Christiane von Stutterheim*, Johannes Gerwien,
Abassia Bouhaous, Mary Carroll and Monique Lambert

What makes up a reportable event in a language? Motion events as an important test domain in linguistic typology

https://doi.org/10.1515/ling-2020-0212

Abstract: Numerous crosslinguistic studies on motion events have been carried out in investigating the scope of the two-fold typology "path versus manner" (Talmy 1985, 2000) and its possible implications. This typological contrast is too narrow as it stands, however, to account for the diversity found both within and across types. The present study is based on what can be termed a *process-oriented perspective*. It includes the analyses of all relevant conceptual domains notably the domain of temporality, in addition to space, and thus goes beyond previous studies. The languages studied differ typologically as follows: *path* is typically expressed in the verb in French and Tunisian Arabic in contrast to *manner of motion* in English and German, while in the temporal domain *aspect* is expressed grammatically in English and Tunisian Arabic but not in German and French. The study compares the representations which speakers construct when forming a reportable event as a response to video clips showing a series of naturalistic scenes in which an entity moves through space. The analysis includes the following conceptual categories: (1) the privileged event *layer* (manner vs. path) which drives the selection of breakpoints in the formation of event units when processing the visual input; (2) the privileged category in spatial framing (figure-based/ground-based) and (3) viewpoint aspect (phasal decomposition or not). We assume that each of these three cognitive categories is shaped specifically by language structure (both system and repertoire) and language use (frequency of constructions). The findings reveal systematic differences both across, as well as within,

***Corresponding author: Christiane von Stutterheim**, Institut für Deutsch als Fremdsprachenphilologie, Universität Heidelberg, Plöck 55, 69117 Heidelberg, Germany,
E-mail: stutterheim@idf.uni-heidelberg.de
Johannes Gerwien, Abassia Bouhaous and Mary Carroll, Institut für Deutsch als Fremdsprachenphilologie, Universität Heidelberg, Plöck 55, 69117 Heidelberg, Germany,
E-mail: gerwien@idf.uni-heidelberg.de (J. Gerwien), bouhaous@idf.uni-heidelberg.de (A. Bouhaous), carroll@idf.uni-heidelberg.de (M. Carroll)
Monique Lambert, Centre CNRS, 59 rue Pouchet, 75017 Paris, France,
E-mail: mon.lambert@wanadoo.fr

typologically related languages with respect to (1) the basic event type encoded, (2) the changes in quality expressed, (3) the total number of path segments encoded per situation, and (4) the number of path segments packaged into one utterance. The findings reveal what can be termed language-specific default settings along each of the conceptual dimensions and their interrelations which function as *language specific attentional templates*.

Keywords: event segmentation, unit formation, typological analysis, motion events, temporal aspect

1 Introduction

The present article explores a new perspective on the extensively researched field of motion events in that it takes into account both spatial as well as temporal cognitive categories which underlie the formation of motion event units across languages. Numerous studies based on Talmy's two-fold spatial typology of verb-framed and satellite-framed languages (Talmy 1975, 1985) have provided the foundations in this field. The crosslinguistic comparisons cover a range of theoretical approaches, as well as empirically based procedures, on the way different systems impact motion event descriptions (Goschler and Stefanowitsch 2013, van der Zee 2003–2013).

Following Talmy's seminal studies, typological insights have been extended in different directions both with respect to the range of languages analyzed as well as the extent to which they fit the verb-framed (path in verb) versus satellite-framed (manner in verb) dichotomy. Based on the empirical findings, the theoretical framework has been modified along different lines, as in studies reported in Slobin (2006), for example, leading to the proposition that there are more than two language types (cf. Croft et al. 2010). Other studies indicate that the relevant conceptual components can be weighted differently, a factor which would explain the language-specific preferences in information selection and perspective taking (Beavers et al. 2009, Berthele 2004). This claim is based on the fact that all languages have means to express motion events drawing on path verbs as well as manner verbs. The questions which follow, and which we set out to address in the present article, shift the focus from the description and analysis of *linguistic form* across languages to investigating the principles which underlie information selection when representing motion events.

In the present study, speakers were asked to describe a series of dynamic events showing situations in which a figure moves along an extensive path with a change in direction. In preparing a response, speakers have to form mental representations of

the visual input in the course of what Levelt (1989) has termed *the conceptualization process*. Conceptualization as a technical psycholinguistic term refers to the process in which information is selected and combined for expression in terms of the linguistic means available to the speaker. We analyze the representations formed during conceptualization as event units, i.e., clusters of information that result from the way ongoing situations in the external world are broken down into segments as defined by changes in the continuous flow of information in the visual input. There is no "objective" procedure in solving this task. There are options as to the conceptual categories which are relevant in event unit formation and the range of information selected for expression. The approach in the present study aims at detecting language-specific patterns during this preverbal process (Lupyan 2015). In focusing on the conceptual categories which are critical in the formation of motion event units across languages, the present study sets out (i) to account for the relative weight of these categories within a language and (ii) to relate the patterns observed to the structural features of the respective linguistic systems. This procedure goes beyond earlier research in two important respects:

- Focus is placed on event sequences and the processes which underlie *event unit formation*, the critical step in conceptualizing visual input for linguistic representation.
- In addition to the use of spatial concepts, the analysis takes into consideration the role of other conceptual domains,[1] notably the temporal domain, when conveying information on motion events.

Since the central question addressed on these ground concerns the processes underlying event construal, focus is placed on language-specific means with their intrinsic features and combinatorial constraints and the role they may play in the way speakers segment, select and represent motion event units. In other words, to what extent do these processes reflect clear preferences with respect to specific semantic domains and specific semantic subcategories within domains?

In addressing these questions, the experimental procedure adopted has to meet the following requirements: The languages selected should differ in the way spatial categories are encoded, as with verb-framed and satellite-framed (Talmy 2000). Furthermore, as an initial study on the role of temporal concepts on motion event construal has indicated (von Stutterheim et al. 2017), aspectual categories may also play a role in determining language-specific patterns. Based on specific divergences and convergences in both the spatial and temporal domain, the

1 We use the term *domain* in reference to an abstract conceptual field, such as *time* or *space*. This is closely aligned with the terminology in cognitive linguistics (e.g., Croft and Cruse 2004).

languages in the present study include English, German, French and Tunisian Arabic. Both French and Tunisian Arabic can be categorized as verb framed whereas English and German are generally categorized as satellite framed. With regard to grammatical aspect, English and Tunisian Arabic cover one category and thus contrast with French and German.

In the elicitation task, video clips showing everyday motion events, such as 'a person walking along a path and into a building' were used as stimuli. These naturalistic stimuli were chosen in order to avoid the potential problem of different degrees of familiarity with an artificial input across the different groups of speakers, as could be the case with cartoon-type scenes or abstract animations. The data obtained from the speakers of the four languages were analyzed with respect to patterns of event unit formation, covering both quantitative as well as qualitative aspects. In quantitative terms, differences were expected to emerge in the number of units formed when representing a given scene, while at a qualitative level, the path segments represented were expected to be based on different conceptual categories. The findings thus contribute to questions concerning relevant typological categories and associated language specific processes when extracting information from the visual input. This leads to the formation of what we term reportable event units (Carroll and von Stutterheim 2011; von Stutterheim and Nüse 2003; von Stutterheim et al. 2017). The notion "reportable event unit" reflects the language-specific principles that circumscribe the type of information which constitutes an event, as a self-contained conceptual unit, when forming the basis for an assertion. For example, a scene from the data base showing someone leaving a building and walking away, is frequently described as 'a man has left' in the perfective in Tunisian Arabic. By contrast, speakers of German always provide information on the ground as in 'a man walks out of a building'.

