Annemarie Verkerk (Max Planck Institute for Psycholinguistics in Nijmegen and University of Nijmegen)

Where Alice fell into:
Motion events from a parallel corpus*

Abstract

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 sixteen Indo-European languages, this paper investigates the validity of these categories. I employ aggregation measures to present visual representations of the relationships between the languages in order to show that although some languages fit well into the category of "satellite-framed" or "verb-framed" language, others clearly do not. In line with these and other results, I propose that the Talmyan classifications only have limited use, and motion research should take into account all motion construction types when describing motion encoding.

Abbreviations in Glosses

Glosses which are not given in the Leipzig Glossing Rules are ANTIC anticausative, AUX auxiliary, EZ ezafe, PRET preterit, DEP dependent verb marking, PART particle and VF verb formative.

1.   Introduction

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

(1)   It was the White Rabbit, **trotting slowly back** again,

(2)   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

In (1) and (2), 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* 'trot'?

In this paper the question I aim to answer is whether motion event typology is better framed in terms of *types of languages* (which has been the traditional approach), or in terms of the range of motion construction types that are used within the language. I suggest that looking at rates of usage of motion construction types is the most viable approach. These motion construction strategies are shown to be used to different extents in a sample of sixteen Indo-European languages. I also suggest 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 will be used as tools to discover whether typological classes in motion event encoding are present.

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 in this paper are English, Dutch, German, Irish, Portuguese, French, Italian, Russian, Polish, Latvian, Lithuanian, Albanian, Armenian, Hindi, Persian and Modern Greek. Five languages, namely Irish, Latvian, Lithuanian, Albanian and Armenian, have not been studied before in the motion event literature that provides claims with regard to the language's satellite-framed or verb-framed nature. These languages were chosen as a

representative sample of Indo-European languages. In future work, this parallel corpus will be used to study the evolution of motion event encoding in the Indo-European language family.

In Section 2, an overview of the semantic elements involved in motion encoding will be presented, as well as their lexical expression, and the various motion encoding strategies that emerge from the different ways these lexical expressions can be combined. In Section 3, the merits of my source of data, parallel corpora, will be discussed briefly. In Section 4, an overview of the motion typology of the sixteen languages that are the focus of this paper will be presented.

2.   Motion events

2.1. Introduction

My approach to the analysis of motion expressions is heavily dependent on the ideas developed by Leonard Talmy (Talmy 1985, 1991, 2000), which I will discuss to the extent that they are relevant for the purposes of this paper. 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 (1) and (2): in the English sentence (1), the way in which the White Rabbit is moving is indicated by the verb *trot*, while it is indicated by the adverbial phrase *com pulinhos vagarosos* in the Portuguese sentence (2). The same semantic information, namely the specific way in which the White Rabbit 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 may be combined to form a range of different syntactic motion event constructions.

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 as proposed by Talmy. Slobin and Hoiting (1994: 498–499) and Slobin (2004, 2005, 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" (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 also developed 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). In this paper, I will present evidence 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.

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 I take into account here as well: figure, path, ground and manner.[1] 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 (1), the White Rabbit 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 (1), the path is the trajectory of the movement of the White Rabbit, who is moving from an undefined place back towards a place where he was before. In my framework, path (or deixis, see below) should always be encoded linguistically for a motion expression to count as a motion event.[2]

─────────

[1] 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 paper. 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 I am looking exclusively at motion events that describe transitional motion, it is not necessary to include it.

[2] One of the reviewers pointed out that viewing path as an obligatory component of motion is theory dependent, and I agree that it is. However, I find it 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

Before taking a closer look at the different encoding possibilities for path, it is necessary to focus on one of Talmy's important concepts, namely that of the satellite.

In Talmy's framework, path can either be expressed in the verb or in the satellite. Talmy (1985: 102) defined satellites as "certain immediate constituents of a verb root other than inflections, auxiliaries, or nominal arguments." 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 (3) below, the path element is a satellite, if not, as in (4), the path element is not a satellite. Beavers et al. (2010: 338) point out that the sentences in (3) and (4) 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 (3), 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.

(3) *John ran in (the house).*
(4) *John ran to *(the store).*   Beavers et al. (2010: 338)

Following Filipović (2007) and Beavers et al. (2010), I reject the strict definition of satellite as put forward by Talmy (1985) and use a broader definition as a replacement: 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 refer to motion with respect to a deictic center, rather than motion that has a certain path. Berthele (2006) points out why deictic verbs should be

————

specifying the path that the baby had, but only the location of the motion ('the room'). In this case, we are saying that the baby crawled inside the room, and no change of location has occurred. Such expressions are not part of my theory of motion events.

Table 1. Deictic verbs encountered in the sample.

