntactic aspects of motion event encoding, such as the expression of ground elements or path satellites. Others refer to lexical aspects of motion event encoding, such as the size and diversity of the manner of motion verb lexicon and the amount of static scene-setting. Other hypotheses regarding motion encoding properties that are associated with construction choice have been made by Cifuentes Férez (2010), Matsumoto (2003), Narasimhan (2003), and Özçalışkan (2004).

Chapter 5 was the first statistical investigation of these hypothesized correlations between different aspects of motion event encoding. It showed that there is some evidence that languages that use the satellite-framed construction more often have a larger manner verb lexicon and that languages that use the verb-framed construction and the path-only construction more often have a larger path verb lexicon. The results presented in this chapter are the first statistical evidence that supports Slobin's notion of 'rhetorical style'. The Phylogenetic Generalized Least Squares (PGLS) correlation analyses conducted in chapter 5 could be used to investigate any of the correlations between the use of certain motion constructions and features commonly associated with the use of those motion constructions.

The results presented in chapter 5 also make clear that motion event encoding does not take place in a vacuum. Motion event encoding is dependent on the lexical items and the constructions that are available within the language (Beavers et al. 2010). In addition, it is likely to be in sync with 'normal' construction types within a language. For instance, is a high amount of path satellites per trajectory in motion also found in other domains, such as perception (i.e., 'Jane glanced down from her study into the hallway where Carol was standing')? Is a high use of manner adverbials matched in other domains, such as communication (i.e., 'Simon growled menacingly that John should leave right now')? Correlations like this will help with discovering the links between different domains and whether we can speak of a 'rhetorical style' of a whole language, not just for motion but for all semantic domains. With samples that include more languages, it also becomes possible to investigate which element changes first, i.e. whether or not the use of motion constructions changes first, giving rise to less static scene-setting or a higher amount of path satellites per trajectory, or vice versa.

The next step forward for the cross-linguistic study of motion event encoding is to start using statistics to validate its claims about cross-linguistic differences and similarities. One option would be to conduct similar studies in different language families, so that the results may be compared. In this way, we can discover whether the same correlations hold up in every language family, or whether the relationships found here are unique. A sampling approach rather than a phylogenetic approach would also be useful to discover cross-linguistic dependencies between motion encoding properties. If we want to have a better understanding of how motion event encoding works and how different elements – the lexicon, semantics, different types of syntactic constructions – are connected, larger cross-linguistic studies are called for.

7.3.4 Lexical evolution

Centuries of linguistic research have taught linguists much about historical phonology, sound changes, and cognacy (Campbell 2004). But the evolution of the lexicon has been considered only in limited ways outside the study of genealogical relationships. The most elaborated works on this are Algeo (1980) and Cannon (1978), as well as work on the growth of semantic subdomains such as color terminology (Berlin and Kay 1969; Kay and Maffi 1999), folk biology (Brown et al. 1976, 1986), cardinal direction terms (Brown 1983), and body part terminology (Brown 1976). There certainly is much work to do for lexical typology.

Recently, some investigations into the rates of lexical evolution have been made. These studies concern themselves mostly with rates of cognate replacement and not with lexical evolution in terms of semantic change. Various factors that influence rates of lexical evolution have been discovered, such as frequency of use (Calude and Pagel 2011; Pagel et al. 2007), complexity (Roberts 2009), and the emergence of new languages (Atkinson et al. 2008). However, these studies look at the lexicon as a whole and discover variability in rates of lexical change on a word-by-word basis.

In contrast, chapter 6 reported results of an investigation of rates of change of manner verbs and path verbs, two classes of motion verbs that have different sizes in different languages depending on the motion constructions used by those languages as reported in chapter 5. The results indicated that manner verbs and path verbs have different rates of change in different branches of the Indo-European tree, i.e. manner verb lexicon size evolves faster in the branches leading to Balto-Slavic, while path verb lexicon size evolves faster in the branch leading to Romance. Whether these differences between rates of change of manner verb lexicon size and path verb lexicon size are actually indicative of processes in language that shape the lexicon can only be determined if future studies address rates of change in other verb classes. In addition, there are many more different lexical classes that relate to motion event encoding that can be investigated in this way: path satellites and manner adverbials, for instance.

Another topic for future study of lexical evolution is change in the structure of semantic domains, such as the division of manner verbs between first-tier and second-tier manner verbs. This does not only include large open classes such as motion verbs, but especially also closed classes such as adpositions, pronouns, and numerals. For instance, how do oppositions in the adpositional system (in – out; up – down; etc.) emerge and change over time (Levinson and Meira 2003; Majid et al. 2011)? How do pronoun systems expand and shrink? How do numeral systems change from having a quinary base to a decimal base? These are all interesting questions that lexical typology has hardly touched upon. Investigations into rates of change as well as change of the content, size, and structure of different semantic subclasses could benefit from an evolutionary approach that models change in cross-linguistic samples.

7.4 Phylogenetic comparative linguistics

From the results presented in this thesis, we can sketch a picture of Proto-Indo-European motion event encoding, as well as change in motion event encoding throughout the Indo-European language family. Given the large variability in the use of different motion event encoding constructions in contemporary languages, it is likely that Proto-Indo-European made use of a set of different motion constructions as well. The estimated use of the different motion construction in Proto-Indo-European indicates a mixed typological type, somewhat like Modern Greek, which is neither radically satellite-framed nor radically verb-framed. At least two classes of motion verbs, manner verbs and path verbs, have been evolving along with the changing usage of motion constructions, moving and changing on the branches of the Indo-European tree that lead to the different contemporary languages.

Future work is required to understand whether the patterns that were presented in this dissertation are unique for the Indo-European language family or apply more generally. Wälchli (2009: 215) suggests that the amount of cross-linguistic variability found in the Indo-European languages is not uncommon. The idea that motion event encoding is stable seems to be a misrepresentation induced by Talmy's (2007) classification of language families rather than languages. Some of the patterns reported in this thesis may be driven by the opposition between Germanic (plus similar Balto-Slavic) and Romance, the two linguistic subgroups that Talmy (1985) originally started out with in his comparison of English and Spanish. Although no evidence for areal influences was found in chapter 4, the question remains whether the variability of motion event encoding in the Indo-European languages is caused by the existence of a satellite-framed linguistic area in North, Central and East Europe and the Caucasus as proposed by Wälchli (2009). The Uralic language family would be the first obvious candidate for further investigation, although the three Uralic languages in which the parallel texts are most likely to be retrievable (Finnish, Estonian, and Hungarian), are not representative for Uralic motion event encoding (Wälchli 2009: 215).

Diachronic change in the encoding of motion in Indo-European was analyzed with the use of phylogenetic comparative methods. By testing hypotheses on a large sample of phylogenetic trees, any dependency on a single tree that might otherwise have been present was removed. The analyses to test for phylogenetic signal that were used in chapters 4 and 5 clearly indicated that the use of methods that accounted for genealogical dependencies was necessary. Not all linguistic features necessarily have a phylogenetic history that has influenced their behavior. However, the analysis of many linguistic features requires the use of phylogenetic comparative methods that take such histories into account. Likewise, the tree model might not always be appropriate for all linguistic features, but the analyses conducted in this dissertation indicate that it was appropriate for the current dataset. In the future, network models might be developed for the comparative study of linguistic features (Nakhleh et al. 2005a; Nelson-Sathi et al. 2010). These would allow for analysis that takes into account phylogenetic as well as geographical dependencies at the same time.

However, perhaps a more important development would be the inclusion of linguistic meta-information into the phylogenetic comparative models, something that has not been done in this thesis. It would be extremely useful if certain pieces of highly specific linguistic information could be included in the priors of the phylogenetic comparative analysis. For the ancestral-state estimation of motion construction usage in Proto-Indo-European, for instance, it would have been useful to for instance include data on the unidirectional process of preverb-verb lexicalizations. Phylogenetic comparative analysis that is able to incorporate relevant information about processes of change that we already know about from etymologies and ancient languages provides the most optimal combination of quantitative and qualitative analysis to study language change.

I hope to have shown in this thesis that the application of phylogenetic comparative methods enables the investigation of diachronic questions in a practical, useful and statistically informed way. As was explained in section 1.3.1, the application of phylogenetic comparative methods is not just a way to deal with Galton's problem. They can be used to infer ancestral states and rates of linguistic change, two aspects of language evolution that otherwise would be more difficult to investigate. They allow for the investigation of the directionality and universality of the pathways of change studied by historical linguists and diachronic typologists alike (DeLancey 1985, 1991; Dunn et al. 2011; Durie 1988; Greenberg 1978, 1980; Levinson and Gray 2012; Matisoff 1976).

Given the large body of knowledge that historical linguists have already gathered on within language-family phylogenetic relations it seems appropriate to focus on comparative studies of language families, rather than eliminating genealogical relationships from the equation by looking for worldwide samples of unrelated languages. This is as true for historical linguistics and diachronic typology as it is for evolutionary biology and comparative anthropology (Dunn et al. 2011; Levinson and Gray 2012; Levinson et al. 2011). If this study had been conducted on a worldwide sample of unrelated languages (which would probably have been impossible to do using parallel texts, as their availability is heavily biased towards European languages), information would have been gathered on twenty unrelated motion event encoding systems. It would have been quite difficult to say anything about why those motion event encoding systems are the way they are, and there would have been no diachronic dimension that allows for the investigation of a Proto-system or rates of change. This is a problem in typological studies in general, as has been observed by Stassen (1997: 492ff) and Croft (2003: 250). Diachronic explanations illuminate the how and why of typological generalizations, and therefore we need adequate methods to generate these explanations. This thesis has demonstrated that phylogenetic comparative methods allow for a historically informed investigation of typological questions. There is no reason why typology should not embrace these methods critically but wholeheartedly in order to account for genealogical and (in the near future, using rooted phylogenetic networks) geographical dependencies. Ultimately, this approach will lead towards a better understanding of the evolution of the diversity of human languages.

References

Acedo Matellán, Victor & Jaume Mateu. (2008). *The path from satellite-framed Indo-European to verb-framed Romance: A lexical-syntactic account*. Paper presented at the X Diachronic Generative Syntax Conference.

Acedo Matellán, Victor & Jaume Mateu. (2010). From satellite-framed Latin to verb-framed Romance: A syntactic approach. *GGT*.

Adams, James Noel. (2003). *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Algeo, John. (1980). Where do all the new words come from? *American Speech*, 55(4), 264-277.

Ameka, Felix K. & James Essegbey. (2004). Serialising languages: Satellite-framed, verb-framed or neither. In Larry Hyman & Ian Maddieson (eds.), *African comparative and historical linguistics: Proceedings of the Thirty-second Annual Conference on African Linguistics*. Lawrenceville, NJ: Africa World Press.

Andriōtēs, Nikolaos Pantelē. (1951). *Etumologiko lexiko tēs koinēs neoellēnikēs*. Athēna: Institut Français d'Athènes.

Aske, Jon. (1989). Path predicates in English and Spanish: A closer look. *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 1-14).

Askedal, John Ole. (2006). Some typological differences between the Modern Germanic languages in a historical and geographical perspective. *Tidsskrift for Sprogforskning* 4: 187-207.

Atkinson, Quentin D. & Russell D. Gray. (2006). How old is the Indo-European language family? Illumination or more moths to the flame? In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages* (pp. 91-109). Cambridge: McDonald Institute for Archaeological Research.

Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon J. Greenhill & Mark Pagel. (2008). Languages evolve in punctuational bursts. *Science* 319: 588.

Atkinson, Quentin D., Geoff Nicholls, David Welch & Russell D. Gray. (2005). From words to dates: Water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103(2): 193-219.

Baayen, R. Harald. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistics* 1: 16-34.

Baicchi, Annalisa. (2005). Translating phrasal combinations across the typological divide. In M. Bertuccelli Papi (ed.), *Studies in the semantics of lexical combinatory patterns* (pp. 487-519). Pisa: Pisa University Press.

Bańkowski, Andrzej. (2000-present). *Etymologiczny słownik języka polskiego*. Warszawa: Wydawnictwo naukowe PWN.

Barkman, Todd J., Mika Bendiksby, Seok-Hong Lim, Kamarudin Mat Salleh, Jamili Nais, Domingo Madulid & Trond Schumacher. (2008). Accelerated rates of floral evolution at the upper size limit for flowers. *Current Biology* 18: 1508-1513.

Bartholomae, Christian. (1904). *Altiranisches Wörterbuch*. Strassburg: Trübner.

Bateman, Richard, Ives Goddard, Richard O'Grady, V. A. Funk, Rich Mooi, W. John Kress & Peter Cannell. (1990). Speaking of forked tongues. *Current Anthropology* 31(1): 1-24.

Battisti, Carlo & Giovanni Alessio. (1950-1957). *Dizionario etimologico italiano*. Firenze: Barbèra.

Beavers, John, Beth Levin & Siao Wei Tham. (2010). The typology of motion expressions revisited. *Journal of Linguistics* 46: 331-377.

Beekes, Robert S. P. (2010). *Etymological dictionary of Greek*. Leiden: Brill.

Beekes, Robert S. P. (2011). *Comparative Indo-European linguistics: An introduction*. Amsterdam: John Benjamins.

Berlin, Brent & Paul Kay. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

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

Berthele, Raphael. (2004). The typology of motion and posture verbs: A variationist account. In Bernd Kortmann (ed.), *Dialectology meets typology: dialect grammar from a cross-linguistic perspective* (pp. 93-126). Berlin: Mouton de Gruyter.

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

Bickel, Balthasar. (2008). A refined sampling procedure for genealogical control. *STUF - Sprachtypologie und Universalienforschung* 61(3): 221-233.

Bielec, Dana. (1998). *Polish: An essential grammar*: London: Routledge.

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

Bloch, Oscar & Walther von Wartburg. (1932). *Dictionnaire étymologique de la langue française*. Paris: Presses Universitaires de France.

Borgerhoff-Mulder, Monique. (2001). Using phylogenetically based comparative methods in anthropology: More questions than answers. *Evolutionary Anthropology* 10: 99-111.

Borgerhoff-Mulder, Monique, Charles L. Nunn & Mary C. Towner. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology* 15: 52-64.

Boryś, Wiesław Stefan. (2005). *Słownik etymologiczny języka polskiego*. Kraków: Wydawn. Literackie.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337: 957-960.

Boyd, Robert, Monique Bogerhoff-Mulder, William H. Durham & Peter J. Richerson. (1997). Are cultural phylogenies possible? In Peter Weingart, Sandra D. Mitchell, Peter J. Richerson & Sabine Maasen (eds.), *Human By Nature: Between Biology and the Social Sciences* (pp. 355-387). Mahwah: Lawrence Erlbaum Associates.

Brachet, Auguste (1873). *An etymological dictionary of the French language.* Oxford: Clarendon Press.

Brosch, Cyril. (2014). *Untersuchungen zu hethitischen Raumgrammatik*. Berlin: de Gruyter.

Brown, Cecil H. (1976). General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist* 3(3): 400-424.

Brown, Cecil H. (1983). Where do cardinal direction terms come from? *Anthropological Linguistics* 25(2): 121-161.

Brown, Cecil H., E. N. Anderson Jr., Brent Berlin, James S. Boster, Thilo C. Schadeberg & Leontine E. Visser. (1986). The growth of ethnobiological nomenclature. *Current Anthropology* 27(1): 1-19.

Brown, Cecil H., Eric W. Holman, Søren Wichmann & Viveka Velupillai. (2008). Automated classification of the world's languages: A description of the method and preliminary results. *STUF - Sprachtypologie und Universalienforschung* 61(4): 285-308.

Brown, Cecil H., John Kolar, Barbara J. Torrey, Tipawan Truong-Quang & Philip Volkman. (1976). Some general principles of biological and non-biological folk classification. *American Ethnologist* 3(1): 73-85.

Brown, Penelope. (2004). Position and motion in Tzeltal frog stories: The acquisition of narrative style. In Sven Strömqvist & Ludo Verhoven (eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 37-57). Mahwah: Lawrence Erlbaum Associates.

Brucale, Luisa. (2011). *Manner of motion verbs in Latin*. Paper presented at the Historical-comparative linguistics in the 21st century conference.

Brucale, Luisa, Claudio Iacobini & Egle Mocciaro. (2011). *Typological change in the expression of motion events from Latin to Romance languages*. Paper presented at the 44th Annual Meeting of the Societas Linguistica Europaea.

Brückner, Aleksander. (1957). *Słownik etymologiczny języka polskiego*. Warszawa: Wiedza Powszechna.

Bryant, David and Vincent Moulton. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2): 255-265.

Buck, Carl Darling. (1949). *A dictionary of selected synonyms in the principal Indo-European languages.* Chicago: University of Chicago Press.

Burnei, Alexandra. (1996). *Dicționarul explicativ al limbii române*. București: Univers Enciclopedic.

Calude, Andreea S. & Mark Pagel. (2011). How do we use language? Shared patterns in the frequency of word use across 17 world languages. *Philosophical Transactions of the Royal Society (B)* 366: 1101-1107.

Camaj, Martin. (1984). *Albanian grammar: with exercises, chrestomathy and glossaries* (Leonard Fox, trans.). Wiesbaden: Harrassowitz.

Campbell, Lyle. (2001). What's wrong with grammaticalization. *Language Sciences* 23(2-3): 113-161.

Campbell, Lyle. (2004). *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Campbell, Lyle. (2008). What can we learn about the earliest human language by comparing languages known today? In Bernard Laks, Serge Cleuziou, Jean-Paul Demoule & Pierre Encrevé (eds.), *Origins and evolution of language: Approaches, models, paradigms* (pp. 79-111). London: Equinox Publishing Limited.

Cannon, Garland. (1978). Statistical etymologies of new words in American English. *Journal of English Linguistics* 12: 12-18.

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

Chen, Liang & Jiansheng Guo. (2010). From language structures to language use: A case from Mandarin motion expression classification. *Chinese Language and Discourse* 1(1): 31-65.

Chen, Shichao, Dong-Kap Kim, Mark W. Chase & Joo-Hwan Kim. (2013). Networks in a large-scale phylogenetic analysis: Reconstructing evolutionary history of Asparagales (Lilianae) based on four plastid genes. *PLoS ONE*, 8(3): e59472.

Cheung, Johnny. (2007). *Etymological dictionary of the Iranian verb*. Leiden: Brill.

Cheverud, James M., Malcolm M. Dow & Walter Leutenegger. (1985). The quantitative assessment of phylogenetic constraints in comparative analyses: Sexual dimorphism in body weight among primates. *Evolution* 39(6): 1335-1351.

Choi, Soonja & Melissa Bowerman. (1992). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. In Beth Levin & Steven Pinker (eds.), *Lexical & conceptual semantics* (pp. 83-121). Cambridge: Blackwell.

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

Cioranescu, Alejandro. (1966). *Diccionario etimológico Rumano*. Tenerife: Universidad La Laguna, Biblioteca Filológica.

Clackson, James. (1994). *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.

Clackson, James. (2007). *Indo-European linguistics: An introduction*. Cambridge: Cambridge University Press.

Claidière, Nicolas & Jean-Baptiste André. (2012). The transmission of genes and culture: A questionable analogy. *Evolutionary Biology* 39(1): 12-24.

Collar, David C., Brian C. O'Meara, Peter C. Wainwright & Thomas J. Near. (2009). Piscivory limits diversification of feeding morphology in Centrarchid fishes. *Evolution* 63(6): 1557-1573.

Collard, Mark, Stephen J. Shennan & Jamshid J. Tehrani. (2006). Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evolution and Human Behavior* 27: 169-184.

Comrie, Bernard, Martin Haspelmath & Balthasar Bickel. (2004). *Leipzig Glossing Rules*. http://www.eva.mpg.de/lingua/resources/glossing-rules.php

Crapo, Richley. (1970). The origins of directional adverbs in Uto-Aztecan languages. *International Journal of American Linguistics* 36(3): 181-189.

Croft, William. (2000). *Explaining language change: An evolutionary approach*. Harlow: Pearson Education Limited.

Croft, William. (2003). *Typology and universals*. Cambridge: Cambridge University Press.

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

Currie, Thomas E. (2013). Cultural evolution branches out: The phylogenetic approach in cross-cultural research. *Cross-Cultural Research* 47(2): 102-130.

Currie, Thomas E., Simon J. Greenhill, Russell D. Gray, Toshikazu Hasegawa & Ruth Mace. (2010). Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature* 467: 801-804.

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

Danesi, Serena. (2013). Particle-verb constructions in Vedic: The case of ápa. *Studi e Saggi Linguistici* 51(2): 57-100.

Deaner, Robert O. & Charles L. Nunn. (1999). How quickly do brains catch up with bodies? A comparative method for detecting evolutionary lag. *Proceedings of the Royal Society London B* 266(1420): 687-694.

Dediu, Dan. (2011). A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society B* 278: 474-479.

DeLancy, Scott. (1985). The analysis-synthesis-lexis cycle in Tibeto-Burman: A case study in motivated change. In John Haiman (ed.), *Iconicity in syntax* (pp. 367-389). Amsterdam: John Benjamins.

DeLancy, Scott. (1991). The origins of verb serialization in Modern Tibetan. *Studies in Language* 15(1): 1-23.

Delbrück, Berthold. (1888). *Altindische syntax*. Darmstadt: Wissenschaftliche Buchgesellschaft.

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

Demiraj, Bardhyl. (1997). *Albanische Etymologien*. Amsterdam: Rodopi.

Derksen, Rick. (2008). *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.

Dickey, Stephen M. (2010). Common Slavic "indeterminate" verbs of motion were really manner-of-motion verbs. In Victoria Hasko and Renee Perelmutter (eds.), *New approaches to Slavic verbs of motion* (pp. 67-109). Amsterdam: John Benjamins.

Dizionari sapere.it. (2012). *Dizionari di Lingua Italiana*. Novara: De Agostini Editore.

Długosz-Kurczabowa, Krystyna. (2003). *Nowy słownik etymologiczny języka polskiego*. Warszawa: PWN.

Dow, Malcolm M. (1991). Statistical inference in comparative research: New directions. *Behavior Science Research* 25: 235-258.

Drummond, Alexei J., Simon Y. W. Ho, Matthew J. Phillips & Andrew Rambaut. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4(5): e88.

Drummond, Alexei J., Marc A. Suchard, Dong Xie & Andrew Rambaut. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29(8): 1969-1973.

Dryer, Matthew S. (1989). Large linguistic areas and language sampling. *Studies in Language* 13: 257-292.

Dryer, Matthew S. (1992). The Greenbergian word order correlations. *Language* 68(1): 81-138.

Dufresne, Monique, Fernande Dupuis & Mireille Tremblay. (2003). Preverbs and particles in Old French. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology* (pp. 30-60): Kluwer.

Dunn, Michael. (2009). Contact and phylogeny in Island Melanesia. *Lingua* 119(11): 1664-1678.

Dunn, Michael, Kate Bellamy, Chundra Cathcart, Andrew Garrett, Julia Ludewig & Cornelia van Scherpenberg. (to appear). IELex: an Indo-European lexical database for phylogenetic analysis. *PloS One*.

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

Dunn, Michael, Angela Terrill, Ger Reesink, R. A. Foley & Stephen C. Levinson. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science* 309(5743): 2072-2075.

Durie, Mark. (1988). Verb serialization and "verbal-prepositions" in Oceanic languages. *Oceanic Linguistics* 27: 1-23.

Dyen, Isidore, Joseph B. Kruskal & Paul Black. (1992). An Indo-European classification: A lexicostatistical experiment. *Transactions of the American Philological Society* 82(1): iii-132.

Eastman, Jonathan M., Michael E. Alfaro, Paul Joyce, Andrew L. Hipp & Luke J. Harmon. (2011). A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65(12): 3578-3589.

Ember, Melvin. (1978). Size of color lexicon: Interaction of cultural and biological factors. *American Anthropologist* 80(2): 364-367.

Engberg-Pedersen, Elisabeth & Frederikke Blytmann Trondhjem. (2004). Focus on action in motion descriptions. In Sven Strömqvist & Ludo Verhoven (eds.), *Relating events in narrative volume 2: Typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Associates.

Ethnologue: Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). (2013). *Ethnologue: Languages of the World, Seventeenth edition*. Dallas: SIL International. Online version: http://www.ethnologue.com

Evans, Nicholas & David Wilkins. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76: 546-592.

Falk, Hjalmar Sejersted & Alf Torp. (1910-1911). *Norwegisch-dänisches etymologisches Wörterbuch*. Heidelberg: Winter.

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.

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

Felsenstein, Joseph. (1981). Evolutionary trees from DNA sequences: A Maximum Likelihood approach. *Journal of Molecular Evolution* 17: 368-376.

Felsenstein, Joseph. (1982). Numerical methods for inferring evolutionary trees. *The Quarterly Review of Biology* 57(4): 379-404.

Felsenstein, Joseph. (1985). Phylogenies and the comparative method. *The American Naturalist* 125(1): 1-15.

Ferrari, Giacomo & Monica Mosca. (2010). Some constructions of path: From Italian to some classical languages. In Giovanna Marotta, Alessandro Lenci, Linda Meini & Francesco Rovai (eds.), *Space in language: Proceedings of the Pisa International Conference* (pp. 317-338). Pisa: Edizioni ETS.

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

Fillmore, Charles J. (1982). Towards a descriptive framework for spatial deixis. In Robert J. Jarvella & Wolfgang Klein (eds.), *Speech, place, and action: Studies in deixis and related topics* (pp. 31-59). Chichester: John Wiley & Sons.