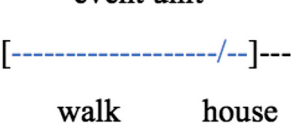
2 Research background

2.1 Different perspectives on the conception of *event units*

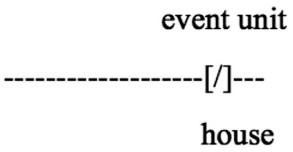
The analytic notion *event* is defined differently not only across different disciplines, but also within the field of linguistics. In cognitive theories, events are viewed as conceptual units defined by perceived changes in quality between two breakpoints in the external world (Newtson and Engquist 1976; Radvansky and Zacks 2014; Zacks and Tversky 2001). With this point of view, event schemata or frames are stored in long term memory and automatically accessed during cognitive processing. Events are defined on a different basis in semantic theory

in that the notion refers to a specific semantic unit that corresponds to a sentence (e.g., Davidson 1980; Kamp 1979). The definition thus relates to quality/time relations (Klein 2010; Koenig 2016), as expressed by verbs and their arguments (Levin and Rappaport 2005). Along these lines, the notion of a *reportable event* is defined as a unit which has a (finite) verb as its core. In the present analysis a distinction is drawn between a narrow and a broad definition. Given a broad definition, events cover all types of situations which are dynamic. In this respect it is an umbrella term for activities, accomplishments and achievements, in the sense of Vendler (1957). Given a narrow reading, it refers to situations in which changes in quality are implied. This could be represented either explicitly, with changes expressed in the predicate (*to enter*), or by reference to the boundary of a dynamic situation, without specification of the quality following the boundary, as in *walk for an hour*. Klein (1995) proposes a theory which treats the semantic correlates of sentences as time/argument relations and distinguishes between 0, 1-, and 2-state contents. Given this point of view, *to enter* would be classified as a 2-state-verb, as with *to walk for an hour*. This means that some form of breakpoint is required if 2-state contents are related to external situations. The present analysis is based on a broad definition of the term *event* and includes sub-categorizations between *1-and 2-state events* (referred to as $event_1$ and $event_2$ in the following, see Klein 1995). It adopts an integrative view with the assumption that mental processes involved in speech production (perception, segmentation, conceptualization and the construction of meaningful units) are based on event models that are shaped linguistically (Gerwien and von Stutterheim, 2018). This implies that speakers of languages which differ in their structural properties may vary in the way they segment, select and encode specific types of events when describing a series of dynamic situations, as presented for example in video clips.

With respect to motion events, the question arises as to which aspects of a state of affairs in which 'a figure moves along a path' will be selected as relevant breakpoints in the formation of event units. For example, a sentence in German such as *eine Frau geht in ein Haus* (a woman goes into a house) encodes the event type by means of the deictic verb *gehen,* a form which does not imply a change in quality. The relevant breakpoint in unit formation, given a 2-state-event, is provided by the prepositional phrase which refers to a potential boundary crossing at the endpoint of the path traced by the figure (into a house). The situation described can consist of an extended 'walking' component which is bounded by a breakpoint at goal.



In French, by contrast, a description such as *une femme entre dans une maison* 'a woman enters in a house' encodes a change of place ($event_2$) by means of a path verb. The event unit encompasses the boundary-crossing component of the situation.



These two representations of a given scenario are not equivalent at the level of event units. The difference lies in the implications given with the event type determined by manner of motion versus path. Based on research in event semantics, these different levels can be viewed as *layers* of one *thick event* (see Bennett 2002; Gerwien and von Stutterheim 2018). A situation in the external world can consist of various occurrences proceeding at the same time in parallel, as with a motion event situation where a person 'is running' and simultaneously 'approaching' some reference point. A situation can include clusters of both changes in quality as well as qualities maintained over a period of time. Based on their language, speakers will select one layer in forming a reportable unit. Talmy (1985, 2000) considers the question of simultaneous ongoing occurrences when referring to manner (and cause) on the basis of what is termed a *co-event* of a path event (see also Matsumoto 2003). This view is not adopted in the present analysis of event unit formation. In contrast to Talmy who views path as the core dimension of a motion event, with others downgraded to the status of *co-events*, the notion *event layers* accounts for the fact that different dimensions are available when construing an event unit from the material extracted from the input. On this basis there is no universal privilege for one *layer* over the other. With respect to motion events, the basic layers consist of a *manner layer* and a *path layer*. Other possible layers include factors such as an *intentional layer* related to the moving figure, or that of a *witness* based on the viewpoint of the observer.[2] Although both layers are inseparably linked in the input in the

2 A study on motion events in Japanese suggests that in Japanese the "witness layer", as reflected in the use of a deictic head verb, may be of central importance (Mano et al. 2019).

examples above, the *manner* layer is selected in German while the *path* layer is selected in French. Different units are formed, depending on the layer accessed in providing the basis for verb selection and thereby the event type.[3] $Event_2$ unit formation requires the identification of a (potential) breakpoint and what counts as a breakpoint will depend on the layer selected. The way in which these different options are manifested in language-specific patterns of information selection is one of the core questions in the empirical study.

2.2 Motion events in cognitive typology

Differences between speakers of verb-framed and satellite-framed languages constitute the starting point in a wide range of studies on language production, comprehension and acquisition of this event type (e.g., Bylund et al. 2013; Carroll et al. 2012; Flecken et al. 2015b; Levinson 2003; Majid et al. 2004; Papafragou et al. 2008). Slobin (2006) assumes that this basic contrast leads speakers of satellite-framed languages to pay more attention to manner of motion, since manner is typically expressed in the main verb with "path" expressed by a satellite, while the reverse holds for speakers of verb-framed languages who are led to focus on direction. He speaks of a "cline of manner salience" along which languages can be typologically ordered (Slobin 2006).

But things are not that systematic: many crosslinguistic analyses show a high degree of variation within language types. For example, languages where path is typically encoded in the verb (French, Italian, and Spanish) also show relatively high frequencies in the use of manner verbs (Cardini 2008, 2012; Kopecka 2009; Naigles et al. 1998). Likewise, English, categorized as satellite-framed, displays frequent use of path verbs (Carroll et al. 2012). This leads to questions concerning the factors which govern the selection of one pattern over the other in actual language use. Drawing on a survey of the literature, Pourcel (2005) points to other sources which drive variation in the conceptualization of motion events, apart

3 Jackendoff (1990:88) describes verbs encoding manner of motion as forms that do not include a path component. Given the fact that manner verbs combine with path adjuncts according to language specific rules, Jackendoff states that in English, for instance, manner verbs can combine MOVE with GO, which is not the case for Spanish or Japanese. "The claim is that languages may differ not only in their syntactic patterns but also in their correspondence rules – the ways of mapping from syntax to conceptual structure" (1980:225). Jackendoff views verb-argument structures as the level at which rules can be formulated. This implies that manner verbs are basically treated as similar across languages. Since the present analysis starts from the perspective of event construal, this allows the integration of the contrasts described in Jackendoff into language specific patterns of event construal in which the function of manner verbs is differentiated as required.

from typological features: (a) the type of entity with respect to the features human/non-human, whereby human figures encourage path over manner in contrast to non-human figures which encourage manner; (b) the type of manner, as with marked forms of motion such as *to limp* or *to dash*, which lead to a manner bias, in contrast to default means of moving (e.g., *to walk* with a bias toward path); and (c) the type of path, whereby a long trajectory often entails a bias toward manner, with a bias toward path given a short trajectory. According to Pourcel, these trends have varying weights in different languages. Durst-Anderson et al. (2013) also present a more differentiated picture. In a comparison of Russian, English and Danish they claim that the "same" representation of a motion event can lead to three different *linguistic images* in terms of mental focus (2013:129). The different *images* can be traced back to different points of departure when selecting content for expression. In Russian this involves the *figure*, in contrast to the *ground* in Danish. In this case, *language-related images* are viewed as homologous to *perceptual and conceptual images* where the grammatical system of a language is the major determining factor. The study focuses on languages with manner verbs at the core of motion event representation. While the basic idea of "different starting points" in event construal is related to a certain extent to the present approach in this study, the restriction to languages with manner of motion as the core concept in motion event representation leaves relevant questions unanswered.

There are relatively few studies to date which focus on segmentation processes of the perceptual input for event unit formation in the context of motion events. Gerwien and von Stutterheim (2018) investigated patterns of event unit formation by French and German speakers in a verbal and non-verbal task. The findings show that the two groups of speakers differ when segmenting the same visual input (video clips showing extended motion events) with respect to the number of utterances produced: French speakers produced more utterances, in comparison to German speakers. They also segmented the visual input more frequently in the non-verbal task. The results were related to typological differences between French and German in the domain of space. The present study elaborates on these findings with the inclusion of more languages and a detailed qualitative analysis of the type of information provided by speakers of the different languages across the relevant conceptual domains.

2.3 Time and space

While Talmy (2000) points to the interdependence between spatial and temporal categories in his framework on spatial cognition, few studies have taken up this

idea and investigated the role of temporal concepts in the construal of motion events. The study by von Stutterheim et al. (2017) on motion events in Standard Arabic and Tunisian Arabic sheds light on the role of aspectual distinctions and the complementary function of time and space in the representation of this type of event. The aspectual systems in these two varieties of Arabic differ in that Standard Arabic has a grammaticalized opposition between the perfective and imperfective, while Tunisian Arabic includes a grammaticalized progressive in addition to the perfective and imperfective. The core finding concerns the implications of aspectual marking for conceptualization in the spatial domain: while Tunisian Arabic, in contrast to Standard Arabic, does not have path verbs to express directed motion, speakers draw on temporal aspect on a systematic basis to express the concept "progression along a route". The findings show that typological variation is not confined to the conceptual categories under focus to date such as figure, path, ground, manner, but should also include options that serve in profiling the course of events over time.