Language	Deictic verbs	Reference
English	*come, go*	Ricca (1993)
Dutch	*komen, gaan*	Ricca (1993)
German	*kommen, gehen*	Ricca (1993)
Irish	*tar, gabh, téigh*	ÓBaoill (1975)
Portuguese	*vir, ir*	Ricca (1993)
French	*venir, aller*	Ricca (1993)
Italian	*venire, andare*	Ricca (1993)
Russian	no deictic verbs	
Polish	no deictic verbs	
Latvian	*nākt*	Wälchli (2001b: 414)
Lithuanian	no deictic verbs	
Albanian	*vij, shkoj*	Ricca (1993)
Armenian	*gal, gnal*	no reference
Greek	*erchomai, pigaino*	Ricca (1993)
Persian	*āmadan, raftan*	Feiz (2011)
Hindi	*ānā, janā*	Kachru (2006: 86–87)

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, I also separate deictic verbs from path verbs in this paper. Deixis is a complicated issue and is characterized by very different solutions cross-linguistically (Wälchli 2009: 230). 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 1.

Table 1 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-*. 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 for more general discussion on the identification of deictic verbs Wälchli (2009: 230ff, 2001a: 311).

The *ground* is defined as an explicitly indicated object that serves as a reference point for the motion in which the figure is engaged. This can be any type of object, from buildings to forests, and from other people, animals to household objects. The ground can also be an extended area or place, such as the air or the sea.

The *manner* is defined as the way in which the action can be carried out. In example (1), the verb *trot* 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, I employ a broad definition 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 *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 manner aspects, such as *gently* in (5). They can also be descriptive phrases that encode aspects of manner. An example of the latter type is given in (5), where the phrase *without even touching the stairs with her feet* indicates the manner in which she floated down.

(5)   *… 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 is given below in (6). 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.

(6) 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 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. I have found that some of the languages in my sample make use of manner plus path verbs. An example is Greek *skarfalono* 'climb up', which expresses both upward motion and a climbing manner.[3]

Although the terms manner and path have been in use for a long time, a classification of verbs into the classes of manner verb, path verb, or manner plus path verb is far from clear. Many verbs are in 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. In other words, many manner verbs seem to have a path preference. 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

---

[3] The existence of manner plus path verbs has also been questioned. Jones (1983) and Beavers et al. (2010: 357) argue that manner plus path verbs do not exist. Beavers et al.'s (2010) put forward in their theoretical model 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.

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 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 Greek *skarfalono* 'climb up', Persian *goriktan* 'run away', Dutch *duiken* 'dive', 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.

To summarize: there are four semantic motion elements, namely figure, path, ground and manner. 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.

2.3. Motion event encoding constructions

In this section I will provide an overview of the strategies encountered in my sample. I will start out with 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 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 (1) and (5). 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 (2) and (6).

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. An example is provided in (7).

(7)   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.'

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 my 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 (8).

(8) 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 indication of path, the manner plus path verb strategy is employed. An example is provided in (9).

(9) 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'

When a deictic verb is the main verb, we have an instance of what I call here the deictic verb strategy, exemplified in (10). Since deictic verbs may 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 (11).

(10) 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 crossed the small brook after the Queen'

(11) 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 my 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 (12). 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 co-ordinated 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.

(12) 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'

From my sample a construction that has not been discussed in the motion event literature has emerged as well. 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 Greek is provided in (13).

(13) 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 (13), the main verb is *paleyo* 'struggle' and the subordinate verb is *skarfalono* 'climb up'. 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 strategies that were 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 verb-framed and subordinate strategies at the same time – a verb-framed example from Greek is included in (14). Verb-framed strategies in Greek are characterized by using a participle form of the verb, which is different from the dependent verb marking in (13).

(14) Greek
      … *e-fyg-e*   *alafropat-ontas:*
      PST-go.away.PST.PFV-3SG   walk.delicately-PTCP.ACT
      'she left walking gently'

Aside from these constructions, an "other" category 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 build 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: 205). 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.

Table 2. Motion encoding strategies.

	Name	Components
1.	satellite-framed strategy	manner verb + path satellite
2.	verb-framed strategy	path verb + participle OR adverbial expression
3.	path only strategy	path verb, no indication of manner
4.	manner only strategy	manner verb, no indication of path
5.	manner plus path verb strategy	only a manner plus path verb
6.	deictic verb strategy	deictic verb, no indication of manner
7.	deictic verb-framed strategy	deictic verb + participle OR adverbial expression
8.	subordinate strategy	any two motion verbs, one is subordinate
9.	coordination strategy	any two motion verbs, coordinated

To summarize this section, an overview of the constructions distinguished in this paper is presented in Table 2.