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

Folli, Raffaella & Gillian Ramchand. (2001). Getting results: Motion constructions in Italian and Scottish Gaelic. In K. Megerdoomian & L. A. Bar-el (Eds.), *Proceedings of the 20th West Coast Conference on Formal Linguistics* (pp. 101-114). Somerville, MA: Cascadilla Press.

Folli, Raffaella & Gillian Ramchand. (2005). Prepositions and results in Italian and English: An analysis from event decomposition. In Henk J. Verkuyl, Henriette de Swart & Angeliek van Hout (eds.), *Perspectives on aspect* (pp. 1-20). Dordrecht: Kluwer.

Fong, Vivienne & Christine Poulin. (1998). Locating linguistic variation in semantic templates. In Jean-Pierre Koening (ed.), *Discourse and cognition: Bridging the gap* (pp. 29-39). Cambridge: Cambridge University Press.

Fortunato, Laura & Fiona M. Jordan. (2010). Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies. *Philosophical Transactions of the Royal Society B* 365: 3913-3922.

Fraenkel, Ernst. (1962-1965). *Litauisches etymologisches Wörterbuch*. Heidelberg: Winter.

Franck, Johannes. (1976). *Franck's Etymologisch woordenboek der Nederlandsche taal*. 's-Gravenhage: Nijhoff.

Frawley, William. (1992). *Linguistic semantics*. Hillsdale: Lawrence Erlbaum Associates.

Freckleton, Rob P., Paul H. Harvey & Mark Pagel. (2002). Phylogenetic analysis and comparative data: A test and review of evidence. *The American Naturalist* 160(6): 712-726.

Freedman, David A. & William S.-Y. Wang. (1996). Language polygenesis: A probabilistic model. *Anthropological Sciences* 104(2): 131-138.
Frisk, Hjalmar. (1954-1972). *Griechisches etymologisches Wörterbuch*. Heidelberg: Winter.
Garland, Theodore & Anthony R. Ives. (2000). Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist* 155(3): 346-364.
Garrett, Andrew. (2006). Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages* (pp. 139-151). Cambridge: McDonald Institute for Archaeological Research.
Gathercole, Virginia C. (1978). Towards a universal of deictic verbs of motion. *Kansas Working Papers in Linguistics* 3: 72-98.
van Gelderen, Elly. (ed.). (2009). *Cyclical change*. Amsterdam: John Benjamins
van Gelderen, Elly. (2011). *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.
Goddard, Cliff. (2001). Lexico-semantic universals: A critical overview. *Linguistic Typology* 5(1): 1-65.
Goetz, Virginia A. (2006). *The development of Proto-Indo-European local adverbs into Germanic prepositions and verbal elements.* PhD thesis, University of Connecticut.
Goldberg, Adele. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Goldberg, Adele & Ray Jackendoff. (2004). The English resultative as a family of constructions. *Language* 80: 532-568.
Gray, Russell D. & Quentin D. Atkinson. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426: 435-439.
Gray, Russell D., David Bryant & Simon J. Greenhill. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society* 365: 3923-3933.
Gray, Russell D., Simon J. Greenhill & Robert M. Ross. (2007). The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory* 2(4): 360-375.
Green, Peter J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4): 711-732.
Greenberg, Joseph H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of grammar*. Cambridge: The MIT Press.

Greenberg, Joseph H. (1978). How does a language acquire gender markers? In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language* (pp. 47-82). Stanford: Stanford University Press.

Greenberg, Joseph H. (1980). Circumfixes and typological change. In Elizabeth C. Traugott, Rebecca Labrum & Susan Shepherd (eds.), *Papers from the 4th International Conference on Historical Linguistics* (pp. 233-242). Amsterdam: John Benjamins.

Greenberg, Marc L. (2010). PIE inheritance and word-formational innovation in Slavic motion verbs in -i-. In Victoria Hasko & Renee Perelmutter (eds.), *New approaches to Slavic verbs of motion* (pp. 111-121). Amsterdam: John Benjamins.

Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. (2009). Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B* 276: 2299-2306.

Grieve, Jack. (2014). A comparison of statistical methods for the aggregation of regional linguistic variation. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech* (pp. 53-88) . Berlin: Walter de Gruyter.

Guo, Jiansheng & Liang Chen. (2009). Learning to express motion in narratives by Mandarin-speaking children. In Jiansheng Guo, Elena Lieven, Susan Ervin-Tripp, Nancy Budwig, Keiko Nakamura & Şeyda Özçalışkan (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 193-208). New York: Psychology Press.

Haig, Geoffrey J. L., Stefan Schnell & Claudia Wegener. (2011). Comparing corpora from endangered language projects: Explorations in language typology based on original texts. In Geoffrey L. J. Haig, Nicole Nau, Stefan Schnell & Claudia Wegener (eds.), *Documenting endangered languages: Achievements and perspectives* (pp. 55-86). Berlin: Walter de Gruyter.

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

Harris, Alice C. & Lyle Campbell. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Harris, Edward F. & Howard L. Bailit. (1988). A principal components analysis of human odontometrics. *American Journal of Physical Anthropology* 75: 87-99.

Harvey, Paul H. & Mark Pagel. (1991). *The comparative method in evolutionary biology*. Oxford: Oxford University Press.

Haug, Dag T. T. & Marius Jøhndal. (2008). Creating a parallel treebank of the old Indo-European bible translations. In Khalid Choukri (ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC '08)*. European Language Resources Association (ELRA).

Haun, Daniel B. M., Fiona M. Jordan, Giorgio Vallortigara & Nicky S. Clayton. (2010). Origins of spatial, temporal and numerical cognition: Insights from comparative psychology. *Trends in Cognitive Sciences* 14: 552-560.

Heine, Bernd, Ulrike Claudi & Friederieke Hünnemeyer. (1991). *Grammaticalization: a conceptual framework*. Chicago: University of Chicago Press.

Heine, Bernd & Mechthild Reh. (1984). *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske Verlag.

Hellquist, Elof. (1922). *Svensk etymologisk ordbok*. Lund: C. W. K. Gleerups Förlag.

Hewson, John & Vit Bubenik. (2006). *From case to adposition: The development of configurational syntax in Indo-European languages*. Amsterdam: John Benjamins.

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

Hofmann, Johann B. & Anton Szantyr. (1965). *Lateinische Syntax und Stilistik*. München: C. H. Beck'sche Verlagsbuchhandlung.

Holden, Clare J. & Ruth Mace. (2003). Spread of cattle led to the loss of matrilineal descent in Africa: A coevolutionary analysis. *Proceedings of the Royal Society London* 270: 2425-2433.

Holton, David, Peter Mackridge & Irene Philippaki-Warbuton. (1997). *Greek: A comprehensive grammar of the modern language*. London: Routledge.

Hopper, Paul & Elizabeth Traugott. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Horn, Paul. (1893). *Grundriss der neupersischen Etymologie*. Strassburg: Trübner.

Hsiao, Hui-Chen Sabrina. (2009). Motion event descriptions and manner-of-motion verbs in Mandarin. PhD thesis, University at Buffalo, State University of New York.

Huang, Shuanfan & Michael Tanangkingsing. (2005). Reference to motion events in six western Austronesian languages: Toward a semantic typology. *Oceanic Linguistics* 44(2): 307-340.

Hübschman, Johann Heinrich. (1883). *Armenische Studien*. Leipzig: Breitkopf & Härtel.

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

Huffman, Stephen. (2013). *The Indo-European Language Family*. Language map made using the World Language Mapping System 16; available on http://www.worldgeodatasets.com

Huson, Daniel H. & David Bryant. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254-267.

Huson, Daniel H. & Celine Scornavacca. (2011). A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution* 3(1): 23-35.

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 & Francesco Rovai (eds.), *Space in language: Proceedings of the Pisa International Conference* (pp. 495-514). Pisa: Edizioni ETS.

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

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

Ibarretxe-Antuñano, Iraide. (2003). What translation tells us about motion: A contrastive study of typologically different languages. *International Journal of English Studies* 3(2): 151-176.

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

Ibarretxe-Antuñano, Iraide. (2009). Path salience in motion events. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 403-414). New York: Psychology Press.

Imbert, Caroline. (2010). Multiple preverbation in Homeric Greek: A typological insight. *CogniTextes* 4.

Jablonka, Eva & Marion J. Lamb. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge: The MIT Press.

Janda, Richard D. (2001). Beyond 'pathways' and 'unidirectionality': on the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* 23(2-3): 265-340.

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

Jordan, Fiona M. (2011). A phylogenetic analysis of the evolution of Austronesian sibling terminologies. *Human Biology* 83(2): 297-321.

Jordan, Fiona M., Russell D. Gray, Simon J. Greenhill & Ruth Mace. (2009). Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B* 276: 1957-1964.

Jordan, Fiona M. & Thomas E. Currie. (submitted). The effect of population size and density on rates of linguistic evolution. *Proceedings of the Royal Society B*.

Joseph, Brian D. (2001). Is there such a thing as 'grammaticalization'? *Language Sciences* 23(2-3): 163-186.

Jovanovic, Jelana & Aida Martinovic-Zic. (2004). Why manner matters: Contrasting English and Serbo-Croatian typology in motion description. In Carol Lynn Moder & Aida Martinovic-Zic (eds.), *Discourse across languages and cultures* (pp. 211-226). Amsterdam: John Benjamins.

Junghänel, Anja. (in preparation). *Coding motion events in Indo-European*.

Kachru, Yamuna. (2006). *Hindi*. Amsterdam: John Benjamins.

Karulis, Konstantīns. (1992). *Latviešu etimoloģijas vārdnīca*. Rīga: Avots.

Kass, Robert E. & Adrian E. Raftery. (1995). Bayes factors. *Journal of the Americal Statistical Association* 90(430): 773-795.

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

Kay, Paul & Luisa Maffi. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* 101(4): 743-760.

Kent, Roland G. (1950). *Old Persian*. New Haven: American Oriental Society.

Kluge, Friedrich. (1883). *Etymologisches Wörterbuch der deutschen Sprache*. Strassburg: Trübner.

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

Kopecka, Anetta. (2009a). Continuity and change in the representation of motion events in French. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 415-426). New York: Psychology Press.

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

Koptjevskaja-Tamm, Maria, Martine Vanhove & Peter Koch. (2007). Typological approaches to lexical semantics. *Linguistic Typology* 11: 159-185.

Koptjevskaja-Tamm, Maria, Dagmar Divjak & Ekaterina Rakhilina. (2010). Aquamotion verbs in Slavic and Germanic: A case study in lexical typology. In Victoria Hasko & Renee Perelmutter (eds.), *New approaches to Slavic verbs of motion* (pp. 315–341). Amsterdam: John Benjamins.

Kortlandt, Frederik. (1981). More evidence for Italo-Celtic. *Ériu* 32(1): 1-22.

Kramer, Johannes. (1981). Die Übernahme der deutschen und der niederländischen Konstruktion Verb + Verbzusatz durch die Nachbarsprachen. In Wolfgang Meid & Karin Heller (Eds.), *Sprachkontakt als Ursache von Veränderungen der Sprach- und Bewusstseinsstruktur: Eine Sammlung von Studien zur sprachlichen Interferenz* (pp. 129-140). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

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

de Lagarde, Paul. (1877). *Armenische Studien*. Göttingen: Dieterisch'sche Verlags-Buchhandlung.

Lambton, Ann K. S. (1957). *Persian grammar*. Cambridge: Cambridge University Press.

Lehmann, Christian. (1985). Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20(3): 303-318.

Lehmann, Christian. (1995). *Thoughts on grammaticalization: a programmatic sketch, vol. I*. München: LINCOM Europa.

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

Lehmann, Winifred P. (1986). *A gothic etymological dictionary*. Leiden: Brill.

Leumann, Manu. (1977). *Lateinische Laut- und Formenlehre*. München: C. H. Beck'sche Verlagsbuchhandlung.

Levin, Beth & Malka Rappaport Hovav. (1991). Wiping the slate clean: A lexical semantic exploration. *Cognition* 41: 123-151.

Levinson, Stephen C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Levinson, Stephen C. & Nicolas Burenhult. (2009). Semplates: A new concept in lexical semantics? *Language* 85(1): 153-174.

Levinson, Stephen C. & Russell D. Gray. (2012). Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences* 16(3): 167-173.

Levinson, Stephen C., Simon J. Greenhill, Russell D. Gray & Michael Dunn. (2011). Universal typological dependencies should be detectable in the history of language families. *Linguistic Typology* 15: 509-534.
Levinson, Stephen C. & Sérgio Meira. (2003). 'Natural concepts' in the spatial topological domain - adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79: 485-516.
Lindenfors, Patrik, Laila Fröberg & Charles L. Nunn. (2004). Females drive primate social evolution. *Proceedings of the Royal Society London B (Supplement)* 271: 101-103.
Lindenfors, Patrik, Liam J. Revell & Charles L. Nunn. (2010). Sexual dimorphism in primate aerobic capacity: A phylogenetic test. *Journal of Evolutionary Biology* 23(6): 1183-1194.
Losos, Jonathan B. (1990). Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: An evolutionary analysis. *Ecological Monographs* 60(3): 369-388.
Losos, Jonathan B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters* 11: 995-1007.
Luraghi, Silvia. (2010). The rise (and possible downfall) of configurationality. In Silvia Luraghi & Vit Bubenik (eds.), *A companion to historical linguistics* (pp. 212-229). London: Continuum Books.
MacBain, Alexander. (1911). *An etymological dictionary of the Gaelic language*. Surling: Eneas Mackay.
Macdonell, Arthur Anthony. (1929). *A practical Sanskrit dictionary with transliteration, accentuation, and etymological analysis throughout*. Londen: Oxford University Press.
Mace, Ruth & Mark Pagel. (1994). The comparative method in anthropology. *Current Anthropology* 35: 549-564.
Machado, José Pedro. (1952). *Dicionário etimológico da língua portuguesa*. Lisboa: Livros Horizonte.
Macrea, Dimitrie. (1958). *Dicţionarul limbii romîne moderne*. Bucureşti: Editura Academiei Republicii Populare Romîne.
Maddison, Wayne P. (1991). Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Zoology* 40(3): 304-314.
Mahootian, Shahrzad. (1997). *Persian*. London: Routledge.
Majid, Asifa, Melissa Bowerman, Sotaro Kita, Daniel B. M. Haun & Stephen C. Levinson. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences* 8(3): 108-114.
Majid, Asifa, Nicholas J. Enfield & Mirjam van Staden. (2006). Parts of the body: Cross-linguistic categorisation. *Language Sciences* 28.

Majid, Asifa, Fiona Jordan, & Michael Dunn. (2011). *The evolution of semantic systems manual*. Max Plank Institute for Psycholinguistics, Nijmegen.

Martins, Emilia P. & Thomas F. Hansen. (1997). Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist* 149(4): 646-667.

Martins, Emilia P. & Jenny Lamont. (1998). Estimating ancestral states of a communicative display: A comparative study of Cyclura rock iguanas. *Animal Behavior* 55: 1685-1706.

Martirosyan, Hrach K. (2010). *Etymological dictionary of the Armenian inherited lexicon*. Leiden: Brill.

Masini, Francesca. (2005). Multi-word expressions between syntax and the lexicon: The case of Italian verb-particle constructions. *SKY Journal of Linguistics* 18: 145-173.

Matasović, Ranko. (2009). *Etymological dictionary of Proto-Celtic*. Leiden: Brill.

Matisoff, James A. (1976). Lahu causative constructions: Case hierarchies and the morphology/syntax cycle in a Tibeto-Burman perspective. In Masayoshi Shibatani (ed.), *The grammar of causative constructions* (pp. 413-442). New York: Academic Press.

Matsumoto, Yo. (1997). Kuukan-idoo no gengo-hyoogen to sono kakuchoo [Linguistic expressions of spatial motion and their extensions]. In Shigenoti Tanaka & Yo Matsumoto (eds.), *Kuukan to idoo no hyoogen* [Expressions of space and motion]. Tokyo: Kenkyuusha.

Matsumoto, Yo. (2003). Typologies of lexicalization patterns and event integration: Clarifications and reformulations. In Shuji Chiba et al. (ed.), *Empirical and theoretical investigations into language: A festschrift for Masaru Kajita* (pp. 403-418). Tokyo: Kaitakusha.

Mayrhofer, Manfred. (1956-1980). *Kurzgefaßtes etymologisches Wörterbuch des Altindischen*. Heidelberg: Winter.

Meyer, Gustav. (1891). *Etymologisches Wörterbuch der albanesischen Sprache*. Strassburg: Trübner.

Miklosich, Franz. (1868). *Vergleichende Grammatik der slavischen Sprachen*. Wien: W. Braumüller.

Moore, John H. (1994). Putting anthropology back together again: The ethnogenetic critique of cladistic theory. *American Anthropologist* 96(4): 925-948.

Mpampiniōtēs, Geōrgios D. (2008). *Lexiko tēs neas Ellēnikēs glōssas*. Athēna: Kentro Lexikologias.

Mufwene, Salikoko S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.

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

Nakhleh, Luay, Don Ringe & Tandy Warnow. (2005a). Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81(2): 382-420.

Nakhleh, Luay, Tandy Warnow, Don Ringe & S. N. Evans. (2005b). A comparison of phylogenetic reconstruction methods on an IE dataset. *The Transactions of the Philological Society* 3(2): 171-192.

Narasimhan, Bhuvana. (1998). *The encoding of complex events in Hindi and English.* PhD thesis, Boston University.

Narasimhan, Bhuvana. (2003). Encoding complex events in Hindi and English. *Proceedings of the 1999 Conference on Perspectives on Argument Structure*. Austin: Texas Linguistic Forum.

Nelson-Sathi, Shijulal, Johann-Mattis List, Hans Geisler, Heiner Fangerau, Russell D. Gray, William Martin & Tal Dagan. (2010). Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proceedings of the Royal Society B* 278: 1794-1803.

Nesset, Tore. (2009). Metaphorical walking: Russian idti as a generalized motion verb. In Victoria Hasko & Renee Perelmutter (eds.), *New approaches to Slavic verbs of motion* (pp. 343-359). Amsterdam: John Benjamins.

Nettle, Daniel. (1999). *Linguistic diversity*. Oxford: Oxford University Press.

Nicholls, Geoff K. & Russel D. Gray. (2008). Dated ancestral trees from binary trait data and its application to the diversification of languages. *Journal of the Royal Statistical Society: Series B* 70(3): 545-566.

Nichols, Johanna & Tandy Warnow. (2008). Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5): 760-820.

Nicolas, Christian. (2005). *Sic enim appello...: essau sur l'autonymie terminologique gréco-latine chez Cicéron*. Leuven: Peeters.

Noonan, Michael. (2003). Motion events in Chantyal. In Erin Shay & Uwe Siebert (eds.), *Motion, direction, and location in languages* (pp. 211-234). Amsterdam: John Benjamins.

Nunn, Charles L. (2011). *The comparative approach in evolutionary anthropology and biology*. Chicago: University of Chicago Press.

Nunn, Charles L., Christian Arnold, Luke Matthews & Monique Borgerhoff-Mulder. (2010). Simulating trait evolution for cross-cultural comparison. *Philosophical Transactions of the Royal Society (B)* 365: 3807-3819.

Ó Baoill, Dónall (1975). Deixis in Modern Irish and certain related problems. *Ériu* 26: 144-161.

OED online (2012). Oxford: Oxford University Press.

O'Meara, Brian C., Cécile Ané, Michael J. Sanderson & Peter C. Wainwright. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60(5): 922-933.

Oh, Kyung-ju. (2009). Motion events in English and Korean fictional writings and translations. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan (eds.), *Crosslinguistic approaches to the psychology of language: research in the tradition of Dan Isaac Slobin* (pp. 253-262). New York: Psychology Press.

Oksanen, Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, Henry H. Stevens & Helene Wagner. (2012). *vegan: Community Ecology Package*. R package available online: http://vegan.r-forge.r-project.org/

Olivieri, Dante. (1965). *Dizionario etimologico italiano, concordato coi dialetti, le lingue straniere e la topo-onomastica*. Milano: Ceschina.

Orel, Vladimir E. (1998). *Albanian etymological dictionary*. Leiden: Brill.

Orel, Vladimir E. (2000). *A concise historical grammar of the Albanian language: reconstruction of Proto-Albanian*. Leiden: Brill.

Östergren, Olof. (1981). *Nusvensk ordbok*. Stockholm: Wahlström & Widstrand.

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

Özçalışkan, Şeyda. (2009). Learning to talk about spatial motion in language-specific ways. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 263-276). New York: Psychology Press.

Özçalışkan, Şeyda & Dan. I. Slobin. (2003). Codability effects on the expression of manner of motion in English and Turkish. In Sumru Özsoy, Didar Akar, Mine Nakipoğlu-Demiralp, Eser Erguvanlı-Taylan 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. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B* 255: 37-45.

Pagel, Mark. (1997). Inferring evolutionary processes from phylogenies. *Zoologica Scriptica* 26: 331-348.

Pagel, Mark. (1999a). Inferring the historical patterns of biological evolution. *Nature* 401: 877-884.

Pagel, Mark. (1999b). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* 48: 612-622.

Pagel, Mark. (2000). Maximum-likelihood models for glottochronology and for reconstructing linguistic phylogenies. In Colin Renfrew, April McMahon & Larry Trask (eds.), *Time depth in historical linguistics* (pp. 189-207). Cambridge: Mc Donald Institute for Archaeological Research.

Pagel, Mark. (2002). Modeling the evolution of continuously varying characters on phylogenetic trees: The case of Homonid cranial capacity. In Norman MacLeod & Peter L. Forey (eds.), *Morphology, shape and phylogeny* (pp. 269-286). Boca Raton: CRC Press.

Pagel, Mark, Quentin D. Atkinson & Andrew Meade. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449(11): 717-721.

Pagel, Mark & Andrew Meade. (2004). A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology* 53(4): 571-581.

Pagel, Mark & Andrew Meade. (2005). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53: 673-684.

Pagel, Mark & Andrew Meade. (2006). Bayesian analysis of correlated evolution of discrete characters by Reversible-Jump Markov Chain Monte Carlo. *The American Naturalist* 167(6): 808-825.

Pagel, Mark, Andrew Meade & Daniel Barker. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53(5): 673-684.

Papafragou, Anna, Christine Massey & Lila Gleitman. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition* 84: 189-219.

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

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

Peyraube, Alain. (2006). Motion events in Chinese: A diachronic study of directional complements. In Maya Hickmann & Stéphane Robert (eds.), *Space across languages: Linguistic systems and cognitive categories* (pp. 121-135). Amsterdam: John Benjamins.

Pfeifer, Wolfgang. (1993). *Etymologisches Wörterbuch des Deutschen*. Berlin: Akademie Verlag.

Philippa, Marlies, Frans Debrabandere & Arend Quak. (2003-present). *Etymologisch woordenboek van het Nederlands*. Amsterdam: Amsterdam University Press.

Pianigiani, Ottorino (1907). *Vocabolario etimologico della lingua italiana*. Roma: Società editrice Dante Alighieri di Albrighi, Segati.

Pokorny, Julius. (1925). *Altirische Grammatik*. Berlin: Walter de Gruyter.

Pokorny, Julius. (1959-1969). *Indogermanisches Etymologisches Wörterbuch*. Bern: Francke.

Pourcel, Stéphanie. (2004). What makes path of motion salient? *Proceedings of the 30th annual meeting of the Berkeley Linguistic Society* 30: 505-516.

Pourcel, Stéphanie & Anetta Kopecka. (2005) Motion expression in French: Typological diversity. *Durham & Newcastle working papers in linguistics* 11: 139-153.

Puşcariu, Sextil (1905). *Etymologisches Wörterbuch der rumänischen Sprache*. Heidelberg: Winter.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ragnarsdóttir, Hrafnhildur & Sven Strömqvist. (2004). Time, space, and manner in Swedish and Icelandic: Narrative construction in two closely related languages. In Sven Strömqvist & Ludo Verhoven (eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 113-141). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

Revell, Liam J., Luke J. Harmon & David C. Collar. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology* 57(4): 591-601.

Revell, Liam J., D. Luke Mahler, Pedro R. Peres-Neto & Benjamin D. Redelings. (2011). A new phylogenetic method for identifying exceptional phenotypic diversification. *Evolution* 66(1):135-46.

Ricca, Davide. (1993). *I verbi deittici di movimento in Europa: una ricerca interlinguistica*. Firenze: La Nova Italia.

Rijkhoff, Jan & Dik Bakker. (1998). Language sampling. *Linguistic Typology* 2: 263-314.

Ringe, Don. (2006). *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.

Ringe, Don, Tandy Warnow & Ann Taylor. (2002). Indo-European and computational cladistics. *Transactions of the Philological Society* 100: 59-129.

Ritt, Nikolaus. (2004). *Selfish sounds and linguistic evolution: A Darwinian approach to language change*. Cambridge: Cambridge University Press.

Rix, Helmut. (2001). *LIV: Lexicon der indogermanischen verben*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Roberts, Murat H. (1936). The Antiquity of the Germanic verb-adverb locution. *The Journal of English and Germanic Philology* 35(4): 466-481.

Roberts, Seán. (2009). *Word complexity predicts rates of lexical evolution*. Paper presented at the Language at Edinburgh Lunch.

Rohlf, F. James. (2006). A comment on phylogenetic correction. *Evolution* 60(7): 1509-1515.

Ronquist, Fredrik, Paul van der Mark & John P. Huelsenbeck. (2009). Bayesian phylogenetic analysis using MrBayes. In Philippe Lemey, Marco Salemi & Anne-Mieke Vandamme (eds.), *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*. Cambridge: Cambridge University Press.