Insights into the systematicity in the range of possible options can be gained by going beyond the domain of space, as revealed in studies on progressive/imperfective markers and event construction when representing a given scenario (Bylund et al. 2013; Durst-Andersen et al. 2013; Schmiedtová and Sahonenko 2008; von Stutterheim et al. 2012). While speakers of languages such as German or Dutch, which do not have grammaticalized aspect, tend to take a holistic viewpoint when describing a motion event, speakers of aspectual languages can select a phase, i.e., a segment of the scene in representing an event. The same scene in which a woman is shown running towards the entrance to a train station may thus be encoded as *Eine Frau läuft in den Bahnhof* (a woman runs in(to) the station) in German, but in Tunisian as (1) *mra teğri* (2) *dāxla li-l-maḥaṭṭa mtā ʿ l-car* ((1) a woman runs IMP (2) enters (PROG) to the bus station, i.e., the woman runs entering the station). The findings provided the starting point in testing the relationship between spatial and temporal properties in the construal of motion event units crosslinguistically on a systematic basis.

3 Typological properties of French, English, Tunisian Arabic and German in the domains of space and time

The selection of French, English, German and Tunisian Arabic as languages in the present study is based on typological differences in the domains of time and space, as outlined above. The relevant contrasts are summarized in Table 1.

Table 1: Contrasts in the domains space and time between the languages under investigation.

	English	French	German	Tunisian
Space	Satellite-framed	Verb-framed	Satellite-framed	Verb-framed
Time	Aspect fully grammaticalized	Aspect not grammaticalized	Aspect not grammaticalized	Aspect fully grammaticalized

The impact of spatial and temporal typological features can be readily cross-tested given the different characteristics of these four languages. German and English correspond in the spatial domain, but differ in the temporal domain. French and German correspond in the temporal domain but differ in the spatial domain, while French and Tunisian correspond in the spatial domain but diverge in the temporal domain; English and French diverge along both dimensions, just as German and Tunisian. A more differentiated picture of the relevant features within the different systems will be drawn in the following.

English

English is typically categorized as *satellite framed* in all typological studies with information on manner of motion in the verb and information on the path in non-verbal constituents such as prepositional phrases or adverbs. Manner verbs can combine with more than one satellite or adjunct, allowing a series of path segments in one reportable unit, as in *a boy is running across the street into the house*.[4] As mentioned above, however, English also possesses a range of path verbs, mainly of Romance origin. With respect to the spatial concepts used to describe the path of motion, a distinction can be drawn in relation to the source from which they are derived. Concepts which circumscribe the path of motion can be derived from features of the ground, as with adjuncts formed by prepositions such as *around, over, along*, in conjunction with a manner verb, or from the figure in motion, as in verbs such as *to head for, to advance, to approach, to enter, to exit* (Carroll 2000; Carroll et al. 2012). These two types of spatial concepts are relevant in English when identifying segments of the trajectory traced through space. With respect to temporal categories, English has a grammaticalized aspectual system with an opposition between the simple and the *-ing* form. Obligatory aspectual marking in specific contexts leads speakers to attend to the phasal structure of a given situation, leading to phasal decomposition, as the case may be.

4 Cf Slobin's studies (e.g., 2004) on complex paths in English descriptions.

German

German is the other satellite-framed language in the set. In encoding motion events, verbs provide information on manner of motion, while adverbs or prepositional phrases are used in reference to path. The language also has forms which are termed *verb particles*. In the present group of languages this specific category exists in German only. Particles refer to features of the path and combine with a manner verb as in *reingehen* 'intowalk', *runterlaufen* 'downwalk'. They merge with the verb into one word in non-finite forms, but are separated given finite forms: *der Mann geht die Treppe hinauf* 'the man walks the stairs thither-up' versus *der Mann ist die Treppe hinaufgegangen* 'the man has the stairs thitherupwalked'. Given the fact that they are separable from the verb, a manner verb can combine with more than one particle. Since event types are determined by the verb selected, one event unit can encompass several segments of a path with a manner verb as predicate. Manner verbs can combine with multiple adjuncts (*Eine Frau läuft über die Straße in einen Laden* 'A woman walks across the street into a shop') or with particles and adjuncts: *Die Frau geht rüber in den Laden* 'The woman walks across into the shop'. With respect to the type of spatial concept, German speakers typically select ground-based concepts, as in the example above. With regard to the temporal domain, German does not encode grammatical aspect.[5] This means that there are no verb-morphological devices with implications for the use of spatial concepts.

French

French is often presented as a typical example of a verb-framed language. Directed motion is expressed by a large set of (reflexive) path verbs which combine with source- or goal-related adjuncts. The spatial concepts are typically derived from the figure in motion (*Une femme se dirige vers une église*. 'a woman directs herself toward a church). French also has a wide range of manner verbs. As noted in the literature, manner verbs in French do not combine with goal related adjuncts (Berthele 2004, 2013; Pourcel and Kopecka 2005). These forms are used in combination with adjuncts that express location, presenting what can be termed a "screenshot" of a figure involved in an activity (*une femme court dans la rue*. 'a woman runs in the street') (cf. Durst-Andersen et al. 2013:130).

5 Note that there are periphrastic verbal forms which express an aspectual meaning (*Die Frau ist am Lesen, die Frau ist dabei zu lesen* the woman is 'at to read', is at reading) but these forms are not grammaticalized in German. There are dialects in which use of these forms is frequent, but even in these varieties there is no grammaticalized opposition between two aspectual forms.

This means that manner verbs differ in function in French, compared to English or German (Carroll et al. 2012; Gerwien and von Stutterheim 2018). In the present study we will see in how far this feature affects event unit formation. With respect to temporal categories, French is similar to German with the absence of grammaticalized aspect, which would require restriction of an assertion to a specific phase of a situation.[6]

Tunisian

Tunisian is classified as a verb-framed language (cf. a detailed description in von Stutterheim et al. 2017). In contrast to French, the number of path verbs in Tunisian is limited as there are no verbs which explicitly express direction with a goal orientation (e.g., *to advance toward, to head for*). This means that although Tunisian differs regarding a core feature of verb-framed languages, it conforms in other respects to the verb-framed pattern. Furthermore, manner verbs cannot be used to encode directed motion. As in French, they combine with information on the location of the figure in motion only (*mra teġri f-el-kayās*, a woman runs (IPF) in the street, cf. Louhichi 2015).

A note is in order concerning the verb *mša*, which is used frequently in the present study and is generally translated as *to walk (mra temši f-el-kayās quddām ʿimāra* 'a woman walks (IPF) in the street in front of a building'). In Tunisian the meaning has bleached in the sense that it is no longer restricted to a specific manner of motion, but can be viewed as a neutral verb. It expresses what Talmy (2000) has termed *motion* without features of either manner or path.[7]

As mentioned above, Tunisian belongs to the group of aspect languages. Finite verbs are marked for aspect which leads to the selection of specific phases of a situation for assertion.[8] The categories expressed are the imperfective (prefixed), perfective (suffixed) and progressive (semi-finite participle active). Use of the

6 We are aware of the fact that there are aspectual distinctions between the *imparfait* and the *passé composé/passé simple*. However, it is not necessary to go into the discussion on the aspectual status of these forms, since speakers use the present tense only in the present data. In this context French does not have a grammaticalized aspectual opposition.

7 It should be noted that the verb *mša* is not equivalent to verbs such as *gehen* (go) in German in the sense that it is used to refer to a concrete motion event of a person 'walking' from A to B. This form is restricted to persons and this form of motion; movement by a vehicle, or a bird, for example cannot be described as *gehen*, in contrast to *mša,* which is not restricted to any specific manner of motion.

8 Details are provided on Tunisian since readers may be less familiar with this system compared to the other languages.

perfective, given the case of path verbs, requires a temporal point of reference which may be located at the beginning of a motion event, thereby signaling entry into a situation (*mša* TA: 'move' ('go away') PF.3SGM), or the termination of an event, signaling a point of completion (*dxal* TA: 'enter' PF.3SGM). Path verbs in the perfective describe a situation as 'goal reached', as in (1).

(1) *qaṭṭos* *dxal* *li* *l-bīt*
cat enter-PF.3SGM to DEF-room
'a cat entered the room'.

The opposition between the imperfective and the progressive can be described as follows: Path verbs in the progressive are interpreted as referring to a path that leads incrementally up to a goal, or away from a point at source. With use of the progressive, the actual state of having reached the endpoint or a goal region is not asserted (2).

(2) *rāğel* *dāxel* *li* *s-supermarché*
man enter-PART.3SGM to DEF-supermarket
'a man is entering the supermarket'.

By contrast, the imperfective cannot be used in the context of goal-oriented motion, as in (3). In motion events, the perspective implied by the use of the imperfective is that of a situation without an envisaged goal.