3.   Parallel corpora

The data on which the current analysis is based come 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 famous 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 1996, 2005; Baicchi 2005 for similar approaches). First of all, using a parallel corpus is very suitable because motion events constitute a mostly lexical topic that is prevalent in natural language (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 paper. In addition, motion events have been largely studied by experimental methods with the parallel corpus approach, it is possible to cover a larger set of languages than would be possible via experimental research, since it is much less demanding of time.

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 I have 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* (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 discussed above). 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. The choice of three different books with different original languages should also make author- and translator-specific biases less pronounced.

From these three books, all descriptions of motion events were extracted. Motion events were defined as "situations in which an animate being moves from one place to another" following Özçalışkan and Slobin (2003: 259).[4] Each motion extract that was picked constituted a single sentence in which (approximately) a single event was being described. Such a sentence could consist of several clauses, as we have seen in (5). 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 (5), this was "floated". Examples of these motion extracts can be found throughout this paper.

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. This was done because it was not feasible to include the full set of motion descriptions – such a sample would have been too large for the purposes of this study. Care was taken to include all the variation that was present in the larger collection, i.e. of each manner verb and each path verb

---

4 Although note that I also included inanimate entities, as discussed in section 2.2.

that occurred in the larger sample, at least one instance was included in the smaller sample.

Because of this 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. This is especially relevant for the use of the deictic strategy. The deictic verbs *come*, *go* and *ir* 'go' were among the most commonly used 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 regard 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. However, the current sentence sample 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 230 sentences that encode voluntary (non-causative) motion. For the analyses presented in this paper, 124 sentences were selected from this set. These sentences are taken from *Alice's Adventures in Wonderland* and *O Alquimista*, the two books that are available for all sixteen languages studied in this paper. The total set of data available for this paper thus consisted of 124 original motion extracts and their translations in a total of 16 languages.

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. This was done in order to have some under-standing of 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, path verb, deictic verb, or manner plus path verb. The originals and translations were then coded for the features that have been discussed in 2.2.

4.   Some results and explanations

In Section 4.1, I will start with an overview of the usage of motion event en-coding strategies in the sixteen Indo-European languages. In Section 4.2, I will discuss some results from different aggregation methods used to pro-vide a more sophisticated view of the data.

4.1. Results on strategy usage

In Figure 1 an overview of the usage of the nine strategies I formulated earlier is presented. This barplot gives the frequency of usage of each strategy (on the y-axis) for each of the sixteen languages (on the x-axis). In both Figure 1 and 2, the first bar labeled "originals" gives the percentage for the original sentences, taken from English for *Alice's Adventures in Wonderland* and from Portuguese for *O Alquimista*. These are provided to give the starting point of the parallel corpus, i.e. the original set of constructions that was used. In Figure 1 and in all figures and analyses to follow, the selection of 124 sentences from the corpus mentioned before was used.

From Figure 1 it becomes clear that all languages use most of the motion encoding strategies available to them, but do so to different extents. 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. It also becomes clear from Figure 1 that the sixteen languages under investigation seem to range themselves along a cline with regard to the use of the satellite-framed strategy. These are used most often in the Russian sample, with over half of the sentences attested in this corpus using the satellite-framed construction. This strategy is used the least in the Albanian sample – in fewer than 20 % of the sentences. The cline in usage of the satellite-framed strategy is paralleled partly by a cline in the path-only strategy, which becomes more common as one moves from the left side to the right side of Figure 1. 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, Portuguese, French, Hindi, Persian and Albanian, but not as much for Armenian. The coordinate strategy is quite often used by Armenian and Hindi, while the remaining strategies are less common.

The encoding patterns that are found in the current data set correspond with what is known about motion descriptions in these languages. In Table 3, an overview of classifications made in the literature on motion events is presented. Several languages in the sample, namely Irish, Lithuanian, Latvian, Armenian and Albanian, have not been described in the literature on motion encoding before, and are therefore not listed in Table 3.

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 and Dutch on one hand and Portuguese,

Table 3. Motion encoding classifications made in the literature.

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

French, Italian and Hindi on the other, with Greek and Persian somewhere in between. However, this is not what we observe in Figure 1. There is a steady decline in the use of satellite-framed strategies and an incline of path-only strategies if we move from the left-most language to the right-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 left side of the plot we find languages which use satellite-framed strategies 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 paper, I will call these languages "manner-salient", reflecting a partial habit to encode manner on the main motion verb.[5] Clear manner-salient languages are Russian, Dutch, Polish, Lithuanian, German, English, and Latvian. On the right side of the plot we find languages which use satellite-framed strategies less often than the "originals", and which use more path-only strategies and verb-framed strategies. I will call these languages "path-salient", reflecting their partial tendency to encode path on the

------

   [5] This term is of course borrowed from Slobin (2004).

main motion verb. Clear path-salient languages are Greek, Italian, French, Portuguese, and Albanian. These terms are a bit more transparent than the traditional terminology, as they reflect the semantics often encoded on the verb.