Saunders, Arpiar. (2005). *Linguistic phylogenetics of the Austronesian family: A performance review of methods adapted from biology*. BA Thesis, Swarthmore College.

Schluter, Dolph, Trevor Price, Arne Ø. Mooers & Donald Ludwig. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution* 51(6): 1699-1711.

Schmitt, Rüdiger. (1981). *Grammatik des Klassish-Armenischen mit sprachvergleichenden Erläuterungen*. Inssbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Schwegler, Armin. (1990). *Analyticity and syntheticity*. Berlin: Mouton de Gruyter.

Schwyzer, Eduard. (1950). *Griechische Grammatik vol. 2: Syntax und syntaktische Stilistik*. München: C. H. Beck'sche Verlagsbuchhandlung.

Sinha, Chris & Tania Kuteva. (1995). Distributed spatial semantics. *Nordic Journal of Linguistics* 18: 167-99.

Skok, Petar. (1971-1974). *Etimologijski rječnik hrvatskoga ili srpskoga jezika*. Zagreb: Jugoslavenska akad. znanosti i umjetnosti.

Skopeteas, Stavros. (2002). *Lokale Konstruktionen im Griechischen: Sprachwandel in funktionaler Sicht*. PhD thesis, Universität Erfurt.

Sławski, Franciszek. (1952-1982). *Słownik etymologiczny języka polskiego*. Kraków: nakład. Towarzystwo miłośników języka polskiego.

Slobin, Dan I. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics* 1(1): 7-26.

Slobin, Dan I. (1996a). From "thought and language" to "thinking for speaking". In John J. Gumperz & Stephen C. Levinson (eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.

Slobin, Dan I. (1996b). Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibitani & Sandra A. Thompson (eds.), *Grammatical constructions: Their form and meaning* (pp. 195-219). Oxford: Clarendon Press.

Slobin, Dan I. (1997). Mind, code, and text. In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type: Dedicated to T. Givón* (pp. 437-467). Amsterdam: John Benjamins.

Slobin, Dan I. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism In Susanne Niemeier & René Dirven (eds.), *Evidence for linguistic relativity* (pp. 107-138). Amsterdam: John Benjamins.

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

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

Slobin, Dan I. (2005a). Linguistic representations of motion events: What is signifier and what is signified? In Costantino Maeder, Olga Fischer & William J. Herlofsky (eds.), *Outside-in-Inside-out* (pp. 307-322). Amsterdam: John Benjamins.

Slobin, Dan I. (2005b). Relating narrative events in translation. In Dorit Diskin Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman* (pp. 115-129). Dordrecht: Kluwer.

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

Slobin, Dan I. & Nini Hoiting. (1994). Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society* (pp. 487-505). Berkeley: University of California.

Speyer, Jacob Samuel. (1896). *Vedische und Sanskrit-Syntax*. Strassburg: K. J. Trübner.

Stassen, Leon. (1997). *Intransitive predication*. Oxford: Oxford University Press.

Stevens, Christopher M. (1992). Grammaticalization in spatial deixis: A case study. In Irmengard Rauch, Gerald F. Carr & Robert L. Kyes (eds.), *On Germanic linguistics: Issues and methods* (pp. 299-313). Berlin: Mouton de Gruyter.

Swadesh, Morris. (1952). Lexico-statistic dating of prehistoric ethnic contacts: With special reference for North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4): 452-463.

Swadesh, Morris. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2): 121-137.

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

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

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

Talmy, Leonard. (2007). Lexical typologies. In Timothy Shopen (ed.), *Language typology and syntactic description: Grammatical categories and the lexicon* (pp. 66-168). Cambridge: Cambridge University Press.

Taylor, John R. (1989). *Linguistic categorization: Prototypes in linguistic theory*. Oxford: Clarendon Press.

Tëmkin, Ilya & Niles Eldredge. (2007). Phylogenetics and material cultural evolution. *Current Anthropology* 48(1): 146-154.

Toner, Gregory, Maxim Fomin, Thomas Torma & Grigory Bondarenko. (2007). *Electronic dictionary of the Irish language*. Dublin: Royal Irish Academy.

Townsend, Charles E. (1968). *Russian word-formation.* New York: McGraw-Hill Book Company.

Tucker, Thomas George. (1931). *A concise etymological dictionary of Latin*. Halle: Niemeyer.

Turner, Ralph Lilley. (1931). *A comparative and etymological dictionary of the Nepali language*. London: K. Paul, Trench and Trubner.

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

Uhlenbeck, Christiaan Cornelis. (1899). *Kurzgefasstes etymologisches Wörterbuch der altindischen Sprache* Amsterdam: Müller.

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

Vasmer, Max. (1953-1958). *Russisches etymologisches Wörterbuch*. Heidelberg: Winter.

Venditti, Chris, Andrew Meade & Mark Pagel. (2011). Multiple routes to mammalian diversity. *Nature* 479: 393-396.

Vendryes, Joseph Jean Baptiste Marie. (1959-present). *Lexique étymologique de l'irlandais ancien*. Dublin: Institute for Advanced Studies.

Verkerk, Annemarie. (2013). Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. *Language Dynamics and Change* 3(2):169-217. Special issue, "Phylogeny and beyond: Quantitative diachronic approaches to language diversity", edited by Michael Dunn.

Verkerk, Annemarie. (2014a). Diachronic change in Indo-European motion event encoding. *Journal of Historical Linguistics* 4(1): 40-83.

Verkerk, Annemarie. (2014b). The correlation between motion event encoding and path verb lexicon size in the Indo-European language family. *Folia Linguistica Historica*.

Verkerk, Annemarie. (2014c). Where Alice fell into: Motion events in a parallel corpus. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology and register analysis: Linguistic variation in text and speech* (pp. 324-354). Berlin: Walter de Gruyter.

Verkerk, Annemarie. (submitted). Where do all the motion verbs come from? The speed of development of manner verbs and path verbs in Indo-European. *Diachronica*.

Verkerk, Annemarie & Frostad, Benedicte H. (2013). The encoding of manner predications and resultatives in Oceanic: A typological and historical overview. *Oceanic Linguistics* 52(1): 1-57.

Viberg, Åke. (1983). The verbs of perception: A typological study. *Linguistics* 21: 123-162.

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

Vidaković, Ivana. (2012). He walked up the pole with arms and legs: Typology in second language acquisition. In Luna Filipović & Kasia M. Jaszczolt (eds.), *Space and time in languages and cultures: Linguistic diversity* (pp. 233-262). Amsterdam: John Benjamins.

Vincent, Nigel. (1999). The evolution of c-structure: prepositions and PPs from Indo-European to Romance. *Linguistics* 37: 1111-1153.

de Vries, Jan. (1971). *Nederlands etymologisch woordenboek*. Leiden: Brill.

Wälchli, Bernhard. (2001a). A typology of displacement (with special reference to Latvian). STUF - *Sprachtypologie und Universalienforschung* 54: 298-323.

Wälchli, Bernhard. (2001b). Lexical evidence for the parallel development of the Latvian and Livonian verb particles. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages vol. 2: Grammar and typology* (pp. 413-441). Amsterdam: John Benjamins.

Wälchli, Bernhard. (2007). Advantages and disadvantages of using parallel texts in typological considerations. *STUF - Sprachtypologie und Universalienforschung* 60: 118-134.

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

Walde, Alois. (1930-1956). *Lateinisches etymologisches Wörterbuch*. Heidelberg: Winter.

Walde, Alois & Julius Pokorny. (1927-1932). *Vergleichendes Wörterbuch der indogermanischen Sprachen*. Berlin: de Gruyter.

von Wartburg, Walther. (1922-2003). *Französisches etymologisches Wörterbuch*. Bonn: Klopp Verlag.

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

Watkins, Calvert. (1966). Italo-Celtic revisited. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects* (pp. 29-50). Berkeley: University of California Press.

Webster, Andrea J. & Andy Purvis. (2002). Ancestral states and evolutionary rates of continuous characters. In Norman MacLeod & Peter L. Forey (eds.), *Morphology, shape and phylogeny* (pp. 247-268). Boca Raton: CRC Press.

Weins, John J. (ed.). (2000). *The phylogenetic analysis of morphological data*. Washington: Smithsonian Institution Press.

Whitney, William Dwight. (1879). *A Sanskrit grammar*. Leipzig: Breitkopf and Härtel.

Wichmann, Søren & Arpiar Saunders. (2007). How to use typological databases in historical linguistic research. *Diachronica* 24: 373-404.

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, Armin von Stechow & Götz Wienold (eds.), *Lexical knowledge in the organization of language* (pp. 301-340). Amsterdam: John Benjamins.

Wilkins, David P. (2004). The verbalization of motion events in Arrernte. In Sven Strömqvist & Ludo Verhoven (eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 143-157). Mahwah: Lawrence Erlbaum.

Wilkins, David P. & Deborah Hill. (1995). When GO means COME: Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6: 209-59.

Windfuhr, Gernot L. (2009). *The Iranian languages*. London: Routledge.

Witkowski, Stanley R., Cecil H. Brown & Paul K. Chase. (1981). Where do tree terms come from? *Man* 16(1): 1-14.

Zlatev, Jordan & Caroline David. (2006). Three ways to travel: Motion events in French, Swedish and Thai. In Augusto Soares da Silva, Amadeu Torres & Miguel Gonçalves (eds.), *Linguagem, cultura e cognição. Estudos de linguística cognitiva* (pp. 119-142). Coimbra: Almedina.

Zlatev, Jordan & Peerapat Yangklang. (2004). A third way to travel: The place of Thai in motion-event typology. In Sven Strömqvist & Ludo Verhoven (eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 159-190). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendices

Appendix 1: List of translations used for the parallel corpus

Appendix 2: An overview of manner verbs in two different corpora

Appendix 3: List of the original sentences included in the parallel corpus sentence samples

Appendix 4: Maximum clade credibility trees used for robustness tests in ancestral state estimation

Appendix 5: Manner verbs in Indo-European

Appendix 6: Complex manner verbs in Indo-European

Appendix 7: Non-motion verbs used as manner verbs in Indo-European

Appendix 8: Number of types of manner verb per novel

Appendix 9: Path verbs in Indo-European

Appendix 10: List of consulted etymological dictionaries in the etymological analysis of motion verbs

Appendix 11: Manner verb etymologies

Appendix 12: Path verb etymologies

Appendix 13: The etymological origins of manner verbs, averaged for all languages and the four big subgroups

Appendix 14: The etymological origins of path verbs, averaged for all languages and the four big subgroups

Appendix 15: The etymological origins of manner verbs and path verbs averaged for all languages and separated by tier

Appendix 16: Paired sample t-tests of the number of manner verbs and path verbs in each etymology category

Appendix 1: List of translations used for the parallel corpus

Albanian

Carroll, Lewis. (2009). *Liza në botën e çudirave* (T. Tafa & E. Tafa, Trans.). Tirana: Vëllezërit Tafa.

Coelho, Paulo. (2004). *Alkimisti* (N. Varfi, Trans.). Tirana: Toena.

Armenian

Carroll, Lewis. (1994). *Alise črašk'nerw aškharčoim ši Alise čayšloi aškharčoim* (S. Seferian & S. Mkrtichian, Trans.). Yerevan: Samson Publishers.

Coelho, Paulo. (2004). *Alk'imikosĕ* (A. Vardazaryan, Trans.). Yerevan: Van Aryan.

Dutch

Carroll, Lewis. (2006). *De Avonturen van Alice in Wonderland en Spiegelland* (P. Bulthuis, A. Kossmann & C. Reedijk, Trans.). Rotterdam: Donker B.V.

Coelho, Paulo. (1994). *De Alchemist* (H. Lemmens, Trans.). Amsterdam: Arbeiderspers.

English

Carroll, Lewis. (1865). *Alice's Adventures in Wonderland*. London: Macmillan.

Carroll, Lewis. (1871). *Through the Looking-Glass and What Alice Found There*. London: Macmillan.

Coelho, Paulo. (2006). *The alchemist* (A. R. Clarke, Trans.). London: Harper.

French

Carroll, Lewis. (2009). *Les aventures d'Alice au pays des merveilles et La traversée du miroir et ce qu' Alice trouva de l'autre côte* (L. Bury, Trans.). Paris: Le livre de poche.

Coelho, Paulo. (2007). *L'alchimiste* (J. Orecchioni, Trans.). Paris: J'ai Lu.

German

Carroll, Lewis. (1999). *Alices Abenteuer im Wunderland* (G. Flemming, Trans.). Stuttgart: Reclam.

Carroll, Lewis. (2000). *Durch den Spiegel und was Alice dort fand* (G. Flemming, Trans.). Stuttgart: Reclam.

Coelho, Paulo. (1996). *Der Alchimist* (C. Swoboda Herzog, Trans.). Zürich: Diogenes.

Hindi

Carroll, Lewis. (2009). *Nanhi Elisa van daralainda mem* (N. Dasgupt, Trans.). Delhi: Aarogiya Nidhi Prakashan.

Coelho, Paulo. (2009). *Elkemista* (Trans. unknown). Delhi: Wisdom Tree.

Irish

Carroll, Lewis. (2007). *Eachtraí Eilíse i dtír na nIontas* (N. J. Williams, Trans.). Westport, Co. Mayo: Evertype.

Carroll, Lewis. (2009). *Lastall den scáthán agus a bhfuair Eilís ann roimpi* (N. J. Williams, Trans.). Westport, Co. Mayo: Evertype.

Coelho, Paulo. (2010). *An tAilceimiceoir* (S. Ó Coileáin, Trans.). Howth: Coisceim.

Italian

Carroll, Lewis. (1978). *Alice nel Paese delle Meraviglie e Attraverso lo Specchio*. (R. Carano & G. Pozzo, Trans.). Turin: Giulio Einaudi.

Coelho, Paulo. (1995). *L'Alchimista* (R. Desti, Trans.). Milan: Bompiani.

Latvian

Carroll, Lewis. (2002). *Alises piedzīvojumi brīnumzemē. Alise Aizspogulijā, un ko viņa tur redzēja* (E. Melbārzde, Trans.). Riga: Zvaigzne ABC Publishers.

Coelho Paulo. (1998). *Alkīmīkis* (E. Raups, Trans.). Riga: Jāna Rozes apgāds.

Lithuanian

Carroll, Lewis. (1991). *Alisa stubuklų šalyje ir Veidrodžio karalystėje* (K. Grigas & J. Lapienytė, Trans.). Vilnius: Vyturys.

Coelho, Paulo. (2004). *Alchemikas* (R. Knizikevičiūtė, Trans.). Vilnius: Leidykla VAGA.

Modern Greek

Carroll, Lewis. (1979). *Mes ston Kathrephtē kai ti vrēke hē Alikē ekei* (S. Kakise, Trans.). Athens: Ekdoseis Hypsilon.

Carroll, Lewis. (2003). *E Alikē stē khōa tōn Phaupátōn* (P. Pampoúdē, Trans.). Athens: Printa Roes .

Coelho, Paulo. (1996). *Ho alchèmistēs* (M.-P. Chidiroglou, Trans.). Athens: Ekdotikos Organismos Libane.

Nepali

Carroll, Lewis. (1993 [2049]). *Anautho teshamaa rilasa* (C. Mani Bandhu, Trans.). Kathmandu: Ekta Books.

Coelho, Paulo. (2002). *Kimiyagar* (I. Thapa, Trans.). Kathmandu: Walden Book House.

Persian

Carroll, Lewis. (1995 [1373]). *alis dær særzamine æǰayeb* (A. Karami Far, Trans.). Tehran: Arghavan Publication.

Coelho, Paulo. (2005). *Kimiagar* (Ā. Hijāzī, Trans.). Tehran: Caravan Books.

Polish

Carroll, Lewis. (2005). *Alicja w Krainie Czarów. Alicja po drugiej stronie zwierciadta* (A. Marianowicz & H. Baltyn, Trans.). Warsaw: Nasza Księgarnia.

Coelho, Paulo. (1995). *Alchemik* (B. Stępień & A. Kowalski, Trans.). Warsaw: Drzewo Babel.

Portuguese

Carroll, Lewis. (2000). *As aventuras de Alice no país das maravilhas e Alice do outro lado do espelho* (N. J. Williams & M. Vale de Gato, Trans.). Lisbon: Universos Mágicos.

Coelho, Paulo. (1988). *O Alquimista*. Lisbon: Pergaminho.

Romanian

Carroll, Lewis. (2006). *Alice în ţara minunilor* (T. Antoniu, Trans.). Bucharest: Editura Cartex 2000.

Carroll, Lewis. (2006). *Alice în ţara oglinzilor* (T. Antoniu, Trans.). Bucharest: Editura Cartex 2000.

Coelho, Paulo. (2003). *Alchimistul* (G. Banu, Trans.) Bucharest: Humanitas.

Russian

Carroll, Lewis. (2007). *Prikljuczenija Alicy w strane chudes. Skwoz' serkalo i chto tam uwidela Alica, ili Alica w zazerkal'e. Pischa dlja uma* (N. Demurova, G. Krujkova & A. Bochenkova, Trans.). Moscow: Eksmo.

Coelho, Paulo. (2008). *Alkhimik* (A. Bogdanovskogo, Trans.). Moscow: ACT.

Serbo-Croatian

Carroll, Lewis. (1989). *Alica u zemlji čudesa* (M. Jurkić-Šunjić, Trans.). Zagreb: Mladosi.

Carroll, Lewis. (2004). *Alica u zemlji čudesa i iza zrcala* (A. Šoljan, Trans.). Zagreb: Školska Knjiga.

Coelho, Paulo. (1998). *Alhemičar* (R. Tatić, Trans.). Belgrade: Paideia.

Swedish

Carroll, Lewis. (2001). *Alice i Underlandet – Alice i Spegellandet* (H. Lundin, Trans.). Stockholm: Rabén & Sjögren Bokförlag.

Coelho, Paulo. (2003). *Alkemisten* (Ö. Sjögren, Trans.). Stockholm: Bazar Förlag AB.

Appendix 2. An overview of manner verbs in two different corpora

This appendix gives an overview of 7 different English manner verbs and their Ancient Greek correspondents in two text corpora (first and fourth column). The English numbers are taken from *Alice's Adventures in Wonderland* and *Through the Looking-Glass and What Alice Found There* (by Lewis Carroll, consisting of 56777 words). The Ancient Greek numbers are taken from the *New Testament* (consisting of 179011 words) available at PROIEL (Haug & Jøhndal 2008). It shows that the *New Testament*, contains approximately the same number of manner verbs (second and fifth column). However, since the *New Testament* contains approximately three times more words as the *Alice* corpus, contains fewer manner verbs per 10.000 words (third and sixth column).

English	no.	10.000	Ancient Greek	no.	10.000
run	45	7.9	*τρέχω* 'to run'	19	1.1
walk	35	6.2	*(περι)πατέω* 'to go about, walk'	85	4.7
jump	20	3.5	*πηδαν; ἅλλομαι* 'to jump, leap'	3	0.2
swim	7	1.2	*νειν*; *κολυμβάω* 'to swim'	1	0.06
fly	5	0.9	*πέτομαι* 'to fly'	3	0.2
crawl	4	0.7	*ερπειν* 'to crawl'	0	0
rush	2	0.4	*ὁρμάω* ' to rush'	5	0.3
sail	1	0.2	*πλέω* 'to sail'	18	1.0

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

Matthew 11.5: *The blind receive their sight, and the lame* ***walk****, the lepers are cleansed, and the deaf hear, the dead are raised up, and the poor have the gospel preached to them.*

Galatians 2.2: *And I went up by revelation, and communicated unto them that gospel which I preach among the Gentiles, but privately to them which were of reputation, lest by any means I should* ***run****, or had* ***run****, in vain.*

Revelations 4.7: *And the first beast was like a lion, and the second beast like a calf, and the third beast had a face as a man, and the fourth beast was like a* ***flying*** *eagle.*

Appendix 3: List of the original sentences included in the parallel corpus sentence samples

Name of sample	Sentences included	Sources
118 sentence sample	included only sentences that encode just path or both path or manner in the originals	*Alice's Adventures in Wonderland* + *O Alquimista*
132 sentence sample	also includes sentences that encoding only manner in the originals	*Alice's Adventures in Wonderland* + *O Alquimista* + *Through the Looking-Glass and What Alice Found There*
192 sentence sample	included only sentences that encode just path or both path or manner in the originals	*Alice's Adventures in Wonderland* + *O Alquimista*
215 sentence sample	also includes sentences that encoding only manner in the originals	*Alice's Adventures in Wonderland* + *O Alquimista* + *Through the Looking-Glass and What Alice Found There*

118-sentence sample
From *Alice's Adventures in Wonderland*:

A07 And was just in time to hear it say, as it **turned a corner**, 'Oh my ears and whiskers, how late it's getting!'
A08 How she longed **to get out of that dark hall**,
A09 when she **got to the door**,
A10 and when she **went back to the table** for it,
A11 so either way I'll **get into the garden**,
A12 It was the White Rabbit **returning**,
A13 She got up and **went to the table** to measure herself by it,
A14 **Come away**, my dears!
A15 On various pretexts they all **moved off**,
A16 By this time she **had found her way into a tidy little room** with a table in the window,
A17 `Then I'll go round and **get in at the window**.'
A18 Mind that loose slate--Oh, it's **coming down**!
A19 Alice turned and **came back again.**
A20 And she opened the door and **went in.**
A21 **After these came** the royal children;

A22 and he called the Queen, who **was passing** at the moment,
A23 the only difficulty was, that her **flamingo was gone across to the other side of the garden**,
A24 The King and Queen of Hearts were seated on their throne when they **arrived**,
A25 The Hatter looked at the March Hare, who **had followed him into the court**, arm-in-arm with the Dormouse.
A26 and she thought at first she would get up and **leave the court**;
A27 just as the Dormouse **crossed the court**,
A28 At this the whole pack **rose up into the air**,
A29 and fortunately was just in time to see it **pop down a large rabbit-hole under the hedge**.
A30 so suddenly that Alice had not a moment to think about stopping herself before she found herself **falling down a very deep well**.
A31 I shall think nothing of **tumbling down stairs**!
A32 Why, I wouldn't say anything about it, even if I **fell off the top of the house**!
A33 when suddenly, **thump! thump! down she came upon a heap of sticks and dry leaves**, and the fall was over.
A34 **away went Alice like the wind**,
A35 and she tried her best to **climb up one of the legs of the table**, but it was too slippery;
A36 and if it makes me grow smaller, I **can creep under the door**;
A37 and **hurried off to the garden door**.
A38 he **came trotting along in a great hurry**,
A39 and **skurried away into the darkness as hard as he could go**.
A40 and she **ran with all speed back to the little door**:
A41 Her first idea was that she had somehow **fallen into the sea**,
A42 and she soon made out that it was only a mouse that **had slipped in** like herself.
A43 The Mouse **gave a sudden leap out of the water**,
A44 For the Mouse **was swimming away from her as hard as it could go**, and making quite a commotion in the pool as it went.
A45 When the Mouse heard this, it turned round and **swam slowly back to her**:
A46 It was high time to go, for the pool was getting quite crowded with the birds and animals that **had fallen into it**:
A47 and the whole party **swam to the shore**.
A48 It was the White Rabbit, **trotting slowly back again**,
A49 from which she concluded that it was just possible **it had fallen into a cucumber-frame**, or something of the sort.
A50 something **comes at me like a Jack-in-the-box**

A51 and **up I goes like a sky-rocket**!

A52 for the next moment a shower of little pebbles **came rattling in at the window**, and some of them hit her in the face

A53 she **ran out of the house**,

A54 They all **made a rush at Alice** the moment she appeared

A55 whereupon the **puppy jumped into the air off all its feet at once**, with a yelp of delight,

A56 and **rushed at the stick**

A57 the puppy **made another rush at the stick**

A58 and **crawled away in the grass**,

A60 and **was going to dive in among the leaves**,

A61 a large pigeon **had flown into her face**,

A62 they **must needs come wriggling down from the sky**! Ugh, Serpent!'

A63 when suddenly a footman in livery **came running out of the wood**--

A64 and **crept a little way out of the wood** to listen.

A65 Alice laughed so much at this, that she had to **run back into the wood** for fear of their hearing her;

A66 At this moment the door of the house opened, and a large plate **came skimming out, straight at the Footman's head**

A68 `Oh, there goes his PRECIOUS nose'; as an unusually large saucepan **flew close by it**, and very nearly carried it off.

A69 and she **hurried out of the room**.

A70 So she set the little creature down, and felt quite relieved to see it **trot away quietly into the wood**.

A71 she got up in great disgust, and **walked off**;

A72 and the little dears **came jumping merrily along** hand in hand, in couples:

A73 and then **quietly marched off after the others**.

A74 it was very provoking to find that the hedgehog had unrolled itself, and **was in the act of crawling away**:

A75 and, as the doubled-up soldiers were always getting up and **walking off to other parts of the ground**,

A76 where Alice could see it **trying in a helpless sort of way to fly up into a tree**.

A77 And the executioner **went off like an arrow**.

A78 thought Alice, as she **went slowly after it**:

A80 `You may go,' said the King, and the **Hatter hurriedly left the court**,

A81 and **came flying down upon her**:

A82 that **had fluttered down from the trees upon her face.**

From *O Alquimista:*

C07 *O falcão conhecia bem*
DEF.ART.M.SG falcon.M know.IND.IPFV.3SG well
a linguagem d-o deserto,
DEF.ART.F.SG language.F of-DEF.ART.M.SG desert.M
e quando paravam, ele ***saía***
and when stop.IND.IPFV.3PL 3SG.M leave.IND.IPFV.3SG
d-o ombro d-o Alquimista
from-DEF.ART.M.SG shoulder.M of-DEF.ART.M.SG Alchemist.M
e voava em busca de alimento.
and fly.IND.IPFV.3SG in search.F of food.M
'The falcon knew the language of the desert very well, and when they stopped he would leave the shoulder of the Alchemist and fly away in search of food.'