(3) *rāğel* *yḥawwes* *fi* *š-šāri*
man go-out-IPF.3SGM in DEF-street
'a man is going out for a walk in the street'.

These constraints result in language-specific combinatorial properties of spatial and temporal concepts. The progressive relates to a goal-related situation, the perfective to a situation where a goal is reached. Change-of-state predicates combine with these two aspectual markers. The aspectual category expressed by the imperfective is confined to one-state predicates. Note that this differs from English since the *-ing* form can combine with change-of-state predicates (*the woman is entering the room*). This means that manner verbs in Tunisian typically occur in the imperfective and exclude reference to a goal. If spatial information is provided, this refers to a location only, as mentioned above (cf. Example 3). Features of Tunisian which are relevant in the present study on event unit formation can be summarized as follows:

- Change-of-state path verbs + PROG/PF ➔ events expressing directed motion.
- No change-of-state manner verbs + IPF ➔ location of figure showing specific properties.

3.1 Implications for unit formation

This admittedly coarse-grained overview of the four systems under discussion will be concluded with a short section on the potential implications of the different systems for event unit formation. There are two key factors in this regard: (i) the status of the verb and (ii) the nature of the grammaticalized forms, given the salience accorded to specific conceptual categories in comparison to others. The verb reflects both the layer selected as well as the basic type of situation expressed in a sentence, which in turn restricts the potential modifications given with arguments and adjuncts. As outlined above, the selection of a path verb such as *to cross* involves conceptualization of a motion event with an entity moving from one side of a ground object to the other. In contrast, a verb such as *to walk* restricts the event to the manner 'walking' and is no longer the same type of event if the figure starts 'running'. *Manner* of motion does not restrict an event to a specific path segment. This means that different patterns of event unit formation can be expected across the different languages, depending on the layer selected and the semantics of the preferred verb type. With regard to spatial concepts, which can be derived from features of the entity or the ground, different criteria are thus relevant in determining break points when forming event units. If preference is given to figure-related concepts such as *orientation*, *direction*, *intention,* a break point is reached when the position of the figure changes along one of these dimensions. However, if spatial categories are derived from the ground, break points will be based on features of the ground such as a boundary or some landmark. The second key factor in event unit formation concerns the implications of grammaticalized aspectual categories.[9] Aspectual marking restricts the content of a reportable event to a specific phase of a situation. A verb in the progressive form can imply a break point which an unmarked verb will leave underspecified. This could lead to different criteria in the formation of motion event units. In non-aspect languages a break point has to be provided by reference to a change in quality in the spatial domain. In aspect languages, by contrast, temporal break points can be implemented in order to fulfill this function. Given these factors, we expect differences between speakers of the languages in this study with respect to the overall number and type of event units expressed, as well as the path segments which the event units comprise.

9 There are other grammatical features which are relevant in the context of motion events such as case marking for location versus direction, or different grammatical features of particles vs adverbs/prepositional phrases which are discussed in Carroll (2000).

4 The empirical study

Speakers were asked to describe short real-world scenes presented in video clips. Two types of stimuli were developed (see Figure 1) in which (i) the critical stimuli show a figure moving continuously along a path in the course of which the figure changes orientation/direction while (ii) the control stimuli show motion events in which a figure moves along a continuous path with no change in direction or orientation. The control stimuli served the purpose of detecting whether there is a general tendency in a given language to provide more information relative to another language, or whether potential differences arise as a result of the complexity of the motion events depicted (changes in direction/orientation, boundary crossings). A third set of video clips served as fillers and showed situations based on "causative actions" (e.g., dropping a purse and then picking it up off the ground). They serve the purpose of disguising the manipulation present in the first two categories as well as the avoidance of unwanted priming effects.

Given the typological contrasts described above, we expect differences with respect to the breakpoints selected in the formation of event units in the critical condition. Factors underlying the formation of event units will have implications for the selection of components when representing a given scene. No significant difference is expected for the control stimuli.

4.1 Method

4.1.1 Participants

Twenty subjects per language group participated in the experiment. Speakers of French were recruited at the Université Paris VIII, English speakers at the University of York,[10] and German speakers at Heidelberg University. Tunisian speakers were recruited in Heidelberg (recently arrived university students, without previous knowledge of German).[11] Participants were matched for social background, with ages ranging between 18 and 40. They received 5€ in compensation. The experiment was conducted by a native speaker of the respective language.

10 We are grateful to Annie Claude Demagny and Leah Roberts for setting up and running the experiment with English-speaking subjects.

11 Tunisian speakers acquire French as a foreign language in school. No speaker was an early bilingual in the two languages.

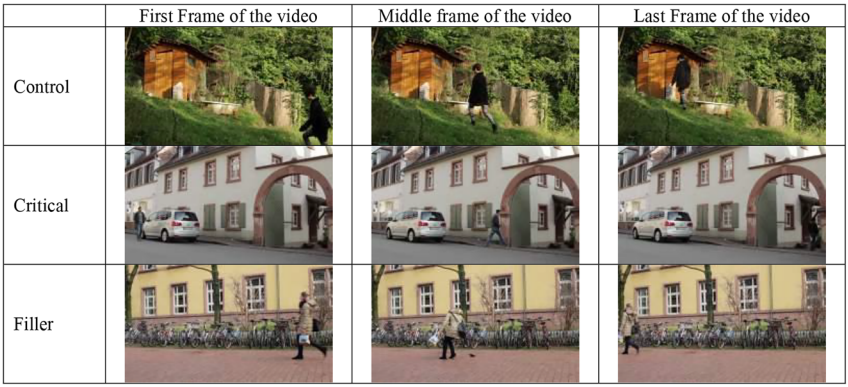


Figure 1: Sample stimuli for the control, critical and filler stimuli.

4.1.2 Stimuli

The critical stimuli (N = 12) consist of video clips showing a figure in motion (duration 5–11 s). The clips were designed as follows: eight clips (a) show a change in direction along two dimensions (either turning right/left, or moving up/down); four clips (b) show a boundary crossing in the first phase, followed by a long trajectory over either the horizontal or vertical axis (see list of stimuli in Appendix A). The control stimuli (N = 8) consist of video clips in which a figure moves continuously along a path with no change in direction or boundary crossing. The remaining videos with causative actions served as filler items (N = 10).

4.1.3 Task

Participants were instructed to describe *what happened?* in the scenes presented by using full sentences (see Appendix B for the instructions in the four languages). They were also explicitly told to focus on the dynamic situation and not on the scenery depicted in the videos.

4.1.4 Procedure

Participants were tested individually at Heidelberg University (Tunisian and German), at the University of York (English), and the University Paris VIII (French). The procedure was the same in all cases. The subjects were seated in front of a computer screen on which the instructions were displayed. The experimenter then asked if they had any questions. This was followed by a practice run in which participants saw three videos, each showing one of the three stimulus categories.

They were again asked if they had questions. The actual experiment was then started and lasted approximately 8 min. Verbal responses were recorded with a mobile phone.

4.2 Data analyses

Verbalizations were audio-recorded and then transcribed by a native speaker of each language. The transcribed data, including all codings reported in the following sections has been published separately (von Stutterheim et al. 2019).

4.2.1 Coding the number of units

The number of reported events was coded for each critical item. Coding was double checked per language group by a second native speaker. Based on the definition given above under 2.1., a *reported event* is defined as a linguistic unit with one finite verb. In French, for example, *un ballon roule et descend les escaliers* 'a ball rolls and descends the staircase' was coded as two units, whereas *une femme sort en courant* 'a woman leaves at running' was coded as one unit. In German, e*in Ball rollt die Treppe runter* 'a ball rolls the staircase down' was coded as one unit. Utterances in English and Tunisian were coded accordingly. Utterances which did not refer to the motion event depicted, or which involved a personal opinion (e.g., *the person seems to be in a hurry*) were discarded. In the critical condition, subjects produced between one and three assertions per item.

4.2.2 Coding event units: Temporal criteria

Two specific cases require separate explanations. (a) One concerns reference to manner of motion without a manner verb; (b) the other concerns the question as to whether a coordination is interpreted as a constituent coordination or an elliptical sentence.

Reference to manner of motion without a manner verb: These are cases where languages do not have a manner verb for a particular type of situation, but express manner by means of a very abstract verb such as *to be* or *to make* along with a nominal expression. This is the case in French with riding a bicycle, which is expressed as *faire du vélo* or *être à vélo*. These cases were counted as reference to manner of motion.