However, there are also languages that do not really fit into one of these two traditional classes. Irish seems to follow a satellite-framed pattern easily[6] and more often than the path-salient languages, but uses a quite large amount of deictic verb strategies. Hindi, unlike other languages traditionally classified as verb-framed, does not use the path-only strategy as much, but especially uses the deictic verb strategy, the coordinate strategy, and the sub-ordinate strategy. Persian also deviates from the path-salient 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, Armenian and Persian are different from manner-salient and path-salient languages in different ways, it seems more useful to classify languages with regard to their usage of the different motion encoding strategies.

In Figure 2, the usage of the three most common strategies used to encode manner are shown separately from the other strategies: the satellite-framed, the verb-framed, and the deictic verb-framed strategy. The variation depicted in Figure 2 seems to be mostly due to the rates of use of manner verbs that declines as we go from the left to the right, as was shown in Figure 1. Verb-framed and deictic verb-framed strategies are used to the same extent both by some of the manner-salient languages (Dutch, English) and some of the path-salient 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 2 is that the Balto-Slavic languages Russian, Polish, Lithuanian and Latvian seem to avoid the usage of verb-framed strategies. There are some instances of the use of verb-framed strategies with manner adverbials, but verb-framed strategies with manner verb participles are quite rare (Russian: none; Polish: 2; Lithuanian: none;

---

6 With 'easily' I mean here that Irish freely uses satellite-framed patterns in bound-ary-crossing situations, unlike path-salient 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.'
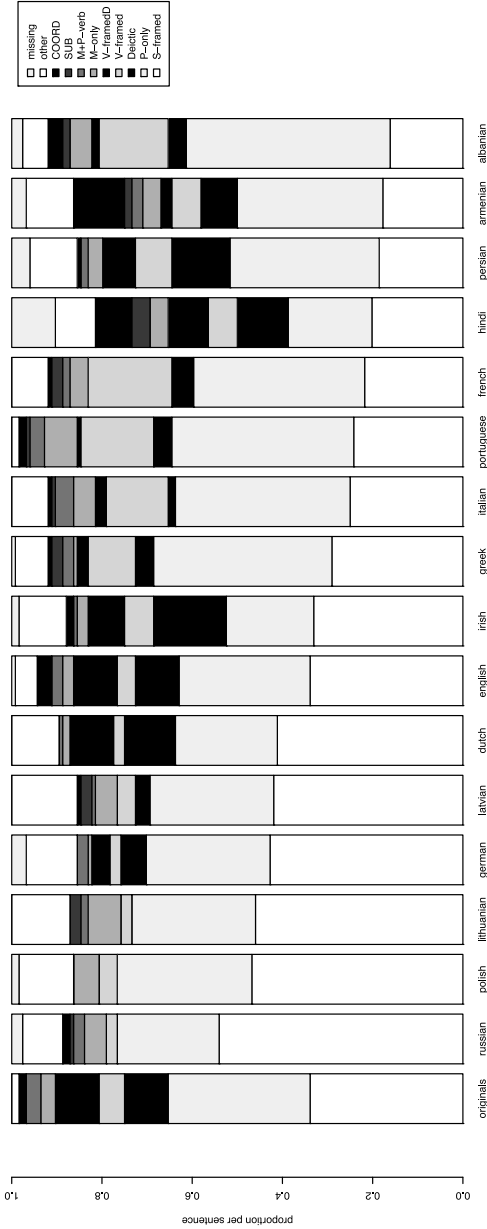
Figure 1. The frequency of the usage of nine different motion encoding strategies in sixteen Indo-European languages.
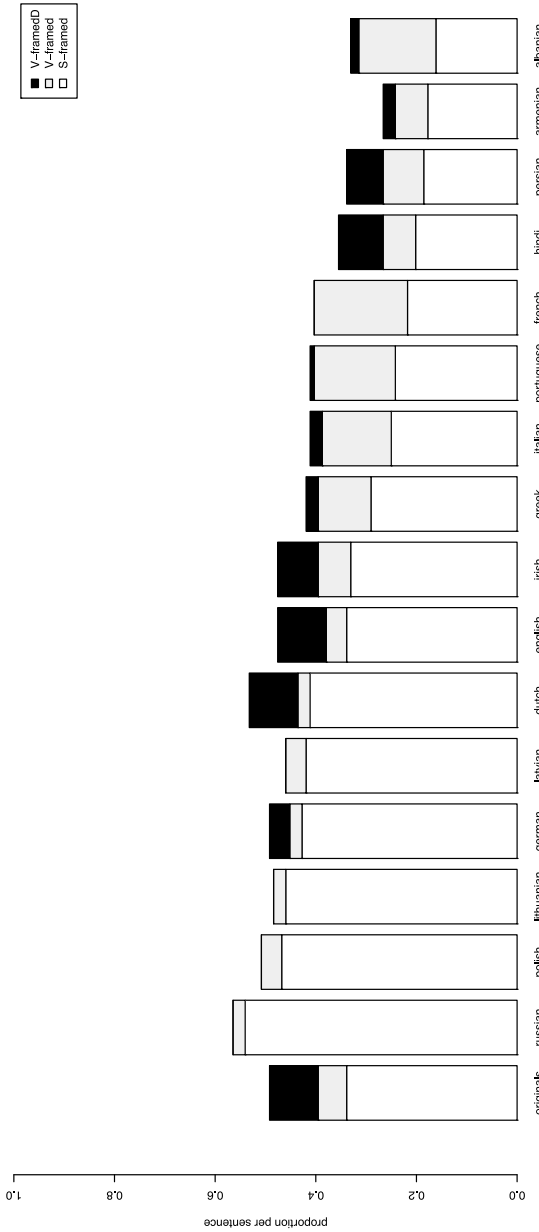
Figure 2. Means of manner expression in sixteen Indo-European languages.[7]