C10 *"O rapaz começou a* ***subir***
DEF.ART.M.SG boy.M begin.IND.PFV.3SG to ascend.INF
e descer a-s escadaria-s
and descend.INF DEF.ART.F-PL flight.of.stairs.F-PL
d-o palácio, *mantendo sempre*
of-DEF.ART.M.SG palace.M maintain.PRS.PTCP always
o-s olho-s fixo-s n-a
DEF.ART.M-PL eye.M-PL fixed.M-PL at-DEF.ART.F.SG
colher.
spoon.F
'The boy began to ascend and descend the stairs of the palace, always keeping his eyes fixed on the spoon.'

C11 *-Você tem que* ***atravessar todo***
2SG have.IND.PRS.3SG to traverse.INF whole.M
o deserto de Saara *- disse*
DEF.ART.M.SG desert.M of Sahara say.IND.PFV.3SG
o recém-chegado.
DEF.ART.M.SG newly-arrived.M
'-You have to cross the entire Sahara desert- said the newcomer.'

C12 ***Chegaram enfim n-o meio de***
arrive.IND.PFV.3PL finally in-DEF.ART.M.SG center.M of
uma grande praça, *onde*
INDF.ART.F.SG large square.F where
funcionava o mercado.
function.IND.IPFV.3SG DEF.ART.M.SG market.M
'They finally arrived in the middle of a large square, where a market was being held.'

C13 *A-o* *mesmo* *tempo,* *via*
at-DEF.ART.M.SG same.M time.M see.IND.IPFV.1SG
passar ***defronte*** ***a*** ***minha***
pass.INF in.front DEF.ART.F.SG 1SG.F.POSS
loja *muita-s* *pessoa-s* *que* *seguiam*
shop.F many.F-PL people.F-PL that proceed.IND.IPFV.3PL
n-a *direção* *de* *Meca.*
in-DEF.ART.F.SG direction.F of Mecca
'At the same time, I see a lot of people that are going towards Mecca passing in front of my shop.'

C14 *O-s* *homens* ***subiam***
DEF.ART.M-PL man.PL ascend.IND.IPFV.3PL
a ***ladeira*** *e* *ficavam*
DEF.ART.F.SG slope.F and become.INF.IPFV.3PL
cansado-s.
tired.M-PL
'The people that ascended the slope became tired.'

C15 *- Já* ***cruzei*** ***muita-s*** ***vez-es***
already cross.IND.PFV.1SG many.F-PL time.F-PL
esta-s ***areia-s*** - *disse* *um*
DEM.PROX.F-PL sand.F-PL say.IND.PFV.3SG INDF.ART.M.SG
cameleiro *certa* *noite.*
camel.driver.M certain.F evening.F
'- I have already crossed these sands many times, - said the camel driver one night.'

C16 *Se* *a* *caravana* *chegava* *em*
if DEF.ART.F.SG caravan.F come.IND.IPFV.3SG in
frente *a* *uma* *pedra,* *ela*
front.F to INDF.ART.F.SG rock.F 3SG.F
a ***contornava;*** *se* *estavam*
OBJ.F.SG go.around.IND.IPFV.3SG if be.IND.IPFV.3PL
diante *de* *um* *rochedo,*
in.front of INDF.ART.M.SG cliff.M
davam *uma* *longa* *volta.*
give.IND.IPFV.3PL INDF.ART.F.SG long.F turn.F
'If the caravan arrives in front of a rock, they would go around it; if it arrives in front of a cliff, it would make a large loop.'

C17

não	*importava*		***quanta-s***		***volta-s***	
NEG	matter.IND.IPFV.3SG		how.many.F-PL		turn.F-PL	
tivesse	***que***	***dar****,*	*a*		*caravana*	
have.SBJV.IPFV.3SG	to	give.INF	DEF.ART.F.SG		caravan.F	
seguia		*sempre*	*em*	*direção*	*a*	
proceed.IND.IPFV.3SG		always	in	direction.F	to	
um	*mesmo*	*ponto.*				
INDF.ART.M.SG	same.M	point.M				

‘It did not matter how many turns they had to make, the caravan always proceeded towards the same point.’

C18

N-uma		*d-esta-s*	*noite-s*		*o*
in-INDF.ART.F.SG		of-DEM.PROX.F-PL	evening.F-PL		DEF.ART.M.SG
cameleiro		***veio***	***até***	***a***	
camel.driver.M		come.IND.PFV.3SG	to	DEF.ART.F.SG	
fogueira	*onde*	*o*	*rapaz*	*e*	*o*
camp.fire.F	where	DEF.ART.M.SG	boy.M	and	DEF.ART.M.SG
Inglês		*estavam*	*sentado-s.*		
Englishman.M		be.IND.IPFV.3PL	seat.PTCP.M-PL		

‘One of those evenings, the camel driver came to the fire where the boy and the Englishman were sitting.’

C19

O	*rapaz*	***se***	***aproximou***
DEF.ART.M.SG	boy.M	REFL.3SG	draw.closer.IND.PFV.3SG
d-a	***moça.***		
of-DEF.ART.F.SG	girl.F		

‘The boy approached the girl.’

C20

O		*Inglês*		***partiu***	
DEF.ART.M.SG		Englishman.M		leave.IND.PFV.3SG	
também,	***em***	***busca***	***d-o***		***Alquimista.***
too	in	search.F	of-DEF.ART.M.SG		Alchemist.M

‘The Englishman left too, in search of the Alchemist.’

C21

N-o		*dia*	*seguinte*	*o*		*rapaz*
on-DEF.ART.M.SG		day.M	next	DEF.ART.M.SG		boy.M
voltou		***para***	***o***		***poço****,*	*para*
return.IND.PFV.3SG		to	DEF.ART.M.SG		well.M	to
esperar	*a*			*moça.*		
wait.INF	DEF.ART.F.SG			girl.F		

‘The next day, the boy returned to the well to wait for the girl.’

C22

Depois	*tornou*		*a*	*encher*	*seu*
then	return.IND.PFV.3SG		to	fill.INF	3SG.M.POSS
cântaro,	*e*	***foi***		***embora****.*	
water.jug.M	and	go.IND.PFV.3SG		away	

‘Then he returned to fill his jug of water and went away.’

C23 *Fátima* ***entrou*** ***n-a*** ***tenda****.*
Fatima enter.IND.PFV.3SG in-DEF.ART.F.SG tent.F
'Fatima entered the tent.'

C24 *Quando o dia chegasse, ela*
when DEF.ART.M.SG day.M come.SBJV.IPFV.3SG 3SG.F
ia ***sair*** *e fazer*
go.AUX.IND.IPFV.3SG go.out.INF and do.INF
aquilo que havia feito durante
DEM.DIST.N that have.AUX.IND.IPFV.3SG do.PTCP during
tanto-s ano-s; mas tudo
so.many.M-PL year.M-PL but everything
havia mudado.
have.AUX.IND.IPFV.3SG change.PTCP
'When the morning would come, she would go out and do the things she had been doing for so many years, but everything would be different.'

C25 *Depois de algum tempo o*
after of some.M.SG time.M DEF.ART.M.SG
falcão ***retornou*** *com a comida.*
falcon.M return.IND.PFV.3SG with DEF.ART.F.SG food.F
'After a while, the falcon returned with food.'

C26 *- perguntou o rapaz a-o*
ask.IND.PFV.3SG DEF.ART.M.SG boy.M to-DEF.ART.M.SG
Alquimista, quando já ***haviam***
Alchemist.M when already have.AUX.IND.IPFV.3PL
se ***distanciado*** *bastante.*
REFL.3PL move.away.PTCP sufficiently
'… - asked the boy to the Alchemist, after they had moved away far enough.'

C27 *- disse o rapaz, quando já*
say.IND.PFV.3SG DEF.ART.M.SG boy.M when already
tinham ***se*** ***afastado***
have.AUX.IND.IPFV.3PL REFL.3PL move.away.PTCP
um ***pouco*** ***d-o*** ***acampamento****.*
INDF.ART.M.SG bit.M from-DEF.ART.M.SG camp.M
'… - said the boy, when they had moved away a bit from the camp.'

C28 *Depois montaram em seu-s cavalo-s,*
then mount.IND.PFV.3PL on 3.M.POSS-PL horse.M-PL
e ***seguiram*** ***em*** ***direção*** ***à-s***
and proceed.IND.PFV.3PL in direction.F to.DEF.ART.F-PL
Pirâmide-s ***d-o*** ***Egito****.*
Pyramid.F-PL of-DEF.ART.M.SG Egypt.M
'Then they mounted their horses and started heading towards the Egyptian pyramids.'

C29 *N-o* *segundo* *dia* *o* *rapaz*
on-DEF.ART.M.SG second.M day.M DEF.ART.M.SG boy.M
foi ***para*** ***o*** ***alto*** ***de***
go.IND.PFV.3SG to DEF.ART.M.SG top.M of
uma ***rocha*** *que* *ficava* *perto*
INDF.ART.F.SG rock.F that remain.IND.IPFV.3SG close
d-o *acampamento.*
of-DEF.ART.M.SG camp.M
'On the second day, the boy went to the top of a rock that was located close to the camp.'

C30 *Quero* *ser* *como* *você,* ***penetrar***
want.IND.PRS.1SG be.INF like 2SG permeate.INF
em ***todo-s*** ***o-s*** ***canto-s,*** *atravessar*
in all.M-PL DEF.ART.M-PL corner.M-PL traverse.INF
o-s *mar-es,* *tirar* *a*
DEF.ART.M-PL sea.M-PL take.away.INF DEF.ART.F.SG
areia *que* *cobre* *meu* *tesouro,* *trazer*
sand.F that cover.IND.PRS.3SG 1SG.M.POSS treasure.M bring.INF
para *perto* *a* *voz* *de* *minha*
to near DEF.ART.F.SG voice.F of 1SG.F.POSS
amada.
beloved.F
'I want to be like you, permeate in all corners, cross the seas, blow away the sane that covers my treasure, bring the voice of my beloved close to me.'

C31 *Contava* *com* *orgulho* *a* *história*
tell.IND.IPFV.3SG with pride.M DEF.ART.F.SG story.F
de *um* *pastor* *que* ***havia***
of INDF.ART.M.SG shepherd.M that have.AUX.IND.IPFV.3SG
deixado ***sua-s*** ***ovelha-s*** *para* *seguir*
leave.behind.PTCP 3SG.F.POSS-PL sheep.F-PL to follow.INF
um *sono* *que* *se*
INDF.ART.M.SG sleep.M that REFL.3SG
repetiu *duas* *noite-s.*
repeat.IND.PFV.3SG two.F night.F-PL
'It told him with pride the story of a shepherd that left behind his sheep to follow a dream that he had two nights in a row.'

C32	*Era*		*tão*	*fascinado*	*por*	*si*
	be.IND.IPFV.3SG		so	fascinated.M	by	3SG.REFL
	mesmo	*que*	*certo*	*dia*	***caiu***	
	self.M	that	certain.M	day.M	fall.IND.PFV.3SG	
	dentro	***d-o***		***lago***	*e*	*morreu*
	inside	of-DEF.ART.M.SG		lake.M	and	die.IND.PFV.3SG
	afogado.					
	drown.PTCP.M					

‘He was so fascinated by his own reflection that one day, he fell into the lake and drowned.’

C33	*Afinal*	*de*	*conta-s,*	*apesar*	*de*	
	after.all	of	count.F-PL	despite	of	
	toda-s	*nós*	*sempre*	***correr-mos***	***atrás***	
	all.F-PL	1PL	always	run.INF-PRS.1PL	after	
	d-ele	***pel-o***		***bosque****,*	*você*	
	of-3SG.M	through-DEF.ART.M.SG		wood.M	2SG	
	era	*o*		*único*	*que*	
	be.IND.IPFV.3SG	DEF.ART.M.SG		only.one.M	that	
	tinha	*a*		*oportunidade*		*de*
	have.IND.IPFV.3SG	DEF.ART.F.SG		opportunity.F		to
	contemplar	*de*	*perto*	*sua*	*beleza.*	
	contemplate.INF	from	close	3SG.F.POSS	beauty.F	

‘After all, you are the ones who are always running after him through the forest, you are the only ones that have had the opportunity to contemplate his beauty from up close.’

C34	*A*	*África*	*ficava*			*a*
	DEF.ART.F.SG	Africa.F	be.situated.IND.IPFV.3SG			at
	apenas	*alguma-s*	*hora-s*	*d-a*		*Tarifa;*
	only	some.F-PL	hour.F-PL	from-DEF.ART.F.SG		Tarifa.F
	e ***era***		***só***	***cruzar***	***o***	
	and be.IND.IPFV.3SG		just	cross.INF	DEF.ART.M.SG	
	pequeno	***estreito***	***n-um***		***barco****.*	
	small.M	strait.M	in-INDF.ART.M		boat.M	

‘Africa is situated only a few hours from Tarifa, one would simply cross the narrow strait in a boat.’

C35	*Transformou-se*				*n-uma*		*pedra*
	transform.IND.PFV.3SG-REFL				in-INDF.ART.F.SG		stone.F
	que	***rolou***	***sobre***	***o***	***pé***	***d-o***	
	that	roll	over	DEF.ART.M.SG	foot.M	of-DEF.ART.M.SG	
	garimpeiro*.*						
	prospector.M						

‘He transformed himself into a stone that rolled over the prospector’s foot.’

C37 *Antes que o rapaz dissesse*
before that DEF.ART.M.SG boy.M say.SBJV.IPFV.3SG
alguma coisa, uma mariposa
any.F thing.F INDF.ART.F.SG moth.F
começou a ***esvoaçar entre ele e***
begin.IND.PFV.3SG to flutter.INF between 3SG.M and
o velho*.*
DEF.ART.M.SG old.man.M
'Before the boy could say anything, a butterfly started to flutter between him and the old man.'

C38 *O rapaz* ***andou durante quarenta***
DEF.ART.M.SG boy.M walk.IND.PFV.3SG during forty
dia-s pel-o deserto*, até chegar*
day.M-PL through-DEF.ART.M.SG desert.M until arrive.INF
a um belo castelo, n-o
to INDF.ART.M.SG beautiful.M castle.M in-DEF.ART.M.SG
alto de uma montanha.
top.M of INDF.ART.F.SG mountain.F
'The boy walked through the desert for forty days, until he arrived at a beautiful castle at the top of a mountain.'

C39 *Já mais tranqüilo, o rapaz*
already more calm.M DEF.ART.M.SG boy.M
pegou a colher e
take.IND.PFV.3SG DEF.ART.F.SG spoon.F and
voltou a ***passear pel-o***
return.IND.PFV.3SG to walk.INF through-DEF.ART.M.SG
palácio*, d-esta vez reparando em toda-s*
palace.M from-DEM.PROX.F time.F pay.attention to all.F-PL
a-s obra-s de arte que
DEF.ART.F-PL work.F-PL of art.F that
pendiam d-o teto e
hang.IND.IPFV.3PL from-DEF.ART.M.SG ceiling.M and
d-a-s parede-s.
from-DEF.ART.F-PL wall.F-PL
'Already calmer, the boy took the spoon and resumed his walking through the palace, this time paying attention to all the works of art that hang from the ceiling and the walls.'

C40

Enfiou	*a*	*mão*	*n-o*	
insert.IND.PFV.3SG	DEF.ART.F.SG	hand.F	in-DEF.ART.M.SG	
alforje	*e*	*ia*	*pegando*	
saddlebag.M	and	go.AUX.IND.IPFV.3SG	take.hold.of.PRS.PTCP	
uma	*d-a-s*	*pedra-s,*	*quando*	*ambas*
one.F	of-DEF.ART.F-PL	stone.F-PL	when	both.F
escorregaram	***por***	***buraco***	***n-o***	
slip.IND.PFV.3PL	through	hole.M	in-DEF.ART.M.SG	
***tecido**.*				
textile.M				

‘He inserted his hand in the saddlebag and wanted to grab one of the stones, when both of them slipped through a whole in the fabric.’

C41

Não	*pensava*	*costurar*	*o*	*buraco -*
NEG	think.IND.IPFV.3SG	sew.INF	DEF.ART.M.SG	hole.M
a-s	*pedra-s*	***poder-iam***	***escapar***	
DEF.ART.F-PL	stone.F-PL	can-IND.COND.3PL	escape.INF	
por	***ali***	*sempre*	*que*	*desejassem.*
through	there.DIST	always	that	desire.SBJV.IPFV.3PL

‘He did not even think about sewing the hole - the stones could escape through it whenever they desired.’

C42

Calçou	*a-s*	*sandália-s*	*nova-s,*	
put.on.IND.PFV.3SG	DEF.ART.F-PL	sandal.F-PL	new.F-PL	
e	***desceu***	***sem***	***fazer***	***qualquer***
and	descend.IND.PFV.3SG	without	make.INF	any
***ruído**.*				
sound.M				

‘He put on his new sandals and descended without making a sound.’

C43

Quando	*ele*	*tirou*	*o*			
when	3SG.M	take.off.IND.PFV.3SG	DEF.ART.M.SG			
casaco,	*pensando*	*em*	*dar*	*de*	*presente*	
coat.M	think.PRS.PTCP	in	give.INF	as	present.M	
para	*um*	*rapaz*	*n-a*	*rua,*		
to	INDF.ART.M.SG	boy.M	in-DEF.ART.F.SG	street.F		
a-s	*duas*	*pedra-s*	***rolaram***			
DEF.ART.F-PL	two.F	stone.F-PL	roll.IND.PFV.3PL			
pel-o	***chão**.*					
over-DEF.ART.M.SG	ground.M					

‘When he took of the coat, thinking to give it to some boy on the street, the two stones rolled over the ground.’

C44 *O* *rapaz* *e* *o* *Inglês*
DEF.ART.M.SG boy.M and DEF.ART.M.SG Englishman.M
haviam *comprado* *camelo-s,* *e*
have.AUX.IND.IPFV.3PL buy.PTCP camel.M-PL and
subiram ***com*** ***uma*** ***certa***
ascend.IND.PFV.3PL with INDF.ART.F.SG certain.F
dificuldade*.*
difficulty.F
'The boy and the Englishman had bought camals and they ascended them with some difficulty.'

C45 *- Tenho* *visto* *a* *caravana*
have.AUX.IND.PRS.1SG see.PTCP DEF.ART.F.SG caravan.F
caminhando ***através*** ***d-o*** ***deserto*** *-*
walk.PRS.PTCP across of-DEF.ART.M.SG desert.M
disse, *por* *fim.*
say.IND.PFV.3SG by end.M
'- I have seen the caravan walking across the desert, - he said finally.'

C46 *De* *repente,* *um* *gavião* ***deu***
of sudden.M INDF.ART.M.SG hawk.M give.IND.PFV.3SG
um ***rápido*** ***mergulho*** ***n-o***
INDF.ART.M.SG fast.M dive.M in-DEF.ART.M.SG
céu ***e*** ***atacou*** ***o*** ***outro****.*
sky.M and attack.IND.PFV.3SG DEF.ART.M.SG other.M
'Suddenly, one of the hawks made a fast swoop in the sky and attacked the other one.'

C48 *Levantou-se,* *e* *começou* *a* ***caminhar***
get.up.IND.PFV.3SG-REFL and begin.IND.PFV.3SG to walk.INF
em ***direção*** ***à-s*** ***tamareira-s****.*
in direction.F to.DEF.ART.F-PL date.palm.F-PL
'He got up and began to walk in the direction of the palm trees.'

C49 *... e* *interpretada-s* ***d-a*** ***maneira*** ***que***
and interpret.PTCP.F-PL of-DEF.ART.F.SG way.F that
caíam*.*
fall.IND.IPFV.3PL
'... and interprets the way in which they fell.'

C50 *Eram* *vinte* *minuto-s* *de* *caminhada*
be.IND.IPFV.3PL twenty minute.M-PL of walk.F
até *sua* *tenda,* *e* *ele* *começou* *a*
until 3SG.F.POSS tent.F and 3SG.M begin.IND.PFV.3SG to
andar*.*
walk.INF
'It was a twenty minutes walk to his tent, and he began to walk.'

C51 *Não pensou um minuto sequer*
NEG think.IND.PFV.3SG one.M minute.M even
em ***fugir****.*
in escape.INF
‘He did not even consider to escape for a minute.’

C52 *o rapaz* ***andou em direção***
DEF.ART.M.SG boy.M walk.IND.PFV.3SG in direction.F
a-o sul*.*
to-DEF.ART.M.SG south.M
‘the boy walked towards the south.’

C54 *O rapaz também* ***deu***
DEF.ART.M.SG boy.M also give.IND.PFV.3SG
um salto, só que para trás*.*
INDF.ART.M.SG jump.M only that towards back
‘The boy made a jump too, only backwards.’

C55 *O Alquimista desfez o*
DEF.ART.M.SG Alchemist.M undo.IND.PFV.3SG DEF.ART.M.SG
círculo n-o chão, e a
circle.M on-DEF.ART.M.SG ground.M and DEF.ART.F.SG
cobra ***correu e desapareceu***
snake.F sprint.IND.PFV.3SG and disappear.IND.PFV.3SG
entre a-s pedra-s*.*
between DEF.ART.F-PL stone.F-PL
‘The Alchemist wiped out the circle on the ground, and the snake quickly slithered away, disappearing between the stones.’

C56 ***Saiu d-a tenda porque***
go.out.IND.PFV.3SG of-DEF.ART.F.SG tent.F because
o Alquimista lhe segurou o-s
DEF.ART.M.SG Alchemist.M 3SG.INDR.OBJ hold.IND.PFV.3SG
braço-s*.*
DEF.ART.M-PL arm.M-PL
‘He got out of the tent only because the Alchemist was holding him by the arms.’

Sentences added to the 118-sentence sample for the 132-sentence sample:

A01 and then **hurried on**,
A02 Alice went on, half to herself, as she **swam lazily about in the pool**,
A03 ‘He took me for his housemaid’, she said to herself as she **ran**.
A04 The three soldiers **wandered about** for a minute or two,
A05 there's the arch I've got to go through next **walking about at the other end of the ground** -
A06 the frightened Mouse **splashed his way through the neighbouring pool**

C01

Começou	*a*	***andar***	***sem***	***pressa***
begin.IND.PFV.3SG	to	walk.INF	without	hurry.F
pel-a	***praça****.*			
across-DEF.ART.F.SG	square.F			

‘He began to walk across the square without any hurry.’

C02

Chamou	*o*	*rapaz*	*e*
call.IND.PFV.3SG	DEF.ART.M.SG	boy.M	and

começaram	*a*	***passear***	***pel-a-s***
begin.IND.PFV.3PL	to	take.a.walk.INF	through-DEF.ART.F-PL

duna-s	***em***	***volta***	***d-o***	***acampamento****.*
dune.F-PL	in	turn.F	of-DEF.ART.M.SG	camp.M

‘He called the boy and they began to walk around through the dunes around the camp.’

C03

Tinha	*visto*	*rei-s*	*e*
have.AUX.IND.IPFV.3SG	see.PTCP	king.M-PL	and

mendigo-s	***pisando***	***aquela-s***	***areia-s***
beggar.M-PL	step.PRS.PTCP	DEM.DIST.F-PL	sand.F-PL

que	*sempre*	*mudavam*	*de*	*forma*	*por*
that	always	change.IND.IPFV.3PL	of	form.F	by

causa	*d-o*	*vento,*	*mas*	*que*
cause.F	of-DEF.ART.M.SG	wind.M	but	that

eram	*a-s*	*mesma-s*	*que*
be.IND.IPFV.3PL	DEF.ART.F-PL	same.F-PL	that

havia	*conhecido*	*quando*	*criança.*
have.AUX.IND.IPFV.3SG	know.PTCP	when	child.F

‘He had seen kings and beggars walking on these sands that were always changing because of the wind, but that also had not changed since he was a child.’

C04

Começaram	*a*	***caminhar***	***pel-a-s***
begin.IND.PFV.3PL	to	walk.INF	through-DEF.ART.F-PL

areia-s*,*	*com*	*a*	*lua*	*ainda*
sand.F-PL	with	DEF.ART.F.SG	moon.F	still

brilhando	*sobre*	*o-s*	*dois.*
shine.PRS.PTCP	above	DEF.ART.M-PL	two.M

‘They began to walk through the sands with the moon still shining above them.’

C05 *Mas passar-á o resto d-o-s*
but spend-IND.FUT.3SG DEF.ART.M.SG rest.M of-DEF.ART.M-PL
seu-s dia-s ***vagando entre***
3SG.M.POSS-PL day.M-PL roam.PRS.PTCP between
a-s tamareira-s e o
DEF.ART.F-PL date.palm.F-PL and DEF.ART.M.SG
deserto*, sabendo que não cumpriu*
desert.M know.PRS.PTCP that NEG fulfill.IND.PFV.3SG
sua Lenda Pessoal, e que agora
3SG.F.POSS legend.F personal and that now
é tarde demais para isto.
be.IND.PRS.3SG late too.much for DEM.PROX.N
'But you will spend the rest of your days wandering around between the palms and through the desert, knowing that you did not fulfill your Personal Legend and that now it is much too late for that.'