Coordination: The question of ellipsis is not easy to decide. In this case we followed the line of argumentation in Bohnemeyer et al. (2007: 502), in which the authors investigated the impact of typological differences on motion event segmentation. They show that lexicalization patterns and the availability of

constructions (e.g., serial verb constructions) lead to differences in the way speakers segment motion events into units. The notion of a *macro event* was introduced in this context and defined as a unit with one temporal operator. We adopt this definition in coding the crosslinguistic data in our study. Given this definition the following Example (4) is viewed as one event unit:

(4) *Peter went out of the house into the shop.*

(5) *? Peter went out of the house at seven into the shop at seven ten.*

If a temporal adverbial is inserted, in particular in initial position, it has scope over all spatial adjuncts (6).

(6) *In the morning Peter went out of the house into the shop.*

This stands in contrast to a coordinated construction such as:

(7) *Peter went out of the house at seven and into the shop at seven ten.*

In (7) the two conjuncts can be temporally located at different times. Coordinations of this type are treated in the present analysis as elliptic sentences (cf. the discussion of verb ellipsis in Klein 1993). This means that structures as in (4) are coded as belonging to one sentence and correspond to one event unit which refers to two segments of the path traced by the figure. If speakers express two path segments as coordinated by a conjunction as in example (7) the structure is analyzed as consisting of two event units. The analyses of the other languages are based on the same criteria.

4.2.3 Coding the conceptual categories expressed

The way in which verbalizations produced by the subjects were coded aimed at identifying possible differences with respect to both the overall number and type of event units expressed as well as the path segments which event units encompass. Categories were thus coded at different levels as follows: (a) the components selected in representing a motion event, i.e., the type of unit formed, and (b) the level of verbal packaging. For coding at level (a), the maximum number of segments was determined per video clip. The total number of segments which speakers could describe amount to 40 for the 12 critical items. The possible numbers vary across the critical items depending on the nature of the path taken by the moving figure (see Appendix C). Coding for each scene includes both the type of segment as well as the number mentioned by each speaker. If a speaker did not refer to a segment by providing spatial information but only described manner of motion, this was coded as *segment 0* (as, for example, in the French utterance *une*

femme fait du vélo, 'a woman rides a bike'). Coding at the level of verbal packaging (b) relates to specific spatial and temporal expressions. In the spatial domain the categories expressed by verbs were coded for Path, Manner, Neutral, while prepositional phrases, particles and adverbials were coded for Location, Path, Direction, Boundary crossing. Furthermore, spatial concepts were differentiated as to whether they are based on the figure or ground, or both. In the temporal domain, aspect marking on the verb (ongoing, progressive, perfective) was coded for the English and Tunisian data.

4.3 Results

4.3.1 Segmentation and unit formation

The first step in the analysis concerned the number of reported event units (defined by the occurrence of a finite verb expressing motion) and the number of path segments expressed (encoded either in a verb or adjunct). Table 2 and Figure 2 both show the quantitative differences in the number of reported events across the languages for the critical and the control items.

To assess the numerical differences in the event units expressed between languages in the critical and the control condition, mixed-effects logistic regression modeling was applied (R version 3.5.0, lme4 version 1.1.17). The number of utterances per item and subject (i.e., per trial) was set as the dependent variable. The fixed factors of interest were condition and language group (both deviation coded).

The first step in the analysis set out to evaluate the main effects of condition and language. This proceeded on the basis of the simplest model, i.e., a model that only accounted for subject and item variability (specifying random intercepts for

Table 2: Expressed event units.

	English		German		French		Tunisian	
	Total	**Mean by subject (SD)**	**Total**	**Mean by subject (SD)**	**Total**	**Mean by subject (SD)**	**Total**	**Mean by subject (SD)**
Control (N = 8)	168	1.08 (0.10)	151	1.02 (0.06)	155	1.05 (0.09)	166	1.19 (0.13)
Critical (N = 12)	463	1.98 (0.31)	302	1.31 (0.28)	401	1.70 (0.29)	425	1.80 (0.33)

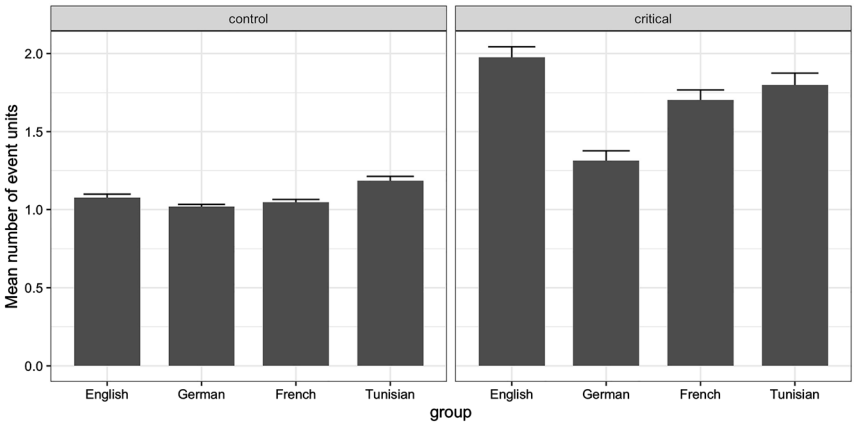


Figure 2: Expressed event units in the critical and control condition over language group.

subjects and items only), comparing this with three further models that additionally included a) condition as the only fixed factor (plus random slopes in addition to random intercepts for subjects), b) language group as the only fixed effect, and c) condition and language group as two fixed effects, and their interaction. Model c) was considered as the final (full) model. As for the random effects structure, this final model specified random intercepts and slopes for subject and random intercepts for items. Note that items were nested under condition. Model comparison revealed that model a) accounted for the variance in the data significantly better than the simplest model ($p < 0.001$), which shows that condition asserted a significant effect. Model b) did not capture the variance in the data significantly better than the simplest model, so language group by itself did not contribute significantly to the variance in the data. However, model c) was found to capture the variance in the data in significantly better terms relative to the simplest model and model a) ($p < 0.001$) thus suggesting an interaction between the two fixed effects.

The second step in the analysis focused on whether the different language groups deviated from the grand means within conditions. For this analysis we retrieved the simple effects from the full model (DV ~ condition * group – group …). The relevant contrasts are displayed in Table 3. Within the control condition, speakers of Tunisian were the only group to significantly deviate from the grand mean in that they produce more utterances than the other speakers did on average irrespective of their language. In the critical condition, English speakers produce significantly more utterances than the average for all other speakers irrespective of their language, while German speakers produced significantly fewer utterances. French and Tunisian speakers do not deviate from the mean in any direction.

Table 3: Model output: Comparing language groups with grand mean within conditions with respect to the number of utterances produced.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.389	0.050	27.669	<2e-16	***
condition1	−0.308	0.048	−6.359	2.99E-06	***
control:English	−0.005	0.037	−0.132	0.8949	
critical:English	0.275	0.058	4.713	1.07E-05	***
control:German	−0.062	0.037	−1.667	0.0962	.
critical:German	−0.383	0.059	−6.55	5.87E-09	***
control:French	−0.041	0.037	−1.095	0.2742	
critical:French	0.006	0.058	0.096	0.9235	
control:Tunisian	0.108	0.038	2.835	0.0048	**
critical:Tunisian	0.102	0.058	1.75	0.0841	.

The finding that Tunisian speakers produced more utterances in the control condition than speakers of other languages is unexpected.[12] However, the fact that this effect is absent in the critical condition suggests that there may be no general bias to produce more detailed verbalizations.

4.3.2 Number of event segments expressed

When the analysis is at the level of the utterance, and manner verbs are used as main predicates, more than one path segment can be expressed by a chain of adjuncts in a single utterance (*A man walks down a road around a corner into a courtyard*.) This factor was taken into account in the next step in the analysis on the number of path segments referred to by each group (the maximum number per

12 As speakers of a language in which aspectual distinctions are fully grammaticalized, some situations are broken down into phases in Tunisian (Example i) in contrast to French (Example ii) where 19/20 describe the situation on the basis of one utterance.

(i) *ṭofla* *tāl'a* *f-eğ-ğbel*
girl monterPART.3SGF in-DEF-mountain
māšya *l-bit*
goPART.3SGF to-house

(ii) *une jeune fille* *monte une colline*
a young girl climbs a hill

The control items reveal language-specific patterns which are not anchored in the spatial but in the temporal domain (see von Stutterheim et al. 2017).

Table 4: Overall number of segments expressed, critical items (the percent value relates to the overall number of possible segments: 40 (possible path segments) × 20 speakers = 800).

	ENG	FR	TUN	GER
Segments	503 (62.9%)	401 (50.1%)	425 (53.1%)	335 (41.9%)

scene was set by several independent raters, see above and Appendix C). Table 4 shows the results.