_____

[7] Only the three most common strategies to encode manner are included in this graph.

Latvian: 2). There seems to be a large pressure for these languages to encode manner on the main verb, as is evident from Figure 2 and illustrated by the Russian example in (15). In this example, the English original has a deictic verb-framed pattern (*came running*), which is translated with a satellite-framed pattern in Russian, Polish, Lithuanian and Latvian.

(15) 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'

Figure 2 also shows that Greek, Italian, French, Portuguese 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 these strategies are quite "heavy" with regard to processing load (Slobin 2004: 229). The native pattern for path-salient languages is that manner information is often not included, but may sometimes be inferred from the context. Adding the same amount of manner information by means of verb-framed strategies 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 2.[8] 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 8) 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 more or less path-salient or more or less manner-salient, for other languages different classifications have to be made.

———————

8 However, note the use of the coordinate strategy and the subordinate strategy that feature a manner verb is not included in Figure 2. Languages that make use of these strategies, such as Armenian and Hindi, therefore encode slightly more manner as is depicted in Figure 2.
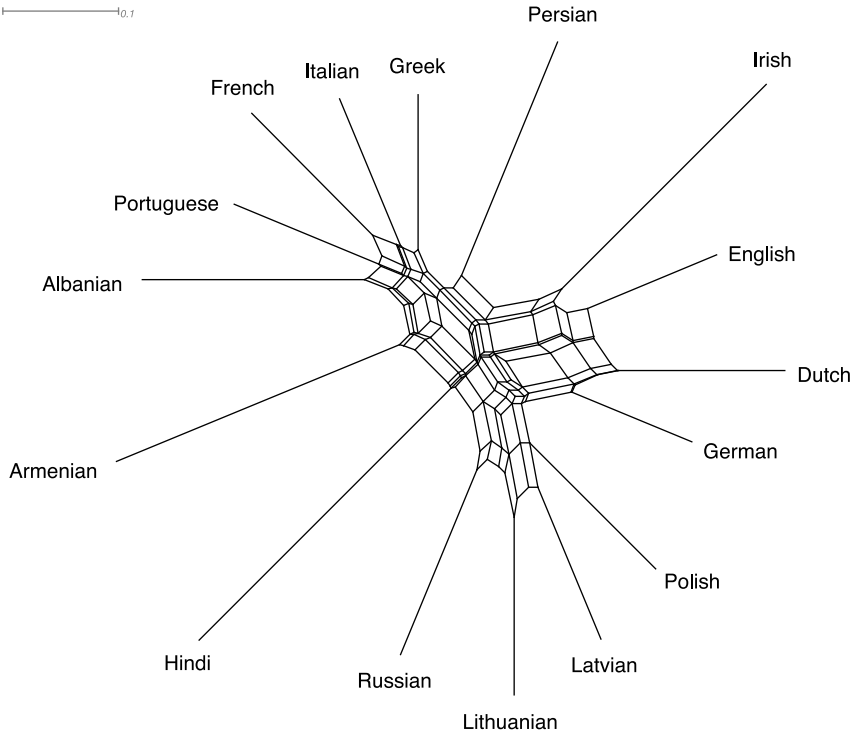
Figure 3. A split graph showing the results of a Neighbor-Net analysis of motion encoding constructions used in 16 Indo-European languages.