C06 *O Alquimista* ***cavalgava***
DEF.ART.M.SG Alchemist.M ride.horseback.IND.IPFV.3SG
n-a sua frente*, com o*
at-DEF.ART.F 3SG.F.POSS front.F with DEF.ART.M.SG
falcão n-o-s ombro-s.
falcon.M on-DEF.ART.M-PL shoulder.M-PL
'The Alchemist rode in front of him, with the falcon on his shoulder.'

C08 *Durante uma semana* ***andaram*** *em*
during INDF.ART.F.SG week.F walk.IND.PFV.3PL in
silêncio, conversando apenas a-s
silence.M converse.PRS.PTCP only DEF.ART.F-PL
precauções necessária-s para evitar
precaution.F.PL necessary.F-PL to avoid.INF
o-s combate-s entre o-s
DEF.ART.M-PL battle.M-PL between DEF.ART.M-PL
clã-s.
clan.M-PL
'They walked in silence for a week, talking only about the precautions needed to avoid the battles between the tribes. '

C09 *O rapaz* ***caminhou*** *duas hora-s*
DEF.ART.M.SG boy.M walk.IND.PFV.3SG two.F hour.F-PL
e meia ***pel-o deserto****, procurando*
and half.F through-DEF.ART.M.SG desert.M seek.PRS.PTCP
escutar atenta-mente o que
listen.INF attentive.F-ADV DEF.ART.M.SG that
seu coração dizia.
3SG.M.POSS heart.M say.IND.IPFV.3SG
'The boy walked through the desert for two and a half hours, listening carefully to what his heart had to say.'

Sentences added to the 118-sentence sample for the 192-sentence sample:

B11 Let's pretend the glass has got all soft like gauze, so that we **can get through**.
B12 She was up on the chimney-piece while she said this, though she hardly knew how she **had got there**.
B13 Mind you **come up – the regular way -**
B14 till they **got to the top of the little hill**.
B15 and Alice looked on with great interest as she **returned to the tree**,
B16 How it happened, Alice never knew, but exactly as she **came to the last peg**, she was gone.
B17 I think I'll **go down the other way**,
B18 At last he said, 'You're travelling the wrong way,' and shut up the window and **went away**.
B19 In another moment she felt the carriage **rise straight up into the air**,
B20 She was rambling on in this way when she **reached the wood**:
B21 till they **came out into another open field**,
B22 She **went on and on**, a long way,
B23 till, on **turning a sharp corner**, she came upon two fat little men,
B24 Alice said very politely, as she **crossed the little brook after the Queen**.
B25 after she had spent a minute or so in vainly **pursuing a large bright thing**,
B26 the `thing' **went through the ceiling as quietly as possible**, as if it were quite used to it.
B27 but she couldn't help saying to herself as she **went**,
B28 Then **came** the horses.
B29 At this moment the Messenger **arrived**:
B30 Who did you **pass on the road**?
B32 that Alice **got behind a tree** to be out of the way of the blows.
B33 And part of the roof came off, and ever so much thunder **got in -**
B34 - and then she **scrambled back into the arm-chair**, taking the kitten and the worsted with her,
B35 this led to a scramble, in which the ball **rolled down upon the floor**,
B36 and really I might have won, if it hadn't been for that nasty Knight, that **came wiggling down among my pieces**.
B37 and had **jumped lightly down into the Looking-glass room**.
B38 the White Queen cried out as she **rushed past the King**,
B39 and she **began scrambling wildly up the side of the fender**.
B40 Alice watched the White King as he **slowly struggled up from bar to bar**,
B41 She just kept the tips of her fingers on the hand-rail, and **floated gently down without even touching the stairs with her feet**;
B42 then **she floated on through the hall**,

B43 And so she did: **wandering up and down**,
B44 and found herself **walking in at the front-door** again.
B45 and then **began slowly walking down the row.**
B46 and then **hurried on to the last**.
B47 So with this excuse she ran down the hill and **jumped over the first of the six little brooks**.
B48 `It's only **a brook we have to jump over**.'
B49 The Gnat amused itself meanwhile by **humming round and round her head**:
B50 But the Gnat only sighed deeply, while two large tears **came rolling down its cheeks**.
B51 she said as she **stepped under the trees**,
B52 Just then a Fawn **came wandering by**:
B53 and here the Fawn **gave a sudden bound into the air**,
B54 and in another moment it **had darted away at full speed**.
B56 the other sulkily replied, as he **crawled out of the umbrella**:
B57 Alice **ran a little way into the wood**, and stopped under a large tree.
B58 in another moment the White Queen **came running wildly through the wood**, with both arms stretched out wide, as if she were flying,
B59 The crow must have **flown away**, I think,
B60 The Queen spread out her arms again, and **went flying after it**,
B61 till it **glided gently in among the waving rushes**.
B62 with flushed cheeks and dripping hair and hands, she **scrambled back into her place**,
B63 she said `Good-bye!' once more, and, getting no answer to this, she **quietly walked away**:
B64 The next moment soldiers **came running through the wood**,
B65 `But he's coming very slowly -- and what curious attitudes he goes into!' (For the messenger **kept skipping up and down, and wriggling like an eel, as he came along**, with his great hands spread out like fans on each side.)
B66 Alice had no more breath for talking, so they **trotted on** in silence, till they came in sight of a great crowd,
B68 And Hatta **went bounding away like a grasshopper**.
B69 There's the White Queen **running across the country**!
B70 She **came flying out of the wood over yonder** -
B71 At this moment the Unicorn **sauntered by them**, with his hands in his pockets.
B72 She started to her feet and **sprang across the little brook in her terror**,
B73 and a Knight dressed in crimson armour, **came galloping down upon her, brandishing a great club**.
B74 the Knight cried, **as he tumbled off his horse**.

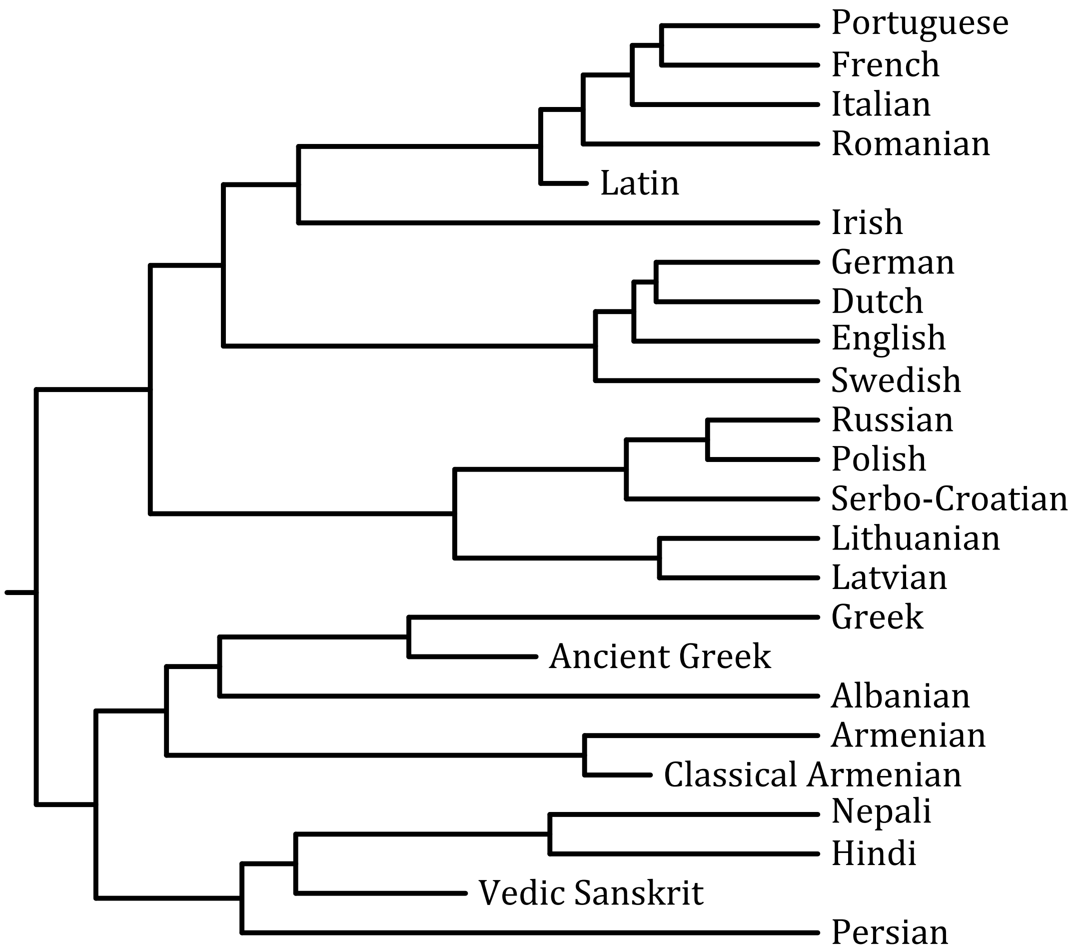
B75 and then the Red Knight mounted and **galloped off**.
B76 Whenever the horse stopped (which it did very often), he **fell off in front**;
B77 and whenever it went on again (which it generally did rather suddenly), he **fell off behind**.
B78 He raised his hands in some excitement as he said this, and instantly **rolled out of the saddle**,
B79 and **fell headlong into a deep ditch**.
B80 Alice **ran to the side of the ditch** to look for him.
B81 So they shook hands, and then the Knight **rode slowly away into the forest.**
B82 So she went on talking to herself, as she watched the horse **walking leisurely along the road**,
B83 she cried as she **bounded across**,
B85 exclaimed Alice, looking about in great perplexity, as first one round head, and then the other, **rolled down from her shoulder**, and lay like a heavy lump in her lap.
B86 but at last, a very old Frog, who was sitting under a tree, got up and **hobbled slowly towards her**
B89 and three of them (who looked like kangaroos) **scrambled into the dish of roast mutton**,
B90 she cried as she jumped up and seized the table-cloth with both hands: one good pull, and plates, dishes, guests, and candles **came crashing down together in a heap on the floor**.

Sentences added to the 192-sentence sample to create the 215-sentence sample (in addition to A01–A06, C01–C06, and C08–C09 added above):

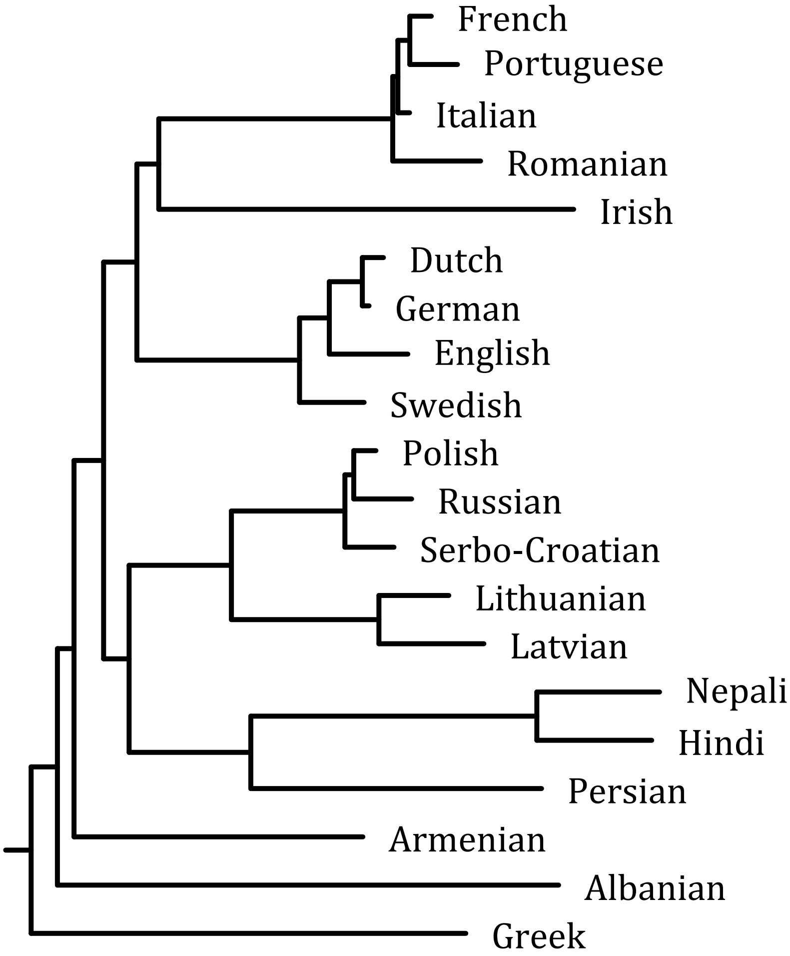
B02 and they **walked on** in silence
B03 And they **ran on** for a time in silence, with the wind whistling in Alice's ears, and almost blowing her hair off her head, she fancied.
B04 And they went so fast that at last they seemed to **skim through the air, hardly touching the ground with their feet**,
B05 **Crawling at your feet**,
B06 So she **wandered on**, talking to herself as she went,
B07 and she found they were in a little boat, **gliding along between banks**:
B08 So the boat was left to **drift down the stream as it would**,
B09 Alice thought, and for a few minutes she **walked on** in silence,
B10 As to the bottles, they each took a pair of plates, which they hastily fitted on as wings, and so, with forks for legs, **went fluttering about in all directions**:

Appendix 4. Maximum clade credibility trees used for robustness tests in ancestral state estimation

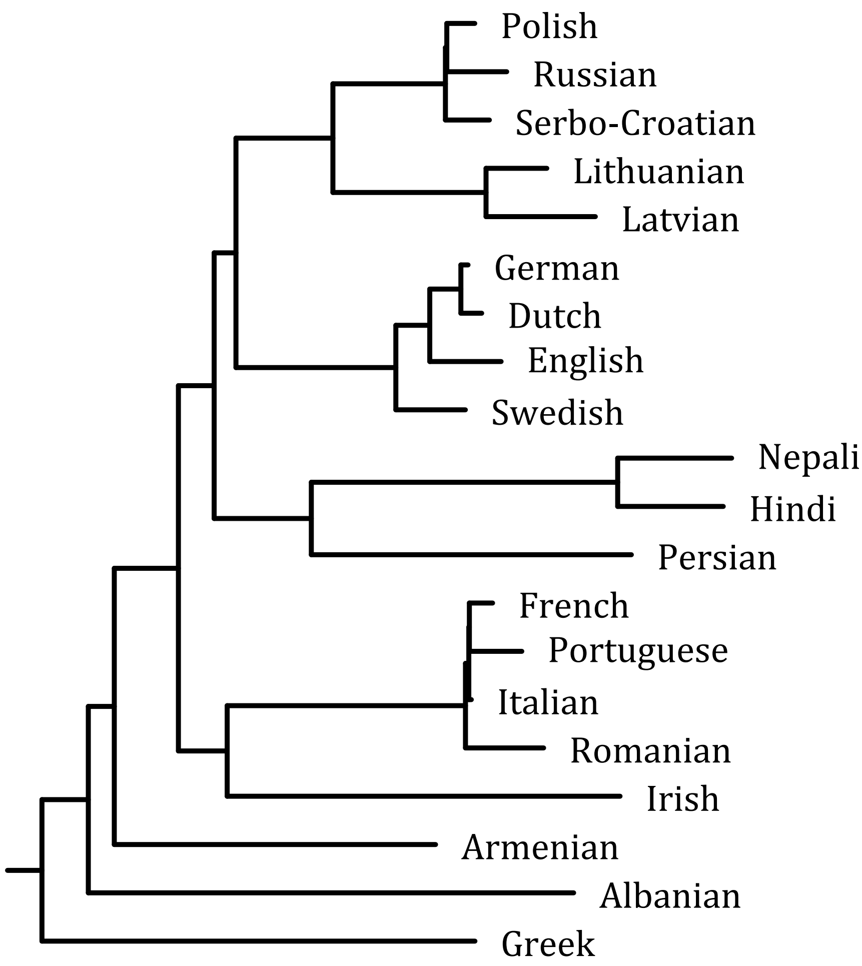
Maximum clade credibility tree of Bouckaert et al. (2012)'s tree sample including four ancient Indo-European languages:



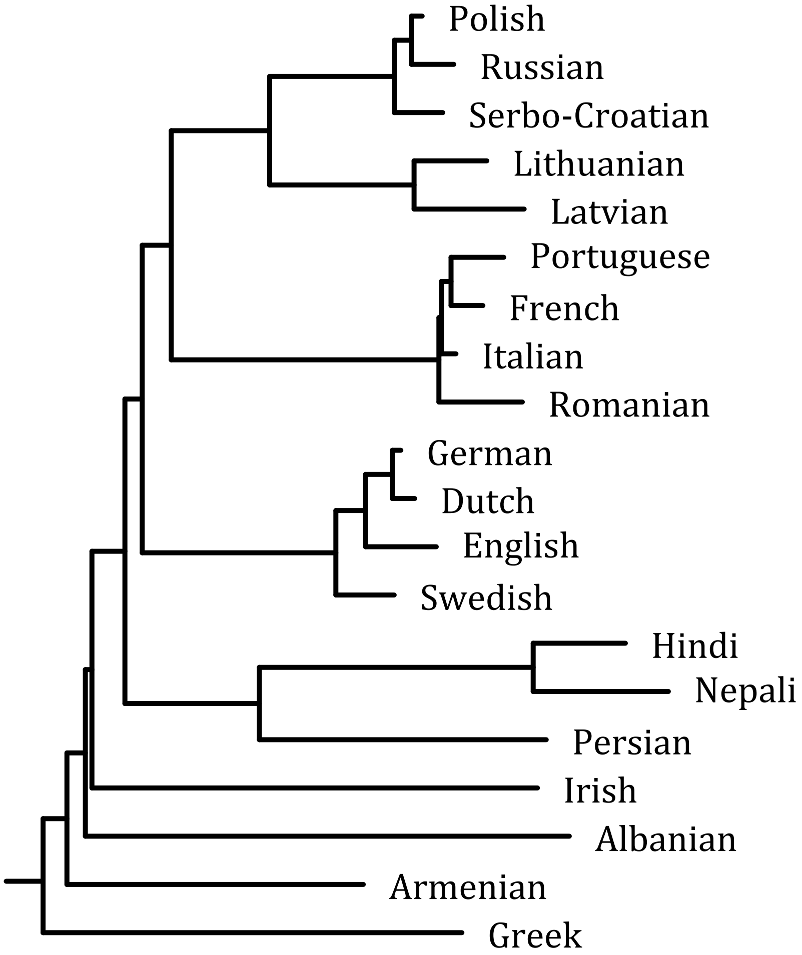
Maximum clade credibility tree of Dunn et al. (2011)'s full tree set:



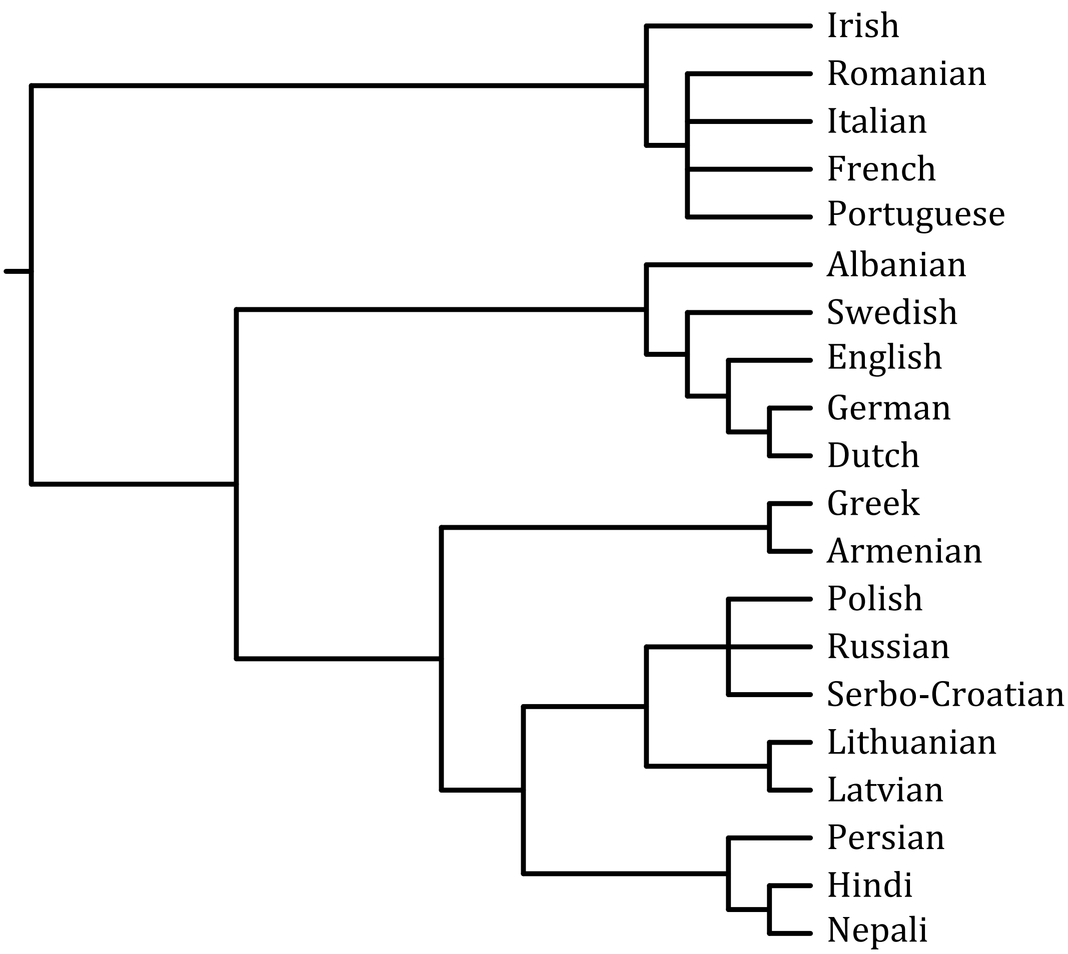
Maximum clade credibility tree of Dunn et al. (2011)'s trees in which Balto-Slavic and Germanic form a subgroup:



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



Nakhleh et al. (2005a)'s tree A:



Appendix 5: Manner verbs in Indo-European

Notes regarding appendices 5-7

Appendices 5 and 6 present an overview of the types of manner verbs used by the 20 languages in the sample. Appendix 7 presents an overview of the non-motion verbs used as manner verbs. In these appendices, types of verbs that occurred in the 132-sentence sample are presented in plain text. Types of verbs that occurred only in the 215-sentence sample are presented in italics. This distinction is made in order to give an overview of manner verbs in these 20 languages that is as extensive as possible, while also making clear what the size of the manner verb lexicon is on the basis of the 132-sentence sample and on the basis of the 215-sentence sample. The number of types of manner verbs each language has have been provided in Table 4.3. The manner verbs have been placed in the following categories (taken from Narasimhan 2003): mode (*walk*), medium (*fly*), velocity (*rush*), attitude and display (*sneak*), contact (*slide*), course (*wander*), and vehicle (*ride*).