Mixed-effects logistic regression modeling was again used to statistically assess the numerical differences. The dependent variable was the number of path segments expressed by each subject per item. Again as an initial step, the simplest model was created, i.e., a model that only specified random intercepts for subjects and items. This model was then compared with the full model, that is, a model which additionally included "language group" (deviation coded) as a fixed factor. Model comparison revealed that the full model captured the variance in the data significantly better than the simplest model ($p < 0.001$). This means that 'language group' exerted a significant influence.

In the next step we evaluated whether any language group significantly deviates from the mean. The results are displayed in Table 5. German participants provided significantly less information on path segments than all other languages while English participants provided significantly more information on path segments compared to all other languages. Tunisian and French do not stand out in any direction.

The quantitative results show significant crosslinguistic differences with respect to the amount of information provided (Table 4) as well as the formation of event units (Table 2), contrary to what would be expected based on previous typological findings. However these studies did not take into account the role of temporal/aspectual distinctions. German and English are located at two opposite

Table 5: Model output: Comparing language groups with grand mean with respect to number of segments expressed.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.778	0.089	20.043	0.000	***
German	−0.322	0.062	−5.153	0.000	***
English	0.364	0.062	5.856	0.000	***
French	−0.070	0.062	−1.117	0.267	
Tunisian	0.027	0.062	0.429	0.669	

poles on a scale of explicitness. Speakers of English produce the highest number of utterances and verbalize the highest number of segments, in contrast to speakers of German with the lowest number of utterances and the lowest number of segments. The French and Tunisian speakers show a similar pattern for both the number of utterances as well as segments.

4.3.3 Types of expressed event segments

The next step in the analysis concerns a closer look at the differences in the verb forms and adjuncts used which lead to contrasts in the content expressed behind the numbers. This applies to the segment(s) selected in representing the scenes as well as how speakers proceed, given the following options (type a and type b):

type a = change of state, identification of a break point or an endpoint, in other words, event$_2$ units

As indicated in the clips, this type encompasses boundary crossings at the source and the goal of a path (e.g., *to walk out of a building/ to leave a building; to walk into a building/ to enter a building*); a change in direction (*to turn into*); passing by an object (*to pass by a car*) in the function of a landmark.

type b = trajectory between breakpoints

A trajectory between break points can be referred to by locating the figure on a path, type (b1) (*une femme marche dans la rue, a woman walks in the street*) or (b2) with information on manner only (*une femme court, a woman runs*) or (b3) by providing information on the trajectory itself (*a woman walks along a road*).

Table 6 presents an overview of the segments that speakers of the four different languages provided in their verbalizations. In order to assess whether the relative differences are statistically significant, another set of logistic regression analyses was carried out. As the dependent variable we calculated empirical logits based on the number of segment type encodings (a, b1, b2, b3) for every response given and the total number of encoded segments per response for each subject.[13] Note that this measure captures the fact that the specific frequency with which a segment type was encoded must be viewed as a relative value (relative to the total number of segment encodings, which differs significantly across languages, as shown above), but at the same time is not bounded

13 Empirical logit: log((number of segment type per response + 0.5)/(total number of expressed segments - number of segment type per response + 0.5))

Table 6: Segments selected for expression in the four groups of speakers.

	Type a	Type b1	Type b2 (zero adjunct)	Type b3	Total
English	79.9% (402)	1.6% (8)	5.6% (28)	13.1% (66)	503
French	61.8% (248)	16.7% (67)	20.9% (84)	0.5% (2)	401
Tunisian	72,5% (308)	3.3% (14)	23.5% (100)	0.7% (3)	425
German	85.1% (285)	3.6% (12)	3.0% (10)	8.1% (27)	335[15]
Grand mean	75.9%	6.2%	13.3%	4.6%	

by 0 and 1. Because the variance of the logit depends on the mean, weights to include in the model were also calculated.[14]

As a first step we evaluated the differences for "type a" segments. Language group was used as a fixed factor (deviation-coded) and random intercepts for subjects and items were added. We determined whether the inclusion of "language group" as a fixed factor explained the variance in the data significantly better relative to the simplest model without it (see above). Model comparison revealed that it did ($p < 0.001$). This means that there was a significant effect of "language group".

Table 7 shows whether or not the number of "type a"-segment encodings differs significantly from the grand mean. As this shows, German as well as English speakers encoded "type a"-segments more frequently in significant terms compared to all speakers on average, irrespective of their language. French speakers deviated significantly from the mean in that they produced significantly fewer "type a"-segments. No significant effect was observed for Tunisian speakers.

Because the numbers for "type b1" segment encodings are relatively low, it was decided not to perform a statistical test.

For the analysis of "type b2"-segment encodings, the same approach was followed as with "type a"-segments (see above). "Language group" again contributed significantly to the variance in the data ($p < 0.001$). Table 8 shows the output of the model. Both speakers of English and German deviated significantly from the mean in that they produced fewer references to "type b2"-segments than speakers irrespective of their language. Both French and Tunisian speakers showed a significantly higher number of references to "type b2" segments.

14 Weights: = 1/(number of segment type per response + 0.5) + 1/(total number of expressed segments – number of segment type per response + 0.5).

15 The difference regarding the number of segments in Table 4 for English and German is due to the fact that some utterances contain references to two segments.

Table 7: Model output for a language specific deviation from the mean with respect to Type A segment encodings.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.623	0.083	7.546	0.000	***
German	0.244	0.048	5.102	0.000	***
English	0.188	0.045	4.220	0.000	***
French	−0.355	0.045	−7.941	0.000	***
Tunisian	−0.077	0.045	−1.699	0.094	.

Table 8: Differences between language groups with respect to Type B2 segments.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	−0.987	0.074	−13.351	0.000	***
German	−0.248	0.059	−4.181	0.000	***
English	−0.360	0.058	−6.208	0.000	***
French	0.274	0.057	4.768	0.000	***
Tunisian	0.334	0.057	5.863	0.000	***

These results point to similarities as well as differences across the four language groups. For all speakers, the highest numbers of utterances apply to scenes in which breakpoints are identified and *motion event$_2$ units* (change of state) are formed. Since the participants were asked to describe *what is happening?* and the clips present situations showing locomotion, use of this event type can be expected across all language groups.

In the situations where a motion event$_2$ is not construed, type b1 and type b2 refer to patterns in which speakers do not refer to directed motion, as in *une femme marche (dans la rue)* 'a woman walks (in the street)'. This type of conceptualization with specification of a location is frequent in French and in Tunisian, but this is significantly less frequent in German and English.

With regard to the number of segments, speakers of German typically select one segment in event representation which is based on an event$_2$ type. As illustrated above, the manner verbs used in German are used in the context of an event type where the phase before a break point is extended: *eine Frau geht ins Haus* 'a woman walks into a house'.[16]

16 Cf von Stutterheim et al. (2012) on the mention of endpoints in different languages.

It is noteworthy that in cases where German speakers do not identify a suitable delimiting breakpoint in the scene itself, they resort to the level of the observer and use the screen as the referential frame in providing a spatial boundary e*in Mann geht aus dem Bild* (a man walks off the picture).

French and Tunisian speakers express a higher number of segments relative to speakers of German (cf. Table 4) and form different conceptual representations for the different components of the situations depicted (cf. Table 6). Changes in place with no specifiable direction do not allow use of a path verb to encode directed motion in these languages. Speakers of French and Tunisian profile this part of the situation in terms of a type b1-event – which can be metaphorically described as a *screenshot* of a path. The figure is described as moving in a specific manner at a specific location. This conceptual frame does not profile directed motion. The spatial information provided given manner of motion in the verb is thus locational, as specified typologically for these languages (see type b2).

English shows a pattern which is similar to German with regard to event type b1 and b2. The languages differ, however, with respect to pattern b3 where information on the ground is provided with respect to the trajectory drawn by the figure (*a woman is walking along a street*). In contrast to German, where speakers frequently represent the scenes on the basis of one unit, this is rare in English where a higher level of detail is provided on the trajectories (cf. Table 4). As will be discussed below, these differences in segmentation can be attributed to the conceptual implications of grammatical aspect as well as the option of expressing direction via the figure (path verbs) in addition to the ground (see below). In contrast to French and Tunisian, English speakers, given a change in place, do not construct a *screenshot* viewpoint with specification of a location only, but represent the events as motion along a path.

The final section in the analysis provides an overview of the differences in information selection and event unit formation as reflected at the level of the expressive devices used.

4.3.5 Verb types

Since the verbs selected as predicates indicate the result of a conceptualization process which solves the task of representing the visual input verbally, the present section gives an overview of the verb types used across the languages.