4.2. Aggregation analysis: A demonstration

The frequency tables presented in Section 4.1. give an indication of how often a strategy is used by each language. However, it does not take into account the relationships between different languages with regard to the choices made 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, the results of a Neighbor-Net analysis on the nine motion encoding constructions distinguished in this paper is presented.[9] A picture emerges that overlaps with the frequency bar plot in Figure 1. Three groupings emerge: Russian, Lithuanian, Latvian, and Polish (Balto-Slavic); Irish, English, and Dutch ("Germanic+Irish"); and Greek, French, Albanian, Portuguese, and Italian ("Romance+Balkan"). German patterns in between the Balto-Slavic and the Germanic+Irish group, while Hindi, Armenian and Persian appear between the Germanic+Irish and the Romance+Balkan group. A 1000-fold bootstrapping on the split graph in Figure 3 results in an almost star-shaped network, but retains the three groupings with one exception: In the 1000-fold bootstrapping, German is included in the Germanic+Irish group.

It is clear from Figure 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 Verkerk (to appear). However, there are also divergences from the phylogenetic pattern: German patterns closely with English and Dutch, as expected, but also seems to be pulled in the direction of the Balto-Slavic languages; and Albanian, a non-Romance language, is placed in the Romance group.

A first interpretation of Figure 3 could be that divergences from the phylogenetic pattern are due to language contact: maybe German is situated more closely to the Balto-Slavic grouping because of contact with its neighbor Polish? 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, Bryant, and Greenhill 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, Bryant, and Greenhill (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

---

[9] 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 meaningful category.

higher delta score than the other Germanic languages (Gray, Bryant, and Greenhill 2010).

For the current analysis presented in Figure 3, the average delta score is 0.36. Languages that have a higher delta score are Armenian (0.41), Hindi (0.39), and Russian (0.37). Languages that have a lower delta score are French (0.31) 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, Bryant, and Greenhill (2010) for Sranan. While Armenian has been influenced by contact with both Indo-European and non-Indo-European languages for centuries, Russian would normally not be characterized as heavily influenced by other languages (see Thomason and Kaufman 1988 for a different view). Also, a contact language like Greek does not have a higher delta score (0.36).[10]

Since language contact does not provide a ready explanation for these patterns, it seems that the higher delta scores for Armenian, Hindi, and Russian suggest a mixed pattern in the type of motion event encoding constructions that are being used. This means that, for a part of the dataset, these languages are similar to certain languages, while for another part of the dataset, they pattern similarly to other languages. The Neighbor-Net analysis presented in Figure 3 can therefore in the first place be interpreted as a map of typological types: a very clear path-salient group (Italian, Greek, French, Albanian, Portuguese) a clear manner-salient group which doesn't use the deictic verb strategy (Russian, Lithuanian, Latvian, Polish), and a manner-salient group which does use the deictic verb strategy (Dutch, Irish, English). The rest of these languages do not immediately belong to one of these groups. Note that if Talmy's (1991) dichotomy would be 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. The Neighbor-Net plot in Figure 3 therefore also supports the suggestion that Talmy's (1991) dichotomy is a reduction of the actual variation that is present in motion encoding.

Figure 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

---

[10] One might suspect that the high delta scores for Hindi and Armenian are caused by the fact that these languages are the only languages of their subgroup 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 both Armenian and Hindi having a score of 0.46.

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, Hindi and Russian 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 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 6 dimensions seems appropriate ($R^2 = 0.69$), but since the first three dimensions already present a clear picture and explain a large portion of the variance ($R^2 = 0.45$), the first three dimensions have been depicted in Figure 4. The first dimension in Figure 4 pulls apart Hindi and Armenian from all the other languages, while the second and third dimension together form a Talmyan cline. The numbers on the axes represent the distances between the languages.

The results in Figure 4 are similar to those in Figure 3. There seems to be a cline from manner-salient languages to path-salient languages in the middle of the plot, with the manner-salient languages in the bottom left and the path-salient languages to the upper right. Hindi and Armenian are removed furthest from all the other languages. As becomes clear from Figure 4, Hindi 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 manner-salient to path-salient languages is not as clearcut as it was in Figure 3: Persian is situated quite close to the manner-salient languages, while Russian is situated quite close to the path-salient languages.