Appendix 5a: Manner verbs in Romance

French	**Portuguese**	**Italian**	**Romanian**
manner type: mode			
ramper ‘crawl’	rastejar ‘crawl’	strisciare ‘crawl’	
			rostogoli ‘tumble’
rouler ‘roll’	rolar ‘ roll’ *rebolar* ‘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’	passeggiare ‘stroll’	plimba ‘stroll’
			tropăi ‘tramp’
boitiller ‘limp’	*coxear ‘limp’*		*șchiopăta ‘limp’*
fouler ‘tread’	pisar ‘tread’		pășì ‘step’
bondir ‘jump, rush’ sauter ‘jump’	pular ‘jump’ saltar ‘leap’	balzare ‘jump’ saltare ~ saltellare ~ salterellare ‘jump’	sări ‘jump’ topăi ‘hop’

French	Portuguese	Italian	Romanian
manner type: animal mode			
	galopar 'gallop'	*galoppare 'gallop'*	
trotter 'trot'	trotar 'trot'	trottare ~ trotterellare 'trot'	
manner type: medium			
flotter 'float'	*flutuar 'float'* *vogar 'sail, float'*		*pluti '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'
manner type: contact			
glisser 'slide'	esgueirar-se 'slip' escorregar 'slip' *deslizar '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'

Appendix 5b: Manner verbs in Germanic

English	Dutch	German	Swedish
manner type: mode			
crawl creep	kruipen 'crawl'	kriechen 'crawl'	krypa 'crawl'
scramble	*klauteren 'scramble'*	*klettern 'scramble'*	
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'

English	Dutch	German	Swedish
saunter	kuieren 'stroll' wandelen 'stroll' trekken 'hike' *slenteren 'saunter'*	schlendern 'stroll' spazieren 'stroll'	vanka 'saunter'
	huppelen 'hop' trippelen 'patter'		
hobble	*hobbelen 'hobble'*	hoppeln 'lollop'	*linka 'limp'*
step	*stappen 'step'*	*schreiten 'step'* treten 'step'	
jump leap *spring* *bound*	springen 'jump' *hoppen 'hop'*	springen 'jump' *hüpfen 'hop'* *setzen 'jump'*	hoppa 'jump' *skutta 'leap'*
manner type: animal mode			
gallop	*galloperen 'gallop'*	*sprengen 'gallop'*	*gallopera 'gallop'*
trot	draven 'trot'	*traben 'trot'* trotten 'trot'	trava 'trot' lunka 'trot'
manner type: medium			
float *drift*	*drijven 'float'* *zweven 'hover'* dwarrelen 'twirl'	*treiben 'float'*	*sväva 'hover'* *driva '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 *dart*	*zich haasten 'hurry'* spurten 'scurry'	eilen 'hurry' schnellen 'hurry' stürzen 'dash'	skynda 'hurry' rusa 'rush'
manner type: attitude and display			
march	marcheren 'march'	marschieren 'march'	marschera 'march'
	sluipen 'sneak'		smyga 'sneak, creep'

English	Dutch	German	Swedish
manner type: contact			
slither slip *glide* *drift*	glijden 'glide' *schuiven 'glide'*	gleiten 'glide'	glida 'glide' ringla 'slither'
manner type: course			
wander	dwalen 'wander' *zwerven 'roam'*	wandern 'wander' umherstreifen 'wander' ziehen 'wander'	vandra 'wander'
manner type: vehicle			
ride	rijden 'ride'	reiten 'ride'	rida 'ride'

Appendix 5c: 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' *ropoti 'crawl'*	līst 'crawl'
	leźć 'scramble'		ropšti 'scramble' *repečkenti 'scramble, crawl'*	*rausties 'scramble'*
	koziołkować się 'tumble'		*versti 'roll, tumble'*	kūleņot 'somersault'
katit'sja 'roll'	toczyć 'roll' *turlać się 'roll'*	kotrljati 'roll'	riedėti 'roll' *risti 'roll'*	ripot 'roll' *velties 'roll'*
			lipti 'climb' sliuogti 'climb'	kāpt 'climb' rāpties 'scramble, climb'
bežat' 'run'	biec 'run'	trčati 'run'	bėgti 'run'	skriet 'run'
idti ~ hodit' 'walk, go'	iść ~ chodzić 'walk, go'	hodati 'walk, stroll'	eiti 'walk, go' vaikščioti 'walk'	iet 'walk, go' *kumurot 'walk'*

Russian	**Polish**	**Serbo-Croatian**	**Lithuanian**	**Latvian**
guljat' 'stroll'	spacerować 'stroll'	šetati 'stroll'	*pėdinti 'pace'*	pastaigāties 'stroll' cilpot 'pace'
plestis' 'trudge'	*człapać 'waddle'* *wlec się 'trudge'*		*kėblinti 'drag oneself along'*	*slāt 'strut'*
kovyljat' 'hobble'	tuptać 'toddle'	gegati 'waddle' *šepesati 'hobble'*	*šlubuoti 'limp'*	tipināt 'toddle' *steberēt 'hobble'*
stupat' ~ stupit' 'tread'		koračati 'tread' *stupiti 'step'*	žengti 'step' mindyti 'tread'	mīt 'tread'
prygnut' ~ prygivat' 'jump' *skakat' 'hop'*	skoczyć 'jump'	skakati ~ skočiti ~ skakutati 'jump'	šokti 'jump'	lēkt ~ ļekāt ~ linkāt 'jump'
manner type: animal mode				
	cwałować 'gallop'	*galopirati 'gallop'*	*šuoliuoti 'gallop'*	*auļot 'gallop'*
trusit' 'trot' rysit' 'trot'		kaskati 'trot'	risnoti 'trot' tursenti 'trot'	
manner type: medium				
		plutati 'float' *lebdjeti 'float'*	*sklęsti 'hover, float'*	lidināties 'flutter, hover'
letet' ~ letat' 'fly'	lecieć 'fly' frunąć 'fly'	letjeti 'fly'	skristi 'fly' *skrieti 'fly'*	lidot 'fly' laisties 'fly'
plavat' ~ pl'it' 'swim'	pływać 'swim'	plivati 'swim'	plaukti 'swim'	peldēt 'swim'
				spraukties 'wring (through)'

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
manner type: velocity				
mčat' ~ mčat'sja 'rush' rvanut'sja 'dash' *nestis' ~ neslas' 'dash'*	spieszyć 'hurry' pędzić 'bolt' gnać 'rush' mknąć 'zoom' smyrgnąć 'dash'	hitati 'hasten' juriti ~ jurnuti 'rush' navaliti 'rush' zuriti 'hurry' strugnuti 'dash' *sunuti 'dart'*	skubėti 'hurry' lėkti 'dash' šmurkštelėti 'dart' nerti 'rush' *mauti 'dash'* *lakstyti 'scurry'*	steigties 'hurry' traukties 'rush' mesties 'dash' drāzties 'dash' šauties 'dart'
manner type: attitude and display				
šagat' 'march' *šestvovat' 'walk in procession'*	maszerować 'march'			soļot 'march'
				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' *vrludati 'wander'*	klaidžioti 'wander'	klīst 'wander'
manner type: vehicle				
ehat' 'ride'	jechać 'ride'	jahati 'ride'	joti 'ride'	jāt 'ride'

Appendix 5d. Manner verbs in Indo-Iranian

Hindi	Nepali	Persian
manner type: mode		
rengnā 'crawl'		ḵazidan 'creep'
luṛhaknā 'roll'	gudnu 'roll' ladnu padnu 'roll'	ġaltāndan ~ ġaltidan 'roll'
caṛhnā 'climb'		

Hindi	Nepali	Persian
dauṛnā 'run' bhāgnā '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'
kūdnā 'jump' chalāṃg lagnā 'jump'	uphranu 'jump'	jastan ~ jast zadan 'jump' jahidan 'leap' kizan 'jump'
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		
		laġzidan 'slide'
manner type: course		
bhaṭaknā 'wander' ghūmnā 'roam'	bhautārinu 'wander' dulnu 'wander'	
manner type: vehicle		
	haknu 'ride'	

Appendix 5e: Manner verbs in Albanian, Armenian, Modern Greek and Irish

Albanian	**Armenian**	**Modern Greek**	**Irish**
manner type: mode			
	soghal 'crawl'	*mpoysoylo 'crawl'* sernomai 'creep'	*lámhacán 'crawl'*
	glorvel 'roll'	kylo 'roll' strifogyrizo 'roll, whirl'	*rothlaigh 'roll'* *rollail 'roll'* *sil 'trickle'*
			dreap 'climb'
vrapoj 'run'	vazel 'run'	trecho 'run'	rith 'run'
eci 'walk'	k'aylel 'walk'	perpato 'walk'	siúil 'walk'
shëtis 'stroll' çapitet 'stride'		*peridiavaino 'stroll'* *alafropato 'trudge'*	
			bacadaíl 'limp'
shkel 'step'		pato 'step'	
kërcej 'jump'	ts'atkel 'jump'	pidao 'jump'	léim 'jump' *preab 'jump'*
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			
nxitoj 'hurry' u sulem 'rush'	shtapel 'hurry' slanal 'rush' sural 'dash' khoyanal 'rush'	speydo 'hurry' viazomai 'hurry' ormo 'rush'	deifir 'hurry' brostaigh 'hurry'
manner type: attitude and display			
u zvarrit 'drag onself'		vadizo 'walk, march'	máirseáil 'march'
manner type: contact			
shkas 'slide'	sahel 'glide'	glistro 'slide'	sleamhnaigh 'slide' sciorr 'slide'

Albanian	**Armenian**	**Modern Greek**	**Irish**
manner type: course			
sillet ‘wander’ baret ‘wander’	t’ap’arrel ‘wander’	trigyrizo ‘wander’ planiemai ‘wander’	bheith ag fánaíocht ‘wander’ *fálróid ‘wander’*
manner type: vehicle			
			marcaigh ‘ride’

Appendix 6: Complex manner verbs in Indo-European

Appendix 6a: Complex 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’ *dare un balzo* *‘jump’* *fare un salto* *‘jump’* *spiccare il salto* *‘leap’*	face un salt ‘jump’
			da buzna ‘rush’
	deixar à deriva ‘leave to drift’		
		mettersi in marcia ‘march’	

Appendix 6b: Complex 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 *give a bound*		einen Satz machen ‘make a jump’	
make a rush			göra en rusning ‘make a rush’

Appendix 6c: Complex 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'

Appendix 6d: Complex manner verbs in Indo-Iranian

Hindi	**Nepali**	**Persian**
	ghodā daudinu 'horse ride'	

Appendix 6e: Complex manner verbs in Albanian, Armenian, Modern Greek and Irish

Albanian	**Armenian**	**Modern Greek**	**Irish**
	vaz tal 'run'		
bëj një kërcim 'jump'	ts'atk anel 'jump' vostyun anel 'jump'	kano ena pidima 'jump'	*tabhair léim 'jump'*
			tabhair fogha 'rush'

Appendix 7: Non-motion verbs used as manner verbs in Indo-European

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

Appendix 7a: Non-motion verbs used as manner verbs in Romance

	French	**Portuguese**	**Italian**	**Romanian**
1.	se tortiller 'wriggle'	contorcer-se 'wriggle' esvoaçar 'flutter'	svolazzare 'flutter'	flutura 'flutter'
2.	*bourdonner 'hum'*		*ronzare 'hum'*	pleoscăi 'splash' *zbârnâi 'whiz'*
3.	avoir un peu de mal 'have a bit of trouble' aider 'help' *démener 'struggle'*			*lupta 'fight'*
4.	*se faufiler 'slip (through)'*			
5.				iuţi 'go faster'

Appendix 7b: Non-motion verbs used as manner verbs in Germanic

	English	**Dutch**	**German**	**Swedish**
1.	flutter wriggle *wiggle*	fladderen 'flutter' kronkelen 'wriggle' *wringen 'wring'*	flattern 'flutter'	fladdra 'flutter' *flaxa 'flutter'* *slingra 'wriggle'*
2.	pop rattle *hum* *crash*	kletteren 'rattle' *zoemen 'hum'*	platschen 'splash' knattern 'rattle' prasseln 'crackle'	plaska 'splash' *surra 'hum'*

	English	Dutch	German	Swedish
2.			*summen 'hum'* *rauschen 'murmur'*	
3.	help *struggle*	*werken 'work'*		hjälpa 'help'
4.	skim	stuiven 'blow' schieten 'shoot' scheren 'shave'	stürmen 'storm'	

Appendix 7c: Non-motion verbs used as manner verbs in Balto-Slavic

	Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
1.	kružit'sja 'flutter'	zatrzepotać 'flutter' *wić się 'wriggle'*	lepršati 'flutter' *đipati 'frisk'* *kobeljati 'wriggle'*	trenkti 'strike' *plasnoti 'flutter'* *pleventi 'flutter'* *klibinti 'wobble'*	ķepuroties 'flounder' laidelēties 'fly about' locīties 'wriggle' griezties 'wriggle' *ietrausties 'snuggle'*
2.		prysnąć 'sprinkle' rąbnąć 'bang' gwizdnąć 'whistle' *bzykać 'hum'*	zviždati 'whistle' *zujati 'hum'*	švilpti 'whistle' švysteleti 'flash' *zvimbti 'hum'* *dardėti 'crash'*	
4.				šauti 'shoot'	*šūpot 'swing'*
5.	*medlit' 'slow down'*				

Appendix 7d: Non-motion verbs used as manner verbs in Indo-Iranian

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

Appendix 7e: Non-motion verbs used as manner verbs in Albanian, Armenian, Modern Greek and Irish

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

Appendix 8: Number of types of manner verb per novel

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

Appendix 9: Path verbs in Indo-European

In this appendix, verbs that occurred in the 118-sentence sample are presented in plain text. Verbs that occurred only in the 192-sentence sample are presented in italics. This distinction is made in order to give an overview of path verbs in these 20 languages that is as extensive as possible, while also making clear which verbs appear only in the 192-sentence sample. The number of path verbs each language has have been provided in Table 5.3. The path verbs have been placed in the following categories (partly taken from Narasimhan 2003):

1. complex endpoint (encodes features of the endpoint, e.g. whether it is an enclosure or a moving object): *enter*, *follow*, *land*
2. complex source-point (encodes features of the source point, e.g. whether it is an enclosure): *exit*, *escape*
3. simple endpoint: *arrive*
4. simple potential endpoint: *head for*, *approach*
5. simple source-point: *leave*
6. mid-point: *pass*
7. direction up: *go up*, *rise*
8. direction down: *go down*, *fall*, *dive*
9. direction forward: *advance*
10. direction around: *go around*, *turn*
11. direction back: *return*
12. direction behind: *go behind*
13. spatial features of the path: *cross*, *penetrate*

Appendix 9a: Path verbs in Romance

French	**Portuguese**	**Italian**	**Romanian**
complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object			
entrer 'enter'	entrar 'enter' enfiar-se 'enter'	entrare 'enter'	intra 'enter'
suivre 'follow' *poursuivre 'pursue'* *traquer 'pursue'*	seguir 'follow' *perseguir 'pursue'*	seguire 'follow' accodarsi 'follow' *inseguire 'pursue'*	urma 'follow' *urmări 'pursue'*
atterrir 'land'	aterrar 'land'	atterrare 'land'	

French	Portuguese	Italian	Romanian
complex source-point; encodes features of the source point, e.g. whether it is an enclosure			
	sair 'go out'	uscire 'go out'	ieși 'go out'
fuir 'flee' s'échapper 'escape'	fugir 'flee' escapulir-se 'escape' escaper 'escape' deixar 'abandon' abandonar 'leave'	lasciare 'leave' scappare 'escape' (s)fuggire 'flee'	scăpa 'escape' părăsi 'abandon'
simple endpoint			
arriver 'arrive' *atteindre 'reach'*	chegar 'arrive'	(rag)giungere 'arrive' arrivare 'arrive' *ficcarsi 'get to'*	ajunge 'arrive'
simple potential endpoint			
se diriger 'head for'	dirigir-se 'go to'		(se) lua 'take oneself' se îndrepta 'go toward'
s'approcher 'approach'	aproximar-se 'approach' *abeirar-se 'draw near'*	avvicinarsi 'approach'	se apropia 'come near' se alătura 'approach'
simple source-point			
partir 'leave' sortir 'leave' quitter 'leave' s'éloigner 'depart'	partir 'leave' afastar-se 'distance oneself' distanciar-se 'distance oneself'	partire 'leave' andarsene 'go away' allontanarsi 'go away' *avviarsi 'set out'*	pleca 'leave' (în)depărta 'leave' porni 'depart' se duce 'leave'
midpoint			
passer 'pass' *franchir 'pass'*	passer 'pass'	passare 'pass'	trece 'pass'
direction up			
monter 'go up' *s'élever 'rise'*	subir 'go up' erguer-se 'rise'	montare 'go up' sollevarsi 'rise' (ri)salire 'go up'	sui 'go up' se ridica 'rise' urca 'go up'
direction down			
(re)tomber 'fall' *basculer 'fall'*	cair 'fall'	cadere 'fall'	cădea 'fall'

French	Portuguese	Italian	Romanian
descendre 'come down'	descer 'descend'	scendere 'descend'	coborî 'descend'
	mergulhar 'dive'		cufunda 'dive'
direction forward			
(s')avancer 'go forward'		avanzare 'move forward'	
direction around			
contourner 'go around'	contornar 'go around'		ocoli 'go around'
	virar 'turn'	(ag)girare 'turn' *svoltare 'turn'*	coti 'turn'
direction back			
repartir 'return' revenir 'return' retourner 'return' rentrer 'return' rejoindre 'return' *regagner 'return'*	regressar 'return' voltar 'return' retornar 'return'	(ri)tornare(se) 'return'	se întoarce 'return'
spatial features of the path			
traverser 'cross' *croiser 'cross'*	atravessar 'cross' cruzar 'cross'	(at)traversare 'cross'	traversa 'cross' străbate 'cross'
pénétrer 'penetrate'	penetrar 'penetrate'	penetrare 'penetrate' *inoltrarsi 'penetrate'*	pătrunde 'penetrate'

Appendix 9b Path verbs in Germanic

English	German	Dutch	Swedish
complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object			
pursue follow	folgen 'follow' *vervolgen 'pursue'*	volgen 'follow' *achtervolgen 'pursue'*	följa 'follow'
			hamna 'land'

English	**German**	**Dutch**	**Swedish**
complex source-point; encodes features of the source point, e.g. whether it is an enclosure			
flee	verlassen 'leave'	ontsnappen 'escape' verlaten 'leave' vluchten 'flee'	fly 'flee'
simple endpoint			
reach get arrive	gelangen 'reach' eintreffen 'arrive' *erreichen 'reach'*	arriveren 'arrive' *bereiken 'reach'*	*anlända 'arrive'*
simple potential endpoint			
	sich nähern 'approach'	*aanzetten 'approach'*	*närma sig 'approach'*
simple source-point			
leave	sich begeben 'set off' sich entfernen 'depart' *weichen 'leave'*	zich begeven 'set off'	avlägsna sig 'remove oneself' lämna 'leave' ta sig 'take oneself'
midpoint			
pass	*überholen 'pass'*	*passeren 'pass'*	*passera 'pass'*
direction up			
rise	sich erheben 'rise'	stijgen 'rise'	*lyfta 'take off'*
direction down			
descend fall	fallen 'fall'	vallen 'fall'	falla 'fall'
	eintauchen 'dive'		
direction around			
turn		omslaan 'turn' *draaien 'turn'*	vika 'turn'
direction back			
return	zurückkehren 'return'	terugkeren 'return'	återvända 'return'
spatial features of the path			
cross	durchqueren 'cross' überqueren 'cross'	oversteken 'cross'	

English	German	Dutch	Swedish
	eindringen 'penetrate'	doordringen 'penetrate'	

Appendix 9c: Path verbs in Balto-Slavic

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

Russian	**Polish**	**Serbo-Croatian**	**Lithuanian**	**Latvian**
blizit’sja ‘approach’	zbliżać się ‘approach’	pridolaziti ‘approach’ se približiti ‘approach’		
simple source-point				
dvinut’sja ‘set out’ brat’sja ‘take oneself’ *skryt’sja ‘go away’*	opuścić ‘leave’ wybrać się ‘set out’ oddalić ‘go away’ ruszyć ‘set out’	otići ~ odlaziti ‘leave’ izaći ‘leave’ poći ‘leave’ se ukloniti ‘go away’ se uputiti ‘set out’ ostaviti ‘leave’ se udaljiti ‘go away’ krenuti ‘set out’	tolti ‘go away’ sileisti ‘let oneself go’	doties ‘set out’ attālināties ‘go away’ atstāt ‘leave’
midpoint				
minovat’ ‘pass’	*minąć ‘pass’*	prolaziti ~ proći ‘pass’		
direction up				
podnjat'sja ~ podnimat’sja ‘go up’		dignuti ~ *dizati* ‘rise’ se penjati ‘rise’ se popeti ‘go up’	kilti ‘rise’	
direction down				
padat’ ~ past’ ‘fall’ valit’sja ‘fall’	padać ‘fall’	padati ~ pasti ‘fall’ se rušiti ‘fall’	kristi ‘fall’ virsti ‘fall’ byrėti ‘fall’	krist ‘fall’ birt ‘fall’
spuskat’sja ~ spustit’sja ‘go down’		sići ~ silaziti ‘go down’		
	nurkować ‘dive’	zaroniti ‘dive’		nirt ‘dive’

Russian	Polish	Serbo-Croatian	Lithuanian	Latvian
direction around				
kružit' 'go round' ogibat' 'go round'	krążać 'go around' zbaczać 'go around'	zaobići 'go around'	lenkti 'go around'	*riņķot 'circle'*
	kręcić 'turn'		sukti 'turn' vingiuoti 'turn'	griezties 'turn'
direction back				
vozvrašat'sja 'return' vernut'sja 'return'	wrócić 'return'	vratiti ~ vraćati 'return'	grįžti 'return'	
direction behind				
		zalaziti 'go behind'		
spatial features of the path				
pereseč' ~ peresekat' 'cross'	przeprawiać 'cross' *przekraczać 'cross'*	prelaziti ~ prijeći 'cross'		šķērsot 'cross'
		prodrijeti 'penetrate'		

Appendix 9d: Path verbs in Indo-Aryan

Hindi	Nepali	Persian
complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object		
samānā 'go in'	pasnu 'go in'	dāḵel šodan 'go in' vāred šodan 'go in' darāmadan 'come in'
complex source-point; encodes features of the source point, e.g. whether it is an enclosure		
nikalnā 'go out'	niskanu 'go out'	ḵārej šodan 'go out'
tyāgnā 'abandon'	chodnu 'abandon'	tark kardan 'abandon'
simple endpoint		
pahuṃcnā 'arrive'	pugnu 'reach'	(az rāh) residan 'arrive'
simple potential endpoint		
		nazdik šodan 'approach'

Hindi	Nepali	Persian
simple source-point		
choṛnā ‘leave’		dur sakhtan ‘go away’ vāgozāštan ‘leave’
		(be) rāh oftādan ‘set out’
midpoint		
guzarnā ‘pass’		gozaštan ‘pass’
direction up		
	uklinu ‘go up’	barḵāstan ‘rise’
direction down		
girnā ‘fall’	khasnu ‘fall’ jharnu ‘fall’ barsinu ‘fall’	oftādan ‘fall’
utarnā ‘go down’	orlinu ‘go down’ dubnu ‘go down’	
direction forward		
calnā ‘go forward, advance’ baṛhnā ‘advance’	aghi badhnu ‘advance’	
direction around		
ghūmnā ‘turn’		dowr zadan ‘go around’
direction back		
	pharkinu ‘return’	bāzgaštan ‘return’ bāzāmadan ‘return’
spatial features of the path		
pār karnā ‘cross’	tarnu ‘cross’ par garnu ‘cross’	
		nofuz kardan ‘penetrate’

Appendix 9e: Path verbs in Albanian, Armenian, Modern Greek and Irish

Albanian	Armenian	Modern Greek	Irish
complex endpoint: encodes features of the endpoint, e.g. whether it is an enclosure or a moving object			
futem ‘go in’ hyj ‘go in’	mtnel ‘go in’	mpaino ‘go in’	
ndjek ‘follow’	hetevel ‘follow’ hetapndel ‘pursue’	akoloytho ‘follow’ kynigo ‘chase’	lean ‘follow’ *tóraigh ‘pursue’*
		egkataleipo ‘abandon’	

Albanian	**Armenian**	**Modern Greek**	**Irish**
complex source-point; encodes features of the source point, e.g. whether it is an enclosure			
dal 'go out'		vgaino 'go out'	
iki 'escape'			éalaigh 'escape'
braktis 'abandon'	t'oghnel 'abandon'		
simple endpoint			
arrij 'arrive' mbërrij 'arrive'	hasnel 'arrive'	ftano 'reach'	sroich 'reach'
simple potential endpoint			
afrohem 'approach' qasem 'approach'	motenal 'approach'	plisiazo 'approach'	
simple source-point			
drejtohem 'head for'		travo 'head for'	
largohem 'depart' u nisem 'set out' lë 'leave'	herranal 'go away'	(xe)feygo 'go away' afino 'leave' apomakryno 'remove oneself'	imigh 'leave' fág 'leave' *scoith 'leave'*
midpoint			
	ants'nel 'pass'	perno 'pass'	cuir do 'pass'
direction up			
ngrihem 'rise' ngjitem 'go up' hipi 'go up'	bardzranal 'rise'	anevaino 'go up' sikono 'rise'	éirigh 'rise'
direction down			
bie 'fall'	ynknel 'fall' t'ap'vel 'fall' teghal 'fall'	pefto 'fall'	tit 'fall'
zbres 'go down'	ijnel 'go down'	katevaino 'go down'	
	suzvel 'dive'		
direction forward			
përparoj 'advance'	arrajanal 'go forward'		

Albanian	**Armenian**	**Modern Greek**	**Irish**
direction around			
rrotullohem ‘go round’	shrjants’el ‘go round’	parakampto ‘go round’	
		strivo ‘turn’	*cuir ‘turn’*
direction back			
kthej ~ kthehem ‘return’	(vera)darrnal ‘return’	gyrizo ‘return, turn’ epistrefo ‘return’	fill ‘return’
spatial features of the path			
kaloj ‘cross’ përshkoj ‘cross’ kapërcej ‘cross’		diaschizo ‘cross’	trasnaigh ‘cross’ cuir de ‘cross’
	t’ap’ants’el ‘penetrate’	dieisdyo ‘penetrate’	

Appendix 10: List of consulted etymological dictionaries in the etymological analysis of motion verbs

Language	Sources used
general IE	Walde and Pokorny (1927-1932); Pokorny (1959-1969); Rix (2001)
Latin	Walde (1930-1956); Tucker (1931); de Vaan (2008)
Sanskrit	Uhlenbeck (1899); Macdonell (1929); Mayrhofer (1956-1980)
Gothic	Lehmann (1986)
Albanian	Meyer (1891); Demiraj (1997); Orel (1998)
Armenian	de Lagarde (1877); Hübschman (1883); Martirosyan (2010)
Dutch	Philippa et al. (2003-present); de Vries (1971); Franck (1976)
English	OED online (2012)
French	Brachet (1873); von Wartburg (1922-2003); Bloch and von Wartburg (1932)
German	Kluge (1883); Pfeifer (1993)
Hindi	Turner (1962-1985)
Irish	MacBain (1911); Vendryes (1959-present); Toner et al. (2007); Matasović (2009)
Italian	Pianigiani (1907); Battisti and Alessio (1950-1957); Olivieri (1965); Dizionari sapere.it (2012)
Latvian	Karulis (1992)
Lithuanian	Fraenkel (1962-1965)
Modern Greek	Andriōtēs (1951); Frisk (1954-1972); Mpampiniōtēs (2008); Beekes (2010)
Nepali	Turner (1931)
Persian	Horn (1893); Bartholomae (1904); Cheung (2007)
Polish	Sławski (1952-1982); Brückner (1957); Bańkowski (2000-present); Długosz-Kurczabowa (2003); Boryś (2005)
Portuguese	Machado (1952)
Romanian	Puşcariu (1905); Macrea (1958); Cioranescu (1966); Burnei (1996)
Russian	Vasmer (1953-1958); Derksen (2008)
Serbo-Croatian	Skok (1971-1974)
Swedish	Falk and Torp (1910-1911); Hellquist (1922); Östergren (1981)