Table 9: Verb types produced in the critical condition (20 participants, 12 items).[17]

Verb type	Group	Overall	Relative	a	b1	b2	b3
Direction	English	145	31.3%	166	0	0	3
Direction	French	248	61.8%	245	3	0	0
Direction	German	75	24.7%	88	0	0	1
Direction	Tunisian	288	67.8%	287	0	1	0
Manner	English	318	68.7%	236	8	28	63
Manner	French	153	38.2%	3	64	84	2
Manner	German	229	75.3%	197	12	10	26
Manner	Tunisian	35	8.2%	3	2	30	0
Neutral	Tunisian	102	24.0%	18	12	69	3

Furthermore, this includes the frequencies with which the different path segments are combined with the different verb types for each language.

The results for German and Tunisian reflect well established typological contrasts in his domain: Speakers of German mainly use manner verbs while Tunisian speakers use path verbs as well as the neutral verb *mša* 'to move'. French and English show both verb types, albeit with a reverse distribution (Table 9). As the table shows, French speakers use manner verbs in expressions that clearly refer to different types of segments, compared to speakers of the other languages.

4.3.5 Tense and aspect in English and Tunisian

English and Tunisian speakers produce the highest number of segments, in overall terms, compared to the other language groups, although they belong to two different typological categories in the spatial domain. In temporal terms, however, the two languages share a relevant structural feature of having fully grammaticalized aspect, thus indicating the possible role of temporal aspect in event segmentation. Depending on the nature of the aspectual oppositions encoded in a system, different phases of a situation will be under focus when conceptualizing content for expression. In English, for example, the perspective that an event can be conceptualized as "ongoing" is grammaticalized with the *-ing* form (*they are/were walking down the street*). Use of relevant distinctions is shown in Table 10 for the speakers of English and Tunisian.

17 In Tunisian there are 10 instances of zero verb utterances due to zero copula in Arabic; in the English data there are two instances of copula constructions which were not included in the counting.

Table 10: Use of verbal forms in the languages with grammaticalized aspect (20 participants, 12 critical items).

	IPF	PF	Prog (PART)	Ongoing	Future	Simple form
English	–	7	–	128	–	328
Tunisian	95	221	84	5	11	–

Although aspectual markings rarely occur in the English data given a series of subsequent events, the activation of the concept "ongoing" at the outset for almost each scene may be relevant for levels of segmentation in overall terms:

(8) *a man is walking down a street, he goes past a car and then turns into an entrance*

The fact that speakers of English are tuned to phasal decomposition of situations by their aspectual system could lead to the differences in granularity in the level of segmentation compared to the German speakers.

The distribution of aspectual forms in Tunisian conforms with findings in a previous study on Arabic languages (von Stutterheim et al. 2017). As outlined in section 3, most items were described on the basis of two utterances: the first with the imperfective form marking either a manner verb, or the neutral verb, while the second utterance is marked by either progressive or perfective aspect. The latter aspectual categories require a breakpoint and thus an $event_2$ unit (see (9)). The data show that speakers can encode $event_{2\ units}$ by means of the temporal markers for *progression* or *perfectivity*, without also having to provide spatial information.

(9) a. *rağel* *yemši*
man move**IPF**.3SGM
b. *tla'* *f-ed-drūğ*
climb**PF**.3SGM in-DEF-stairs
c. *w* *dxal*
and enter**PF**.3SGM

4.4 Discussion of the results

The results for the different levels of analyses on the formation of event units will be discussed for each language from the point of view of the production process which starts with the visual input and ends with the verbal output. The perspective adopted covers the following theoretical concepts: The term "event layers" serves

in differentiating the factors that speakers treat as privileged information when constructing representations of the scenes presented. Each layer subsumes specific "sets of qualities" which are relevant in processing the input in terms of spatial as well as temporal categories, in particular temporal aspect. In the spatial domain, a distinction is drawn between features which can be categorized as figure- or ground-based. Those which relate to the figure in motion involve the direction and orientation of the moving entity, while ground-based features include landmarks, spatial boundaries, and other specific properties of the ground traversed. The categories activated in the construction of a reportable event start with the "privileged" layer in event formation (either *path* or *manner of motion*) which determines the spatial frame of reference (figure- or ground-based), thereby leading to different "qualities" within these domains as well as potential changes. With regard to the role of aspect, the present study includes two systems in which aspectual distinctions are expressed by grammatical means (English and Tunisian), and two other languages in which grammatical aspect is not relevant (German and French); (cf. Bylund et al. 2013; von Stutterheim et al. 2012; von Stutterheim et al. 2017).

Starting with speakers of German, the implications for decisions in motion event construal which are structurally based can be described as follows. Given the default case, manner of motion is under focus in conjunction with ground-based categories in the spatial domain with no preconditions for the temporal domain. Speakers of German construe a motion event focusing on manner of motion of the moving entity, whereby changes in the orientation of the figure will attract less attention, along with potential phases of the situation in relation to the observer's view point. Since there are no grammaticalized categories which profile different phases of a given situation, the verbal forms in German correspond to a holistic viewpoint (cf. von Stutterheim et al. 2012). Given the fact that occurrences which can be described as *walk* or *run* do not entail an inherent change, speakers will therefore attend to properties of the ground which can serve as breakpoints since the temporal domain cannot serve this function. Ground-related concepts are thus the prime candidates in forming a motion event unit. This form of event conceptualization is the preferred pattern in German. This conclusion is based on the empirical findings: Speakers of German mainly use manner verbs in combination with verb particles and spatial adjuncts, with a low number of cases with zero adjuncts (cf. Table 6). The semantic feature "directed" motion is salient in event conceptualization in German given the role of verb particles (*hinein-fahren* 'tither-in-drive'). These forms are prominent given the following features: (i) verb particles support the selection of a breakpoint, thus leading to a change of state event; (ii) they combine with simple finite verbs to form event representations which correspond to a holistic view of the situation. The forms thus lead to

representations of complex motion events on the basis of a single event unit. This pattern in event conceptualization can be described as a condensed representation of the situation, whereby a trajectory is combined with a spatial breakpoint. The high number of descriptions in the German data based on one utterance reflects this strategy.

French speakers tend to access both "layers" frequently in motion event construal with a rate of 61.8% for path verbs and 38.2% manner verbs. However, the distribution of these two types is constrained (cf. Table 9). In order to express directed motion French speakers select the "path layer" to form a reportable event unit. This requires attention to the moving figure with regard to both orientation and direction: Direction can be expressed, for example, as in *une femme monte une pente* 'a woman ascends a hill' or by a (potential) breakpoint, as in *une femme entre dans une cour* 'a woman enters in a yard'. However, if there is no evidence from the input on the direction taken by the figure, speakers of French select manner of motion. As with directionality, the event unit is represented as a quality of the moving figure. Since a spatial concept concerning direction cannot be derived from the figure, the only figure-based concept which then remains is its spatial location: *une femme marche dans la rue* 'a woman walks in a street' or no spatial reference: *une jeune fille fait du vélo* 'a girl cycles'. The relatively high number of utterances, and the relatively high number of segments expressed (in comparison to German), can be attributed to the fact that the motion events presented in the visual input include changes in direction as well as longer passages without an evident source or goal. The first one requires segmentation with regard to changes in direction as in (1) *Il'ya une femme qui sort d'un bâtiment* 'there is a woman who leaves a building' (2) *qui descend vite les marches* 'who descends fast the stairs'. The second one leads to segmentation involving a description of a figure moving in a specific manner and a unit encoding a directed motion event: (1) *une femme marche* 'a woman walks' (2) *puis monte des marches d'escaliers* 'then ascends the steps of the stairs'. The numbers for event types (Table 6) reflect this pattern in that only 62% (type a segments) of the units are directional $event_2$-types, while the remaining 38% relate to 'minus directed motion' (type b1, b2 segments). The fact that there are only two 'type b3' occurrences (i.e., where information on the trajectory is ground-based (*a woman walks along a road*) relates to the status of ground information in French. If manner of motion is encoded in the verb focus is placed on *location* but not on ground-based information concerning *direction*. This is typically not accessed in this context by French speakers when forming a reportable unit.

The results for the speakers of Tunisian can be described as follows: in the conceptualization process from the visual input to verbal encoding, prominence is accorded to the layer "path". If the input allows the identification of a spatial breakpoint (starting point, goal, boundary crossing), speakers select an $event_2$ type

(see Table 6: 72.5%, type a segments). Given a longer trajectory without changes in orientation or direction (the remaining responses) one would expect specification of manner of motion. However, the form selected does not relate to a specific manner. The verb *mša*, which was originally a manner verb that can be translated as *to walk*, has bleached and is now a general motion verb. This is where aspectual categories come into play: in the case of phasal decomposition, unit formation proceeds on the basis of temporal properties leading to $event_1$-/$event_2$-types. Tunisian speakers attend to quality changes over time in relation to a temporal viewing point. Units are constructed which refer to progression from or towards a breakpoint, if a possible breakpoint is identifiable without the figure having reached one. An event unit of this type is marked by means of progressive aspect. If the event unit/temporal interval includes a change of state, a 2-state event is formed which is then marked aspectually by the perfective. If there is no evidence of a (potential) breakpoint, speakers do not access information relating to the layer "path of motion". As in this case in French, speakers access the motion/manner layer in order to construct a reportable unit. As mentioned above for French, if the information extracted does not involve directed motion but a quality of the figure which is not associated with orientation or direction (it's location), this type of unit is typically specified on the basis of an adjunct (b1). However, this form of construal is not observed in the data for Tunisian, given the high number of zero adjuncts (cf. Table 6) with utterances of the type *rağel yemši* (a man moves). This points again to the prominence of the temporal dimension as the basis when forming a specific reportable unit: the imperfective constitutes the aspectual category which is relevant in this context. Core concepts in the temporal domain support the level of segmentation observed for speakers of Tunisian.