By using the Neighbor-Net and the MDS analysis, conflicting typological signals were found in the following languages: Russian, Armenian, and Hindi (which had a higher delta score), German (which is situated in between the
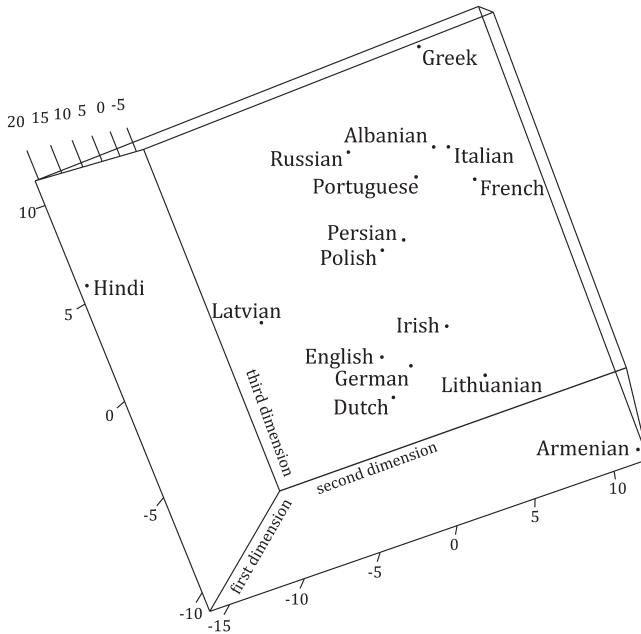
Figure 4. A 3D map of the first three dimensions of a classical
multidimensional scaling analysis on motion encoding constructions
used in 16 Indo-European languages.

Germanic and the Balto-Slavic languages in Figure 3), and Persian (which is
located most closely to the Germanic and Balto-Slavic languages in Figure 4).
Some reasons for this are presented below.

Russian was one of the three languages with a high delta score, and while
it was included in the Balto-Slavic group in Figure 3, it was situated on the
edge of that group, showing some affiliation with the path-salient group.
This was also evident from Figure 4. Looking more closely at the motion
encoding strategies chosen for individual sentences, I could not discern a
clear reason why Russian behaves differently from the other Balto-Slavic lan-
guages. Given more data, this could become clearer, or the difference could
disappear.[11]

---

[11] I have also conducted a Neighbor-Net analysis including also the data from *Through
the Looking-Glass and what Alice found there.* This analysis was conducted using data on
13 languages, including Russian. In that analysis, Russian is also placed on the edge
of the Balto-Slavic group. Close scrutiny of the data included in this larger analysis
does not provide a single clear reason for Russian's position either.

As has become clear in Figure 1 and Figure 4, both Hindi and Armenian cannot be said to belong to either the group of manner-salient or path-salient languages. Neither language employs the satellite-framed strategy very much, but both employ the coordinate and deictic verb strategy relatively often. Even though Hindi and Armenian are still quite dissimilar, there are a few cases where they match, which seems to give rise to the placement of Armenian and Hindi closely together in Figure 3. Both the Neighbor-Net and the MDS analysis give a clear assessment of Armenian and Hindi as not belonging to either of the two typological types.

German is positioned on the edge of the Germanic group in Figure 3, displaying some affinity with the Balto-Slavic group. One of the reasons for this seems to be the difference in the use of deictic verbs between Dutch and German. Dutch uses the deictic verb strategy more often than German, and a closer inspection of the data reveals that for a set of cases, an original path-only or a deictic verb strategy is translated with a path-only or a satellite-framed strategy in German. Because most of the Balto-Slavic languages do not have deictic verbs, German is pulled slightly into the direction of the Balto-Slavic languages.

The mixed typological nature of Persian has also become clear from Figure 3 and 4. The placement of Persian in between the path-salient and the manner salient languages in Figure 3 and close to manner-salient languages such as Dutch and English in the MDS analysis in Figure 4 might be due to Persian's use of deictic verbs, which is similar to that of Irish, English, and Dutch. The position of both German and Persian seems to be dependent on the use of the deictic verb strategy. As far as I am able to tell from the current dataset, the use of the deictic verb strategy seems to act independently from the Talmy typology of manner-salient and path-salient languages.

Even though all of the patterns I discussed here are very tentative and need 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 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.

5. Conclusion

In this paper, I have presented data on motion encoding from sixteen Indo-European languages, five 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. However, the biggest difference between Talmy's (1991) approach and my own is that Talmy proposes a dichotomy of language types, while I have tried to show 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 path-salient class and within the manner-salient class, and cannot account for languages that do not belong to either of these classes.

The aggregation methods employed in Section 4.2. 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 Armenian, Hindi, 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 paper with the help of aggregation methods will be the focus of future work. In that future work 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 (Verkerk to appear).

The Neighbor-Net analysis and the MDS analysis conducted in Section 4.2. are not only useful for an assessment of Talmy's (1991) dichotomy. Generally, these methods are used to make the groupings present in the data explicit. These 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 the different dialects. Explanations for these relationships can then be sought using different methods, for instance using the multivariate spatial analysis proposed by Grieve (this volume) 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.