Appendix 11: Manner verb etymologies

The verbs listed in Appendix 11 and Appendix 12 were classified according to the following scheme (see section 6.3):

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

Language	Tier	Manner verbs
Albanian	1	baret 'wander' (1a), eci 'walk' (3), fluturoj 'fly' (1c), kërcej 'jump' (1b), notoj 'swim' (1c), nxitoj 'hurry' (1a), sillet 'wander' (3), shëtis 'stroll' (1c), u sulem 'rush' (3), vrapoj 'run' (4b)
	2	çapitet 'stride' (1a), shkas 'slide' (3), shkel 'step' (3), u zvarrit 'drag onself' (6)
Armenian	1	glorvel 'roll' (3), khoyanal 'rush' (1b), k'aylel 'walk' (6), loghal 'swim' (3), shtapel 'hurry' (1c), slanal 'rush' (3), sural 'dash' (6), t'ap'arrel 'wander' (1a), t'rrch'el 'fly' (3), ts'atkel 'jump' (3), vazel 'run' (1b)
	2	sahel 'glide' (6), soghal 'crawl' (3), vargel 'trot' (6)
Dutch	1	baantjes trekken 'swim' (5), dwalen 'wander' (3), hollen 'run' (1a), kuieren 'stroll' (1a), lopen 'walk, run' (2), rennen 'run' (2), rollen 'roll' (1c), springen 'jump' (3), spurten 'scurry' (1c), vliegen 'fly' (3), wandelen 'stroll' (2), zwemmen 'swim' (2)
	2	draven 'trot' (3), dwarrelen 'twirl' (3), glijden 'glide' (2), huppelen 'hop' (2), klimmen 'climb' (2), kruipen 'crawl' (2), marcheren 'march' (1c), rijden 'ride' (3), sluipen 'sneak' (2), trekken 'hike' (1a), trippelen 'patter' (3)
English	1	fly (3), hurry (2), jump (1a), leap (2), roll (1c), run (2), rush (1a), scurry (1a), swim (2), walk (2), wander (2)
	2	climb (2), crawl (1c), creep (2), march (1c), ride (3), slip

English	2	(1c), slither (3), step (3), trot (1c), tumble (1c)
French	1	bondir ‘jump, rush’ (1a), cheminer ‘walk’ (1b), courir ‘run’ (2), errer ‘roam’ (2), filer ‘dash’ (2), foncer ‘dash’ (1b), marcher ‘walk’ (1a), nager ‘swim’ (2), se promener ‘stroll’ (4b), rouler ‘roll’ (1b), sauter ‘jump’ (3), voler ‘fly’ (3)
	2	fouler ‘tread’ (2), glisser ‘slide’ (1c), ramper ‘crawl’ (1c), trotter ‘trot’ (1c)
German	1	eilen ‘hurry’ (2), fliegen ‘fly’ (3), gehen ‘walk, go’ (2), laufen ‘walk, go, run’ (2), rennen ‘run’ (2), rollen ‘roll’ (1c), schlendern ‘stroll’ (3), schnellen ‘hurry’ (1a), schwimmen ‘swim’ (2), spazieren ‘stroll’ (1c), springen ‘jump’ (3), stürzen ‘dash’ (2), umherstreifen ‘wander’ (4b), wandern ‘wander’ (2), ziehen ‘wander’ (3)
	2	gleiten ‘glide’ (2), hoppeln ‘lollop’ (2), kriechen ‘crawl’ (2), marschieren ‘march’ (1c), purzeln ‘tumble’ (1a), reiten ‘ride’ (3), segeln ‘sail’ (2), treten ‘step’ (3), trotten ‘trot’ (1c)
Hindi	1	bhāgnā ‘run’ (2), bhaṭaknā ‘wander’ (2), chalāṃg lagnā ‘jump’ (5), dauṛnā ‘run’ (3), ghūmnā ‘roam’ (2), jhapaṭnā ‘dash’ (2), kūdnā ‘jump’ (3), lapaknā ‘dart’ (2), luṛhaknā ‘roll’ (2), tairnā ‘swim’ (2), uṛnā ‘fly’ (4a)
	2	cahlakdamī karnā ‘walk slowly’ (5), caṛhnā ‘climb’ (3), ḍubkī laganā ‘swoop’ (5), mārc karnā ‘march’ (5), rengnā ‘crawl’ (2), ṭahalnā ‘stroll’ (2)
Irish	1	bheith ag fánaíocht ‘wander’ (5), brostaigh ‘hurry’ (3), deifir ‘hurry’ (1b), eitil ‘fly’ (1b), léim ‘jump’ (3), rith ‘run’ (3), siúil ‘walk’ (3), snámh ‘swim’ (3)
	2	dreap ‘climb’ (3), marcaigh ‘ride’ (1b), máirseáil ‘march’ (1c), sciorr ‘slide’ (6), sleamhnaigh ‘slide’ (1b), sodar ‘trot’ (1a)
Italian	1	balzare ‘jump’ (1a), camminare ‘walk’ (1b), correre ‘run’ (2), filare ‘dash’ (2), nuotare ‘swim’ (3), passeggiare ‘stroll’ (1a), precipitarsi ‘rush’ (2), rotolare ‘roll’ (2), saltare ‘jump’ (3), trottare jump‘ (1c), vagare ‘wander’ (2), volare ‘fly’ (3)
	2	cavalcare ‘ride’ (2), scivolare ‘slip’ (1a), strisciare ‘crawl’ (1b)
Latvian	1	drāzties ‘dash’ (2), iet ‘walk, go’ (2), klīst ‘wander’ (2), laisties ‘fly’ (1a), lēkt ~ ļekāt ~ linkāt ‘jump’ (3), lidot ‘fly’ (1a), mesties ‘dash’ (2), pastaigāties ‘stroll’ (6), peldēt

Latvian	1	‘swim’ (3), ripot ‘roll’ (1b), skriet ‘run’ (3), steigties ‘hurry’ (3), šauties ‘dart’ (2), traukties ‘rush’ (3)
	2	cilpot ‘pace’ (1b), jāt ‘ride’ (3), kāpt ‘climb’ (2), kūleņot ‘somersault’ (1b), lidināties ‘flutter, hover’ (1a), līst ‘crawl’ (3), ložņāt ‘sneak’ (6), mīt ‘tread’ (3), rāpties ‘scramble, climb’ (3), soļot ‘march’ (1a), spraukties ‘wring (through)’ (2), tipināt ‘toddle’ (2)
Lithuanian	1	bėgti ‘run’ (3), eiti ‘walk, go’ (2), klaidžioti ‘wander’ (2), lėkti ‘dash’ (3), nerti ‘rush’ (2), plaukti ‘swim’ (3), riedėti ‘roll’ (3), skristi ‘fly’ (3), skubėti ‘hurry’ (3), šmurkštelėti ‘dart’ (1a), šokti ‘jump’ (1a), vaikščioti ‘walk’ (3)
	2	joti ‘ride’ (3), lipti ‘climb’ (3), mindyti ‘tread’ (3), rėplioti ‘crawl’ (3), risnoti ‘trot’ (1b), ropšti ‘scramble’ (3), sliuogti ‘climb’ (3), slysti ‘slide’ (3), tursenti ‘trot’ (1a), žengti ‘step’ (3)
Modern Greek	1	kylo ‘roll’ (1a), kolympo ‘swim’ (1b), ormo ‘rush’ (1b), perpato ‘walk’ (4b), peto ‘fly’ (3), pidao ‘jump’ (1b), planiemai ‘wander’ (1c), speydo ‘hurry’ (3), strifogyrizo ‘roll, whirl’ (1a), trecho ‘run’ (1a), trigyrizo ‘wander’ (1a), vadizo ‘walk, march’ (1b), viazomai ‘hurry’ (1b)
	2	glistro ‘slide’ (6), kalpazo ‘gallop’ (1b), pato ‘step’ (1a), sernomai ‘creep’ (1b)
Nepali	1	badbadaudai cha ‘run’ (5), bhautārinu ‘wander’ (2), dagurnu ‘run’ (3), daudinu ‘run’ (3), dulnu ‘wander’ (3), gudnu ‘roll’ (2), hatāra cha ‘hurry’ (5), hidnu ‘walk’ (2), ladnu padnu ‘roll’ (5), paudi khelnu ~ paudinu ‘swim’ (1c), udnu ‘fly’ (4a), uphranu ‘jump’ (4a)
	2	haknu ‘ride’ (2), lamkinu ‘stride’ (1b)
Persian	1	davidan ‘run’ (3), ġaltāndan ~ ġaltidan ‘roll’ (2), jahidan ‘leap’ (2), jastan ~ jast zadan ‘jump’ (2), kizan ‘jump’ (1b), parridan ‘fly’ (1b), parvāz kardan ~ nemudan ‘fly’ (5), rāh raftan ‘walk’ (5), shetaftan ‘hurry’ (3), šenā kardan ‘swim’ (5)
	2	ḵazidan ‘creep’ (2), laġzidan ‘slide’ (1c), qadam gozāštan ~ zadan ‘step’ (5)
Polish	1	biec ‘run’ (3), błąkać się ‘wander’ (3), frunąć ‘fly’ (1a), gnać ‘rush’ (3), iść ~ chodzić ‘walk, go’ (2), lecieć ‘fly’ (3), mknąć ‘zoom’ (3), pędzić ‘bolt’ (1a), pływać ‘swim’ (3), skoczyć ‘jump’ (3), smyrgnąć ‘dash’ (6), spacerować ‘stroll’ (1c), spieszyć ‘hurry’ (2), toczyć ‘roll’ (3), wędrować

Polish	1	‘wander’ (1c)
	2	cwałować ‘gallop’ (6), czołgać się ‘crawl’ (1a), jechać ‘ride’ (3), koziołkować się ‘tumble’ (1b), leźć ‘scramble’ (2), maszerować ‘march’ (1c), pełzać ‘crawl’ (2), ślizgać się ‘slip’ (2), sunąć ‘glide’ (3), tuptać ‘toddle’ (1a)
Portuguese	1	andar ‘walk’ (4a), caminhar ‘walk’ (1b), correr ‘run’ (2), nadar ‘swim’ (3), passear(-se) ‘stroll’ (2), precipitar-se ‘rush’ (2), pular ‘jump’ (1b), rolar ‘roll’ (1b), saltar ‘leap’ (3), vagar ‘wander’ (2), voar ‘fly’ (3)
	2	cavalgar ‘ride’ (2), escorregar ‘slip’ (4a), esgueirar-se ‘slip’ (4b), pisar ‘tread’ (2), rastejar ‘crawl’ (1b), trotar ‘trot’ (1c)
Romanian	1	alerga ‘run’ (4a), fugi ‘hurry’ (3), înota ‘swim’ (3), merge ‘walk’ (1a), plimba ‘stroll’ (4a), rătăci ‘wander’ (1a), sări ‘jump’ (3), umbla ‘wander’ (3), zbura ‘fly’ (4a), zori ‘rush’ (1b)
	2	călări ‘ride’ (1b), (a)luneca ‘slide’ (4b), păşi ‘step’ (1b), rostogoli ‘tumble’ (1b), topăi ‘hop’ (1b), tropăi ‘tramp’ (1b)
Russian	1	bežat' ‘run’ (3), bresti ~ brodit’ ‘wander, trudge’ (3), guljat’ ‘stroll’ (1a), idti ~ hodit’ ‘walk, go’ (2), katit’sja ‘roll’ (2), letet' ~ letat' ‘fly’ (3), mčat’ ~ mčat’sja ‘rush’ (3), plavat’ ~ pl'it' ‘swim’ (3), prygnut’ ~ prygivat’ ‘jump’ (2), rvanut’sja ‘dash’ (3)
	2	ehat' ‘ride’ (3), lezt’ ‘crawl’ (2), plestis’ ‘trudge’ (1a), rysit’ ‘trot’ (1b), skol’zit’ ~ skol’znut’ ‘glide’ (2), stupat’ ~ stupit’ ‘tread’ (3), šagat’ ‘march’ (1b), trusit’ ‘trot’ (1a)
Serbo-Croatian	1	hitati ‘hasten’ (2), hodati ‘walk, stroll’ (3), juriti ~ jurnuti ‘rush’ (2), kotrljati ‘roll’ (1a), letjeti ‘fly’ (3), lutati ‘wander’ (2), navaliti ‘rush’ (4b), obilaziti ‘wander’ (4b), plivati ‘swim’ (3), šetati ‘stroll’ (2), skakati ~ skočiti ~ skakutati ‘jump’ (3), strugnuti ‘dash’ (3), trčati ‘run’ (2), zuriti ‘hurry’ (2)
	2	gegati ‘waddle’ (1a), jahati ‘ride’ (3), kaskati ‘trot’ (1a), kliznuti ~ klizati ‘slip’ (1b), koračati ‘tread’ (6), puzati ‘crawl’ (2)
Swedish	1	flyga ‘fly’ (3), gå ‘walk, go’ (2), hoppa ‘jump’ (2), kila ‘scurry’ (2), promenera ‘walk’ (1c), rulla ‘roll’ (1c), rusa ‘rush’ (3), simma ‘swim’ (2), skynda ‘hurry’ (2), springa ‘run’ (3), vandra ‘wander’ (2), vanka ‘saunter’ (1c)

Swedish	2	glida ‘glide’ (1c), klättra ‘climb’ (1c), krypa ‘crawl’ (2), lunka ‘trot’ (2), marschera ‘march’ (1c), rida ‘ride’ (3), ringla ‘slither’ (1c), smyga ‘sneak, creep’ (3), trava ‘trot’ (3)

Appendix 12: Path verb etymologies

Language	tier	path verbs
Albanian	1	arrij ‘arrive’ (1c), bie ‘fall’ (3), braktis ‘abandon’ (1c), kaloj ‘cross’ (1c), kapërcej ‘cross’ (4b), kthej ~ kthehem ‘return’ (4b), largohem ‘depart’ (1b), lë ‘leave’ (3), mbärrij ‘arrive’ (4b), ngjitem ‘go up’ (4b), ngrihem ‘rise’ (4b), u nisem ‘set out’ (1a), përshkoj ‘cross’ (4b)
	2	afrohem ‘approach’ (1b), dal ‘go out’ (3), drejtohem ‘head for’ (1b), futem ‘go in’ (1a), hipi ‘go up’ (3), hyj ‘go in’ (1a), iki ‘escape’ (3), ndjek ‘follow’ (4b), përparoj ‘advance’ (4b), qasem ‘approach’ (1a), rrotullohem ‘go round’ (1b), zbres ‘go down’ (4b)
Armenian	1	bardzranal ‘rise’ (1b), darrnal ‘return’ (3), hasnel ‘arrive’ (3), herranal ‘go away’ (1b), teghal ‘fall’ (6), t’ap’vel ‘fall’ (6), t’oghnel ‘abandon’ (3), ynknel ‘fall’ (3)
	2	ants’nel ‘pass’ (3), arrajanal ‘go forward’ (1b), hetapndel (1b), hetevel ‘follow’ (1b), ijnel ‘go down’ (3), motenal ‘approach’ (3), mtnel ‘go in’ (1b), shrjants’el ‘go round’ (1b), suzvel ‘dive’ (1b), t’ap’ants’el ‘penetrate’ (1a)
Dutch	1	arriveren ‘arrive’ (1c), zich begeven ‘set off’ (2), oversteken ‘cross’ (4b), stijgen ‘rise’ (2), terugkeren ‘return’ (4b), vallen ‘fall’ (3), verlaten ‘leave’ (4b)
	2	doordringen ‘penetrate’ (4b), omslaan ‘turn’ (4b), ontsnappen ‘escape’ (4b), vluchten ‘flee’ (1b), volgen ‘follow’ (2)
English	1	arrive (1c), cross (1b), fall (3), get (1a), leave (2), reach (1a), return (1c), rise (2)
	2	descend (1c), flee (2), follow (2), pass (1c), pursue (1c), turn (1c)
French	1	arriver ‘arrive’ (2), s’éloigner ‘depart’ (1b), monter ‘go up’ (2), partir ‘leave’ (2), quitter ‘leave’ (1b), rejoindre ‘return’ (4b), rentrer ‘return’ (4b), repartir ‘return’ (4b), retourner ‘return’ (4b), revenir ‘return’ (4a), sortir ‘leave’ (1a), (re)tomber ‘fall’ (1a), traverser ‘cross’ (2)
	2	s’approcher ‘approach’ (2), atterrir ‘land’ (1b), (s’)avancer ‘go forward’ (2), contourner ‘go around’ (4b), descendre ‘come down’ (4a), se diriger ‘head for’ (1c), s'échapper ‘escape’ (2), entrer ‘enter’ (2), fuir ‘flee’ (3), passer ‘pass’ (2), pénétrer ‘penetrate’ (1c), suivre ‘follow’ (3)

German	1	sich begeben 'set off' (2), durchqueren 'cross' (4b), eintreffen 'arrive' (4b), sich entfernen 'depart' (1b), sich erheben 'rise' (4b), fallen 'fall' (3), gelangen 'reach' (4b), überqueren 'cross' (4b), verlassen 'leave' (4b), zurückkehren 'return' (4b)
	2	eindringen 'penetrate' (4b), eintauchen 'dive' (2), folgen 'follow' (2), sich nähern 'approach' (1b)
Hindi	1	choṛnā 'leave' (2), girnā 'fall' (2), pahuṃcnā 'arrive' (4a), pār karnā 'cross' (5), tyāgnā 'abandon' (2)
	2	baṛhnā 'advance' (2), calnā 'go forward, advance' (3), ghūmnā 'turn' (2), guzarnā 'pass' (6), nikalnā 'go out' (4a), samānā 'go in' (4a), utarnā 'go down' (4a)
Irish	1	cuir de 'cross' (5), éirigh 'rise' (4b), fág 'leave' (4b), fill 'return' (3), imigh 'leave' (4b), sroich 'reach' (4b), tit 'fall' (4b), trasnaigh 'cross' (1b)
	2	cuir do 'pass' (5), éalaigh 'escape' (3), lean 'follow' (1a)
Italian	1	allontanarsi 'go away' (1b), andarsene 'go away' (1a), arrivare 'arrive' (2), cadere 'fall' (3), giungere 'arrive' (1a), lasciare 'leave' (2), montare 'go up' (1c), partire 'leave' (2), salire 'go up' (2), sollevarsi 'rise' (4a), tornare(se) 'return' (2), traversare 'cross' (2)
	2	accodarsi 'follow' (1b), atterrare 'land' (1b), avanzare 'move forward' (2), avvicinarsi 'approach' (1b), entrare 'enter' (2), fuggire 'flee' (3), girare 'turn' (2), passare 'pass' (2), penetrare 'penetrate' (2), scappare 'escape' (2), scendere 'descend' (4a), seguire 'follow' (3), uscire 'go out' (4a)
Latvian	1	atstāt 'leave' (4b), attālināties 'go away' (6), birt 'fall' (3), doties 'set out' (1a), ierasties 'arrive' (4b), krist 'fall' (2), nonākt 'reach' (4b), pamest 'abandon' (4b), šķērsot 'cross' (1b), tikt 'get' (2)
	2	bēgt 'flee' (3), griezties 'turn' (3), iekļūt 'get in' (4b), izkļūt 'get out' (4b), nirt 'dive' (2), sekot 'follow' (3)
Lithuanian	1	atsidurti 'get' (4b), byrėti 'fall' (3), grįžti 'return' (3), kilti 'rise' (3), kristi 'fall' (2), palikti 'abandon' (3), patekti 'get' (3), tolti 'go away' (1b), virsti 'fall' (3)
	2	lenkti 'go around' (2), sekti 'follow' (3), sileisti 'let oneself go' (3), sukti 'turn' (3), traukti 'go to' (2), vingiuoti 'turn' (3)
Modern Greek	1	afino 'leave' (4b), anevaino 'go up' (4b), apomakryno 'remove oneself' (4b), diaschizo 'cross' (4b), egkataleipo

Modern Greek	1	'abandon' (4b), epistrefo 'return' (4b), feygo 'go away' (3), ftano 'reach' (6), gyrizo 'return, turn' (1b), pefto 'fall' (3), sikono 'rise' (1a)
	2	akoloytho 'follow' (1a), dieisdyo 'penetrate' (4b), katevaino 'go down' (4b), kynigo 'chase' (1b), mpaino 'go in' (4b), parakampto 'go round' (4b), perno 'pass' (1b), plisiazo 'approach' (1b), strivo 'turn' (1a), travo 'head for' (1b), vgaino 'go out' (4b)
Nepali	1	barsinu 'fall' (2), chodnu 'abandon' (1c), jharnu 'fall' (2), khasnu 'fall' (2), par garnu 'cross' (5), pharkinu 'return' (2), pugnu 'reach' (2), tarnu 'cross' (3), uklinu 'go up' (4a)
	2	aghi badhnu 'advance' (5), dubnu 'go down' (2), niskanu 'go out' (4a), orlinu 'go down' (4a), pasnu 'go in' (4a)
Persian	1	barḵāstan 'rise' (4b), bāzāmadan 'return' (4b), bāzgaštan 'return' (4b), dur sakhtan 'go away' (5), oftādan 'fall' (3), (be) rāh oftādan 'set out' (5), (az rāh) residan 'arrive' (3), tark kardan 'abandon' (5), vāgozāštan 'leave' (4b)
	2	dāḵel šodan 'go in' (5), darāmadan 'come in' (4b), dowr zadan 'go around' (5), gozaštan 'pass' (4a), ḵārej šodan 'go out' (5), nazdik šodan 'approach' (5), nofuz kardan 'penetrate' (5), vāred šodan 'go in' (5)
Polish	1	dostać 'get' (4b), dotrzeć 'get' (4b), oddalić 'go away' (4b), opuścić 'leave' (4b), padać 'fall' (3), porzucić 'abandon' (4b), przeprawiać 'cross' (4b), przybyć 'arrive' (4b), ruszyć 'set out' (2), wrócić 'return' (3), wybrać się 'set out' (4b)
	2	ganiać się 'chase' (3), krążać 'go around' (1b), kręcić 'turn' (2), następować 'follow' (4b), nurkować 'dive' (2), skierować się 'go to' (6), zbaczać 'go around' (4b), zbliżać się 'approach' (4b)
Portuguese	1	abandonar 'leave' (1c), afastar-se 'distance oneself' (6), atravessar 'cross' (2), cair 'fall' (3), chegar 'arrive' (2), cruzar 'cross' (1b), deixar 'abandon' (2), erguer-se 'rise' (4a), partir 'leave' (2), regressar 'return' (1b), retornar 'return' (4b), voltar 'return' (2)
	2	aproximar-se 'approach' (2), aterrar 'land' (1b), descer 'descend' (4a), distanciar-se 'distance oneself' (1b), contornar 'go around' (1c), dirigir-se 'go to' (4a), enfiar-se 'enter' (1b), entrar 'enter' (2), escapar 'escape' (2), escapulir-se 'escape' (1c), fugir 'flee' (3), mergulhar 'dive' (1b), passar 'pass' (2), penetrar 'penetrate' (2), sair 'go

Portuguese	2	out' (2), seguir 'follow' (3), subir 'go up' (4a), virar 'turn' (3)
Romanian	1	ajunge 'arrive' (4a), cădea 'fall' (3), depărta 'leave' (1b), se duce 'leave' (3), se întoarce 'return' (4a), părăsi 'abandon' (1c), pleca 'leave' (2), porni 'depart' (1c), se ridica 'rise' (4a), străbate 'cross' (4b), sui 'go up' (4a), traversa 'cross' (1c), urca 'go up' (3)
	2	se alătura 'approach' (1b), se apropia 'come near' (2), coborî 'descend' (1c), coti 'turn' (1b), cufunda 'dive' (4a), ieşi 'go out' (4a), se îndrepta 'go toward' (1a), intra 'enter' (2), (se) lua 'take oneself' (2), ocoli 'go around' (1b), pătrunde 'penetrate' (4a), scăpa 'escape' (2), trece 'pass' (4a), urma 'follow' (1b)
Russian	1	dvinut'sja 'set out' (2), padat' ~ past' 'fall' (3), pereseč' ~ peresekat' 'cross' (4b), podnjat'sja ~ podnimat'sja 'go up' (4b), valit'sja 'fall' (3), vernut'sja 'return' (3), vozvrašat'sja 'return' (4b)
	2	blizit'sja 'approach' (1b), brat'sja 'take oneself' (3), kružit' 'go round' (1b), minovat' 'pass' (3), ogibat' 'go round' (4b), pravit'sja 'direct oneself to' (1b), spuskat'sja ~ spustit'sja 'go down' (4b)
Serbo-Croatian	1	dignuti 'rise' (2), izaći 'leave' (4b), krenuti 'set out' (2), napustiti 'abandon' (4b), ostaviti 'leave' (4b), otići ~ odlaziti 'leave' (4b), padati ~ pasti 'fall' (3), se penjati 'rise' (1a), poći 'leave' (4b), se (po)peti 'go up' (1a), prelaziti ~ prijeći 'cross' (4b), se rušiti 'fall' (3), stići 'arrive' (4b), se udaljiti 'go away' (4b), se ukloniti 'go away' (4b), se uputiti 'set out' (4b), vratiti ~ vraćati 'return' (3)
	2	se približiti 'approach' (4b), pridolaziti 'approach' (4b), prodrijeti 'penetrate' (4b), prolaziti ~ proći 'pass' (4b), sići ~ silaziti 'go down' (4b), unići ~ ući 'enter' (4b), se uvaliti 'enter' (4b), zalaziti 'go behind' (4b), zaobići 'go around' (4b), zaroniti 'dive' (4b)
Swedish	1	återvända 'return' (4b), avlägsna sig 'remove oneself' (1a), falla 'fall' (3), lämna 'leave' (2)
	2	fly 'flee' (2), följa 'follow' (2), hamna 'land' (1b), ta sig 'take oneself' (2), vika 'turn' (3)