References

Baicchi, Annalisa   2005 Translating phrasal combinations across the typological divide. In: Marcella Bertuccelli Papi (ed.), *Studies in the Semantics of Lexical Combinatory Patterns*, 487–519. Pisa: Pisa University Press.

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

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

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

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

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

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

Filipović, Luna   2007 *Talking about Motion: A Crosslinguistic Investigation of Lexicaliz-ation Patterns.* Amsterdam: John Benjamins.

Folli, Raffaella   2008 Complex PPs in Italian. In: Anna Asbury, Jakub Dotlačil, Berit Gehrke, and Rick Nouwen (eds.), *Syntax and Semantics of Spatial P*, 197–220. Amster-dam: John Benjamins.

Gray, Russell D., David Bryant, and Simon J. Greenhill   2010 On the shape and fab-ric of human history. *Philosophical Transactions of the Royal Society* 365: 3923–3933.

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

Huson, Daniel H. and David Bryant   2006 Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

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

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

Kachru, Yamuna   2006 *Hindi*. Amsterdam: John Benjamins.

Kopecka, Anetta   2006 The semantic structure of motion verbs in French: Typologi-cal perspectives. In: Maya Hickmann and Stéphane Robert (eds.), *Space in Languages: Linguistic Systems and Cognitive Categories*, 83–102. Amsterdam: John Benjamins.

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

Narasimhan, Bhuvana   2003 Encoding complex events in Hindi and English. *Pro-ceedings of the 1999 Conference on Perspectives on Argument Structure.* Austin: Texas Lin-guistic Forum.

Nesset, Tore   2009 Metaphorical walking: Russian *idti* as a generalized motion verb. In: Victoria Hasko and Renee Perelmutter (eds.), *New Approaches to Slavic Verbs of Motion*, 343–359. Amsterdam: John Benjamins.

ÓBaoill, Dónall   1975 Deixis in Modern Irish and certain related problems. *Ériu* 26: 144–161.

Özçalışkan, Şeyda and Dan. I. Slobin   2003 Codability effects on the expression of manner of motion in English and Turkish. In: A.S. Özsoy, D. Akar, M. Nakipoğ-lu-Demiralp, E. Erguvanlı-Taylan and A. Aksu-Koç (eds.), *Studies in Turkish lin-guistics: Proceedings of the Tenth International Conference on Turkish Linguistics*, 259–270. Istanbul: Bogazici University Press.

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

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

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

Ricca, Davide   1993 *I verbi deittici di movimento in Europa: una ricerca interlinguistica.* Firenze: La Nova Italia.

Slobin, Dan I.   1996 Two ways to travel: Verbs of motion in English and Spanish. In: Masayoshi Shibitani and Sandra A. Thompson (eds.), *Grammatical Constructions: Their Form and Meaning*, 195–219. Oxford: Clarendon Press.

Slobin, Dan I.   2003 Language and thought online: Cognitive consequences of linguistic relativity. In: Dedre Gentner and Susan Goldin-Meadow (eds.), *Advances in the Investigation of Language and Thought*, 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 and Ludo Verhoven (eds.), *Relating Events in Narrative: Typological and Contextual Perspectives*, 219–257. Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, Dan I.   2005 Relating narrative events in translation. In: Dorit Ravid and Hava B. Shyldkrot (eds.), *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman*, 115–129. Dordrecht: Kluwer.

Slobin, Dan I.   2006 What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In: Maya Hickmann and Stéphane Robert (eds.), *Space in Languages: Linguistic Systems and Cognitive Categories*, 59–82. Amsterdam: John Benjamins.

Slobin, Dan I. and Nini Hoiting   1994 Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: University of California.

Talmy, Leonard   1985 Lexicalization patterns: Semantic structure in lexical forms. In: Timothy Shopen (ed.), *Language Typology and Syntactic Description: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.

Talmy, Leonard   1991 Path to realization: A typology of event conflation. *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*, 66–168. Cambridge: Cambridge University Press.

Thomason, Sarah G. and Terrence Kaufman   1988 *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

Verkerk, Annemarie   to appear Diachronic change in Indo-European motion event encoding. *Journal of Historical Linguistics*.

Wälchli, Bernhard   2001a A typology of displacement (with special reference to Latvian). *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 and Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic Languages vol. 2: Grammar and Typology*. Amsterdam: John Benjamins.

Wälchli, Bernhard   2007 Advantages and disadvantages of using parallel texts in typological considerations. *STUF* 60: 118–134.

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

Zlatev, Jordan and Peerapat Yangklang   2004 A third way to travel: The place of Thai in motion-event typology. In Sven Strömqvist and Ludo Verhoven (eds.), *Relating Events in Narrative: Typological and Contextual Perspectives*, 159–190. Mahwah, NJ: Lawrence Erlbaum Associates.