Appendix 13: The etymological origins of manner verbs, averaged for all languages and the four big subgroups

etymology type	mean overall	Germanic	Romance	Balto-Slavic	Indo-Iranian
1a. unspecified / semantic shift	0.11	0.09	0.11	0.15	0
1b. derived non-verbal element	0.12	0	0.23	0.07	0.08
1c. borrowed	0.10	0.23	0.08	0.02	0.05
2. subgroup	0.25	0.40	0.27	0.25	0.37
3. IE	0.29	0.26	0.19	0.43	0.18
4a. inherited preverb+verb	0.03	0	0.08	0	0.07
4b. modern preverb+verb	0.02	0.01	0.05	0.02	0
5. complex verb	0.04	0.01	0	0	0.25
6. unknown	0.03	0	0	0.04	0

Appendix 14: The etymological origins of path verbs, averaged for all languages and the four big subgroups

etymology type	mean overall	Germanic	Romance	Balto-Slavic	Indo-Iranian
1a. unspecified / semantic shift	0.05	0.06	0.05	0.03	0
1b. derived non-verbal element	0.12	0.10	0.17	0.08	0
1c. borrowed	0.05	0.13	0.09	0	0.02
2. subgroup	0.21	0.30	0.35	0.14	0.28
3. IE	0.18	0.11	0.11	0.31	0.09
4a. inherited preverb+verb	0.07	0	0.16	0	0.23
4b. modern preverb+verb	0.26	0.30	0.07	0.42	0.10
5. complex verb	0.05	0	0	0	0.25
6. unknown	0.02	0	0.01	0.02	0.02

Appendix 15: The etymological origins of manner verbs and path verbs averaged for all languages and separated by tier

etymology type	1st tier manner	2nd tier manner	total manner	1st tier path	2nd tier path	total path
1a. unspecified / semantic shift	0.11	0.11	0.11	0.06	0.05	0.06
1b. derived non-verbal element	0.09	0.14	0.11	0.08	0.19	0.13
1c. borrowed	0.07	0.14	0.10	0.06	0.05	0.06
2. subgroup	0.29	0.23	0.27	0.19	0.22	0.20
3. IE	0.32	0.26	0.30	0.18	0.15	0.17
4a. inherited preverb+verb	0.03	0.01	0.02	0.05	0.10	0.07
4b. modern preverb+verb	0.03	0.02	0.02	0.33	0.18	0.26
5. complex verb	0.04	0.03	0.04	0.03	0.05	0.04
6. unknown	0.02	0.06	0.03	0.03	0.01	0.02

Appendix 16: Paired sample t-tests of the number of manner verbs and path verbs in each etymology category

Etymology type	Variables	Mean	s.d.	t	p
1a. unspecified / semantic shift	manner verbs	2.10	1.55	3.32	0.004**
	path verbs	1.00	1.17		
1b. derived non-verbal element	manner verbs	2.00	2.00	-0.75	0.46
	path verbs	2.35	2.30		
1c. borrowed	manner verbs	1.80	2.04	1.53	0.14
	path verbs	1.05	1.70		
2. subgroup	manner verbs	4.90	3.49	1.33	0.20
	path verbs	3.70	3.56		
3. IE	manner verbs	5.50	3.24	4.37	0.0003***
	path verbs	3.10	2.29		
4a. inherited preverb+verb	manner verbs	0.40	0.88	-2.78	0.01**
	path verbs	1.30	2.20		
4b. modern preverb+verb	manner verbs	0.40	0.60	-4.02	0.0007***
	path verbs	4.75	5.19		
5. complex verb	manner verbs	0.65	1.35	-0.16	0.87
	path verbs	0.70	2.05		
6. unknown	manner verbs	0.60	1.05	1.56	0.14
	path verbs	0.35	0.59		

* = $p \leq 0.05$; ** = $p \leq 0.01$; *** = $p \leq 0.001$.

Samenvatting

Mensen die verschillende talen spreken kunnen op verschillende wijzen uitdrukking geven aan beweging. Vergelijk bijvoorbeeld de twee krantenkoppen in (1) en (2):

1) Nederlands
Waaghals Nik Wallenda loopt op staalkabel over Grand Canyon[18]

2) Frans
Le funambule Nik Wallenda traverse le Grand Canyon[19]

De Nederlandse zin in (1) geeft aan dat Nik Wallenda over een staalkabel *naar de andere kant* van de Grand Canyon is *gelopen*, terwijl de Franse kop in (2) enkel aangeeft dat Nik Wallenda de Grand Canyon is *overgestoken*. Dezelfde gebeurtenis, namelijk de oversteek van Nik Wallenda op een stalen koord, wordt dus door de twee krantenkoppen op verschillende wijzen uitgedrukt. De Nederlandse kop gebruikt een werkwoord dat de manier van beweging uitdrukt, *lopen*, en een voorzetsel, *over*. De Franse kop maakt gebruik van een werkwoord dat het pad van beweging uitdrukt, *traverser* 'oversteken', maar geeft niet expliciet aan dat Nik Wallenda over een koord loopt. Dit verschil lijkt op het eerste gezicht vreemd: zelf bewegen en bewegingen van anderen waarnemen zijn centrale pijlers van de menselijke cognitie. Dus waarom zouden mensen in verschillende talen zo verschillend over beweging praten?

Uit eerder onderzoek blijkt dat het verschil tussen de Nederlandse en de Franse krantenkop niet toevallig is, maar de voorkeur van deze twee talen wat betreft de uitdrukking van beweging laat zien. Talen zoals het Nederlands doen dit het liefste met een werkwoord dat de manier van beweging uitdrukt (zoals *lopen* in (1) en een voorzetsel of bijwoord dat het pad van beweging uitdrukt (zoals *over* in (1)). Deze talen worden ook wel '**satellite-framed**' genoemd. Talen zoals het Frans geven de voorkeur aan een werkwoord dat het pad van beweging uitdrukt (zoals *traverser* 'oversteken' in (2)) en laten de manier van beweging weg, of zetten dat in een bijwoord of bijzin. Talen zoals het Frans worden ook wel '**verb-framed**' genoemd.

---

[18] artikel gepubliceerd op 24-06-2013: http://www.nieuws.nl/opmerkelijk/20130624/Waaghals-Nik-Wallenda-loopt-op-staalkabel-over-Grand-Canyon
[19] artikel gepubliceerd op 24-06-2013: http://www.lemonde.fr/ameriques/article/2013/06/24/le-funambule-nik-wallenda-traverse-le-grand-canyon_3435154_3222.html

Dit proefschrift is een historische oftewel diachronische studie naar de uitdrukking van beweging in taal in de Indo-Europese taalfamilie met behulp van comparatieve fylogenetische methoden zoals die in de evolutionaire biologie gebruikt worden. Hoofdstuk 1 is een algemene inleiding waarin allereerst het doel en de relevantie van het onderzoek uiteengezet worden. Hoewel er onderzoek verricht is naar de uitdrukking van beweging in veel verschillende talen, is er nog relatief weinig onderzoek gedaan naar de diachronische factoren en processen die invloed hebben op de uitdrukking van beweging in taal. Om deze processen beter te begrijpen, is er voor gekozen om onderzoek te doen naar de uitdrukking van beweging in taal in één taalfamilie: de Indo-Europese taalfamilie.

De keuze voor een enkele taalfamilie in plaats van bijvoorbeeld een steekproef van talen uit verschillende taalfamilies is gemaakt omdat ik op deze manier kan kijken hoe talen die nauw verwant zijn, verschillen en op elkaar lijken. De keuze voor de Indo-Europese taalfamilie is gemaakt omdat we relatief veel weten over deze talen. Dat betekent dat de resultaten van de analyses in dit proefschrift vergeleken kunnen worden met de bestaande literatuur over de uitdrukking van beweging in deze talen. Daarnaast zijn de Indo-Europese talen erg divers wat betreft de uitdrukking van beweging, en is het interessant om te ontdekken hoe deze verschillen tot stand zijn gekomen vanuit een diachronisch perspectief. De rest van Hoofdstuk 1 is gewijd aan een inleiding op comparatieve fylogenetische methoden. Dit zijn statistische methoden die gebruikt kunnen worden om comparatieve data, zoals informatie over de uitdrukking van beweging in verschillende talen, te modelleren op een fylogenetische boom. Een fylogenetische boom is een representatie van de geschiedenis van een groep verwante talen.

Hoofdstuk 2 introduceert de dataset die in de rest van het proefschrift geanalyseerd wordt. Om vergelijkbare informatie over alle talen te hebben, is er gebruik gemaakt van een parallel corpus. Een parallel corpus bestaat uit een originele tekst en de vertalingen van deze tekst in verschillende talen. De teksten die zijn gekozen voor deze studie zijn drie boeken: *Alice's Adventures in Wonderland*, *Through the Looking-Glass and what Alice found there* (beide door Lewis Carroll) en *O Alquimista* ['De Alchemist'] (door Paulo Coelho). De twee *Alice* boeken zijn oorspronkelijk in het Engels (een **satellite-framed** taal) geschreven, terwijl *O Alquimista* in het Portugees (een **verb-framed** taal) geschreven is. Voor het parallelle corpus zijn de vertalingen van deze boeken in twintig Indo-Europese talen gebruikt.

Het corpus bestaat uit een selectie van 215 zinnen die beweging uitdrukken en de vertalingen van deze zinnen in deze twintig talen. De volgende talen zijn opgenomen in het corpus: Engels, Nederlands, Duits, Zweeds (Germaans), Frans, Italiaans, Portugees, Roemeens (Romaans), Russisch, Pools, Servo-Kroatisch, Litouws, Lets (Balto-Slavisch), Perzisch, Hindi, Nepali (Indo-

Iraans), Iers (Keltisch), Modern Grieks, Albaans, en Armeens. Na de beschrijving van het corpus geeft Hoofdstuk 2 een overzicht van een aantal centrale categorieën wat betreft de uitdrukking van beweging in taal, zoals verschillende soorten werkwoorden, en de verschillende constructies die in het corpus gebruikt worden om beweging uit te drukken, zoals de satellite-framed constructie (zie (1)) en verb-framed constructie (zie (2)).

Hoofdstuk 3 geeft een overzicht van de gebruiksfrequentie van de verschillende constructies in de twintig Indo-Europese talen. In dit hoofdstuk concentreer ik me voornamelijk op de vraag of de typologie die traditioneel is gebruikt, namelijk een dichotomie tussen **satellite-framed** en **verb-framed** talen, toereikend is om deze twintig talen te kunnen classificeren. Ik geef eerst een overzicht van hoe vaak de verschillende constructies gebruikt worden in deze talen. Hierbij wordt duidelijk dat een aantal talen inderdaad op traditionele wijze geclassificeerd kan worden, te weten: Nederlands, Engels, Duits, Zweeds, Russisch, Pools, Litouws, en Lets maken vaak gebruik van de **satellite-framed** constructie, zij kunnen '**satellite-framed**' genoemd worden; terwijl Frans, Italiaans, Portugees, Roemeens, Albaans, en Grieks vaak gebruik maken van de **verb-framed** constructie, zij kunnen '**verb-framed**' genoemd worden.

Voor de andere talen geldt echter dat ze niet in deze dichotomie passen: Servo-Kroatisch, Nepali, Hindi, Perzisch, Armeens en Iers gebruiken regelmatig andere, aanvullende constructies, maar vormen geen homogene groep omdat iedere taal andere constructies prefereert. Dit resultaat wordt ondersteunt door een **Neighbor-Net** analyse en een **multidimensional scaling** (**MDS**) analyse. De traditionele dichotomie is dus niet toereikend om het gebruik van de verschillende constructies in alle onderzochte talen te analyseren. Ik stel dan ook voor om extreem '**satellite-framed**' en extreem '**verb-framed**' als uiteinden van een continue schaal te zien, waarop talen ook een tussenliggende positie kunnen innemen. Daarnaast bestaan er ook andere schalen: talen kunnen bijvoorbeeld meer of minder gebruik maken van constructies die deixis uitdrukken.

In Hoofdstuk 4 presenteer ik de resultaten van een studie naar de veranderingen die hebben plaatsgevonden in het gebruik van de verschillende constructies. Omdat in Hoofdstuk 3 duidelijk is geworden dat een simpele classificatie van talen in de twee categorieën '**satellite-framed**' en '**verb-framed**' niet toereikend is, worden de gebruiksfrequenties van individuele constructies en de resultaten van een principale-componentenanalyse gebruikt. Op basis van de literatuur onderscheidt ik twee hypotheses wat betreft de uitdrukking van beweging in het Proto-Indo-Europees: 1) Proto-Indo-Europees was **satellite-framed** en 2) Proto-Indo-Europees was een mix tussen **satellite-framed** en **verb-framed**.

Om een zo goed mogelijk overzicht te geven van de diachronische veranderingen in de uitdrukking van beweging in de Indo-Europese taalfamilie, wordt ook een overzicht gegeven van de literatuur wat betreft veranderingen in het **preverb** systeem. Proto-Indo-Europees heeft een gereconstrueerd systeem waarin zogenaamde **preverbs**, adverbiale partikels die vrij door de zin heen kunnen bewegen, het pad van beweging uit kunnen drukken. Veranderingen in dit **preverb** systeem hebben een grote rol gespeeld in hoe de moderne Indo-Europese talen beweging uitdrukken. In de Romaanse en Indo-Iraanse talen zijn deze **preverbs** eerst prefixen op werkwoorden geworden, en later compleet met deze werkwoorden samengesmolten. De Balto-Slavische talen hebben nog altijd een systeem van pad prefixen. De Germaanse talen hebben een soortgelijk systeem waarbij het pad van beweging op separabele prefixen wordt uitgedrukt (zoals in het Nederlands: *oversteken*, *terugkeren*, *doordringen*).

Het resultaat van de fylogenetische analyse toont aan dat Proto-Indo-Europees waarschijnlijk een mix van **satellite-framed** en **verb-framed** was, met een lichte neiging naar het **satellite-framed** uiteinde van de schaal. Dit resultaat komt overeen met de literatuur over de **preverb** systemen in de oudste Indo-Europese talen. Het is waarschijnlijk dat het Proto-Indo-Europees een productief **preverb** systeem had, dat gebruikt kon worden in de **satellite-framed** constructie, maar ook andere constructies gebruikte voor de uitdrukking van beweging. De talen die een productief **preverb** systeem hebben kunnen behouden (Germaans en Balto-Slavisch), zijn meer **satellite-framed** geworden. Talen die hun productieve **preverb** system zijn verloren (Romaans en Indo-Iraans), zijn meer **verb-framed** geworden.

Hoofdstuk 5 gaat dieper in op de relatie tussen de constructies die uitdrukking geven aan beweging en de werkwoorden die in deze constructies gebruikt worden, namelijk werkwoorden die de manier van bewegen uitdrukken (**manner verbs**) en werkwoorden die het pad van beweging uitdrukken (**path verbs**). De specifieke hypotheses waarop de nadruk wordt gelegd zijn 1) talen die meer gebruik maken van de **satellite-framed** constructie hebben meer **manner verbs** en 2) talen die meer gebruik maken van de **verb-framed** constructie hebben meer **path verbs**.

De resultaten van de fylogenetische analyse (**Phylogenetic Generalized Least Squares**) tonen aan dat er een positieve correlatie is tussen het gebruik van de **satellite-framed** constructie en de grootte van de **manner verb** en **path verb** lexicons: talen die zich dichter bij het **satellite-framed** uiteinde van de schaal bevinden hebben een groter **manner verb** lexicon en een kleiner **path verb** lexicon. Daarnaast is er een positieve correlatie tussen het gebruik van de **verb-framed** constructie en de grootte van de **manner verb** en **path verb** lexicons: talen die zich dichter bij het **verb-framed** uiteinde van de schaal bevinden hebben een kleiner **manner verb** lexicon en een groter **path verb** lexicon. De volgende stap om deze correlaties verder te onderzoeken is om te kijken in welke richting er veranderingen optreden: worden talen bijvoorbeeld eerst **satellite-framed** en krijgen ze dan een groter **manner verb** lexicon, zoals in de literatuur is voorgesteld, of verwerven talen eerst meer **manner verbs** en worden ze als een reactie daarop meer **satellite-framed**? Hoewel er op deze vraag in de huidige studie nog geen antwoord gegeven kan worden, maakt zij wel voldoende mogelijkheden voor vervolgonderzoek zichtbaar.

In Hoofdstuk 6 wordt nader ingegaan op de relatie tussen de constructies die beweging uitdrukken en de werkwoorden die in deze constructies voorkomen. Aangezien in Hoofdstuk 5 is gedemonstreerd dat er een verband is tussen het type constructie dat gebruikt wordt (**satellite-framed** en **verb-framed**) en het lexicon van werkwoorden van beweging (**manner verbs** en **path verbs**), rijst hier de vraag op hoe dit verband historisch gezien tot stand komt. In Hoofdstuk 6 worden de resultaten van twee analyses weergegeven: een onderzoek naar de etymologie van de 367 **manner verbs** en 366 **path verbs** die gevonden zijn in het corpus, en een fylogenetisch onderzoek naar het tempo van verandering van de grootte van de **manner verb** en **path verb** lexicons. Beide analyses tonen aan dat de evolutie van **manner verbs** en **path verbs** op verschillende wijzen verloopt.

De etymologische studie toont aan dat er een aantal opvallende verschillen zijn tussen de etymologieën van **manner verbs** en **path verbs**. Allereerst lijken de etymologieën van **manner verbs** in de verschillende talen meer op elkaar dan die van **path verbs**. Daarnaast zijn **manner verbs** in het algemeen ouder; zij hebben vaker etymologieën die tot op een Indo-Europees niveau teruggaan. **Path verbs** hebben in alle Indo-Europese talen in ieder geval een aantal etymologieën die gelexicaliseerde **preverb**-werkwoord combinaties zijn. In het merendeel van de talen zijn dit moderne gelexicaliseerde **preverb**-werkwoord combinaties, terwijl de Romaanse en Indo-Iraanse talen ook geërfde, en dus oudere, gelexicaliseerde **preverb**-werkwoord combinaties hebben.

Het tempo van verandering werd onderzocht middels een Bayesiaanse fylogenetische analyse. Deze analyse toont aan dat in sommige Indo-Europese subgroepen de grootte van het **manner verb** lexicon sneller evolueert, terwijl in andere subgroepen de grootte van het **path verb** lexicon sneller evolueert. De resultaten van de fylogenetische studie en de etymologische studie tonen aan dat **manner verbs** en **path verbs** verschillende etymologische oorsprongen hebben en dat de grootte van deze twee klassen op verschillende wijze veranderd is in de Indo-Europese taalfamilie.

Hoofdstuk 7 besluit het proefschrift. Hierin geef ik een overzicht van de belangrijkste bevindingen, breng ik de verschillende thema's van het proefschrift nader bij elkaar, en waar de mogelijkheden voor vervolgonderzoek liggen. Allereerst wordt een kort overzicht van de conclusies gegeven. Hoofdstuk 3 toonde aan dat niet alle talen in het corpus ingedeeld kunnen worden in de traditionele dichotomie van **satellite-framed** en **verb-framed** talen, en dat deze beter als uiteinden van een continue schaal gezien kunnen worden waarop talen een tussenliggende positie kunnen innemen. Hoofdstuk 4 besprak de fylogenetische analyse van het Proto-Indo-Europees, die uitwees dat het Proto-

Indo-Europees midden op deze schaal geplaats kan worden. Hoofdstuk 5 presenteerde bewijs dat aantoonde dat er een relatie is tussen het gebruik van bepaalde constructies en de grootte van de **manner verb** en **path verb** lexicons. Hoofdstuk 6 onderbouwde deze relatie verder door aan te tonen dat in sommige Indo-Europese subgroepen het **manner verb** lexicon sneller evolueert, terwijl in andere subgroepen het **path verb** lexicon sneller evolueert.

In Hoofdstuk 4–6 is daarmee duidelijk geworden dat verandering in de uitdrukking van beweging in de Indo-Europese talen voornamelijk door twee processen beïnvloed is: de morfo-syntactische veranderingen met betrekking tot het **preverb** systeem en de rol die contact met zowel Indo-Europese als niet-Indo-Europese talen heeft gespeeld in de totstandkoming en het handhaven van **satellite-framed** en **verb-framed** zones. In de literatuur was men al eerder tot de conclusie gekomen dat Noord en Centraal Europa vooral **satellite-framed** is, terwijl Zuid Europa **verb-framed** is. De Germaanse en Balto-Slavische talen hebben hun systeem van (separabele of niet-separabele) pad prefixen kunnen behouden omdat zij in nauw contact stonden met elkaar en met soortgelijke talen (zoals bijvoorbeeld de Fins-Oegrische talen). In Zuid Europa vinden we verschillende **verb-framed** talen die hun **preverb** systeem verloren hebben en die met elkaar in contact zijn geweest: de Romaanse talen, Baskisch, Albaans, Grieks, en Turks. In Hoofdstuk vier werd ook een **partial Mantel test** uitgevoerd die test of talen die dichterbij elkaar gesproken worden beweging op soortgelijke wijze uitdrukken. De resultaten hiervan zijn niet statistisch significant, maar desondanks is contact de meest voor de hand liggende verklaring voor de veranderingen die hebben plaatsgevonden in de verschillende Indo-Europese taalgroepen.

In de rest van Hoofdstuk 7 wordt ingegaan op de verschillende lijnen die in dit proefschrift zijn uitgezet om verder onderzoek naar de uitdrukking van beweging in de talen van de wereld te doen. Het meest voor de hand liggend is vervolgonderzoek naar de factoren die het gebruik van de verschillende constructies bepalen en naar de processen die het lexicon beïnvloeden. Een van de factoren die invloed heeft op welke constructie gekozen wordt is de semantiek van de beschreven gebeurtenis: gebeurtenissen die 'natuurlijker' zijn (bijvoorbeeld: *Suzanne rende de straat over*) hebben wellicht andere voorkeuren dan gebeurtenissen die minder 'natuurlijk' zijn (bijvoorbeeld: *Suzanne danste de straat over*). Onderzoek naar deze en andere factoren kan meer licht werpen op de taal-interne diversiteit die talen hebben om beweging uit te drukken.

Daarnaast is het verassend om te zien dat er binnen de historische taalkunde ontzettend veel gebruik gemaakt wordt van woordenlijsten om fylogenetische relaties tussen talen te bestuderen, maar dat er eigenlijk maar zeer weinig vergelijkend onderzoek is gedaan naar de processen waar het lexicon aan onderhevig is. Deze studie toont dan ook aan dat er op dit front vele mogelijkheden zijn om meer inzicht te verwerven over waarom talen bepaalde woorden en semantische klassen hebben. Ook het onderzoek naar de interne structuur van semantische domeinen zoals **manner verbs** of **path verbs** staat nog in de kinderschoenen.

Dit proefschrift toont daarnaast aan dat het toepassen van fylogenetische analyses op taalkundige vraagstukken kan leiden tot nieuwe inzichten, zowel om diachronische veranderingen in de uitdrukking van beweging te analyseren als om andere diachronische vraagstukken te onderzoeken. Hoewel er een grote hoeveelheid onderzoek naar diachronische verandering in syntactische patronen en typologie bestaat, lijkt er geen consensus te bestaan over de methodologie. Fylogenetische comparatieve methoden worden nog zelden gebruikt, maar zouden gevestigde methoden van de historische taalkunde en de diachronische typologie kunnen worden. Dit is vooral afhankelijk van de ontwikkeling van statistische modellen door taalwetenschappers die syntactische en typologische verandering op optimale wijze beschrijven. Er is geen reden waarom de typologie deze methoden niet zou kunnen omarmen om genealogische (en in de nabije toekomst, door gebruik te maken van fylogenetische netwerken) geografische verwantschappen in acht te kunnen nemen. Uiteindelijk kunnen deze methoden een bijdrage leveren aan een beter begrip van de evolutie van de diversiteit van talen.

Biographical note

Annemarie Verkerk was born in Oss (the Netherlands) in 1985. She studied *Linguistics* at Radboud University Nijmegen (propaedeutics 2005, BA 2008, *bene meritum*, MA 2009, *cum laude*) and *Structure and Variation of the Languages of the World* at Leiden University (for the academic year 2008-2009). Her MA thesis was a typological investigation of secondary predication in 63 languages. In September 2009, Annemarie started her Ph.D project *The evolutionary dynamics of motion event encoding* at the Max Planck Institute for Psycholinguistics in Nijmegen. During this project she spent one semester at Tandy Warnow's lab at the University of Texas at Austin (United States). Since December 2012, she holds a postdoctoral position on lexical evolution and diachronic typology at the School of Biological Sciences at the University of Reading (United Kingdom).

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. *Oliver Müller*
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*
38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*
42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*
67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*
69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*
76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*
78. Contributions of executive control to individual differences in word production. *Zeshu Shao*
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*
80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*
81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*
82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*
83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*
84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*
85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*
86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*
87. The acquisition of morphophonological alternations across languages. *Helen Buckler*
88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*