As with speakers of German, the layer typically accessed in English concerns manner of motion. Ground-based categories are also prominent in the process of event unit formation. However, in contrast to German, speakers of English also access the path layer where path is defined on the basis of directional concepts: If features of the ground serve to structure the route taken, with no change in direction (*x rides his bicycle across a courtyard and through a gate),* or when the change in direction is supported by features of the ground (x *walks past a fountain and up some steps*), then the "manner layer" is selected in forming an event unit. When features of the ground are not well-defined, speakers have the option of selecting figure-related breakpoints to encode changes in direction or goal-directed orientation. This is expressed by means of path verbs (*to turn, to head for, to enter, to pass*). Speakers of English can thus proceed on the basis of these two perspectives in the domain of space, according attention to both figure and ground. These strategies lead to the high number of segments observed. Note that this pattern differs from the findings for French and Tunisian where directed

motion events are construed at the level of the path of motion as figure-based. A further factor in this regard concerns the temporal domain and the role of aspect with the underlying concept "event is ongoing" as marked by the *-ing* form. This leads speakers to decompose situations into phases as shown in several studies on the role of aspectual categories in event construal (cf. Athanasopoulos and Bylund 2013; Flecken et al. 2014; von Stutterheim et al. 2012). Both factors, aspectual distinctions and two potential spatial viewpoints (figure and ground) lead to a high level of resolution in event unit formation. They define what can be termed language-specific patterns in allocation of attention which are relevant in event unit formation.

5 General discussion

Mental representations of motion events are constructed on the basis of categories covering different conceptual domains: figure, space, and time. While these domains can be regarded as providing a universal "pool" of concepts which are involved in the representation of motion events, the specific relevance and function of these categories can vary significantly across languages. Previous studies have pointed to gradual differences in the saliency of certain components across languages (cf. Feist 2016; Slobin 2006). In research on typological contrasts, this claim is supported by relevant differences in the distributional patterns of specific lexical and syntactic categories. This is reflected in the analysis of the linguistic means typically used in spontaneous descriptions of motion scenes by speakers of different languages. However, previous analyses have focused on concepts related to the relative weight of *manner* versus *path*, taking both concepts as two poles on a scale (Slobin 2006). The basic claim is *that if a domain is elaborated in linguistic expression, users of this language will continually attend to and elaborate that domain cognitively (*2006: 77). With this approach, which has also been followed by other researchers (Durst-Andersen et al. 2013; Fanego 2012; Feist 2016; Pourcel 2005 [18]), the choice between the two dimensions "manner" and "path" is viewed as a preference without looking into the factors which determine the choice of one or the other dimension within one language. Given that all languages provide forms to express manner as well as path, this question remains unanswered.

18 It is symptomatic that Durst-Andersen et al. (2013) view "manner" as the salient dimension in motion event construal, with Danish as a Germanic language which privileges the manner layer. This contrasts with Ibarretxe-Antunano (2010) who focuses on "a cline of path saliency" (p. 404) given a path language as her language background.

Psycholinguistic studies support the suggested differences in saliency, coupled with differences in allocation of attention during conceptualization that can be systematically linked to the linguistic system of the participant languages (Pourcel 2005 on memory; Flecken et al. 2014; Papafragou et al. 2008; on visual attention; and Flecken et al. 2015a on categorization). However, these studies focus on categorization, patterns of attention, and memory performance based on stimulus material depicting a single motion event (one path segment only). The results thus relate to preferences based on categories which cover path versus manner.

The present study set out to address the following questions: (i) what factors drive speakers of a given language to select one concept over the other, in particular with respect to the less salient option in certain contexts, and (ii) what determines the possible position of a language on the continuum within typologically related languages. In other words, what are the principles behind "more" or "less" of path versus manner saliency?

In order to address the relevant range of questions, we adopted a process-oriented approach. It starts at the point where all components relevant for the construction of mental representations that serve linguistic encoding (spatial and temporal categories) are extracted from a visual stimulus. The study reconstructs how these components are integrated in order to form what has been termed a *reportable event unit* (Carroll and von Stutterheim 2011; von Stutterheim and Nüse 2003; von Stutterheim et al. 2017). In the view adopted here, specifications concerning information selection proceed in a top-down manner.

The integration process requires decisions with respect to what we term the "layer" selected when representing an event with its implications for potential breakpoints. This takes into account the function of both temporal and spatial conceptual categories in event conceptualization. In other words, relevant factors which have been shown to vary between languages in the domain of space could also be implied by contrasts in the domain of time. The analyses of the verbal responses show how the languages under investigation differ with respect to event unit formation, given the information encoded verbally as a response to the visual input. Differences relate (1) to the basic event type encoded in the verbs used, (2) the changes in qualities expressed, (3) the total number of path segments encoded per situation, and (4) the number of path segments packaged into one utterance. Given our starting point, the differences in verbalizations were interpreted in terms of different conceptual representations constructed on the basis of the language-specific patterns of attention that speakers apply during the construction process.[19] More specifically, these attentional patterns capture language-specific

19 Oral language production is typically based on processes which are not conscious and deliberate, but run automatically and unconsciously (cf. Evans 2008 "model of dual processing").

defaults described in relation to the theoretical concepts outlined above (cf. Section 4.4 for details on each language). They are automatically accessed in language production and function as a language specific *attentional template*. In line with previous accounts (Evans 2008; Radvansky and Zacks 2014; Thierry 2016) we assume that these attentional templates ensure the required speed in processing information with regard to segmentation and selection when forming reportable events in online event descriptions.

The analyses also showed that languages differ not only with respect to unit formation in contexts that trigger the respective default patterns: If the external input does not immediately allow for event unit formation based on defaults in patterns of attention, a strategy may be initiated which can be viewed as second in rank. These procedures in event construal are tuned to language-specific verbal means (lexical and grammatical). In other words, the selection of a secondary strategy is also language-specific. A case in point is illustrated by speakers of French in the current study: when the figure in motion is under focus, but there is no relevant information which would allow the preferred pattern in specifying *direction* (path encoded in the verb-/figure-based), speakers select the manner layer, and direction typically remains unspecified. However, if information on the path taken becomes available in the course of the stimulus, information processing for a second event unit draws on the default pattern of attention, and direction is explicitly expressed by means of a path verb. French and German speakers may both select the manner layer in event construal, but while manner of motion represents the default case in German, this functions as an option in French, which is subordinate to the default layer "path". It thus entails a crosslinguistic difference in event unit formation and represents a hierarchical order and not a function of graded salience.

The findings confirm the impact of language (lexical repertoire, grammar) and language use (frequency of occurrence of linguistic structures in relation to a specific input) on pattern building at the conceptual level which leads to the observed preferences. The focus on event unit formation in the present study underlines how event segmentation, information selection and organization ensures that the resulting representation is tuned to an efficient use of the means available in each language, given the specific temporal as well as spatial categories. In the case of locomotion, these patterns may be formed given the different functions of the conceptual domains involved. This will be discussed again below in relation to the core concepts figure, ground and temporal aspect.

The comparison of the four languages shows how speakers proceed when "exploiting" the continuous stream of perceived information in the formation of a reportable event. The *figure* in motion and its direction plays a critical role in the

pattern in French, a factor which also implies intentionality on the part of the figure (going up or down some stairs, or entering a building presupposes some guiding purpose). Since path verbs which encode direction typically relate to the *figure* in motion, references to features of the ground are not at issue in the system in determining direction, as described in previous studies. If a possible direction is unclear, French speakers select manner of motion, with relatively few references to features of the ground, *le long de* 'along', or *à côté de* 'beside', or the direction taken *vers x* 'toward x'. The temporal-aspectual domain is not constitutive in motion event construal in French, in contrast to Tunisian, the other language which privileges the path layer.

For speakers of Tunisian, the *temporal-aspectual domain* is prioritized in motion event construal, while the selection of spatial information was shown to depend on the nature of the specific temporal interval encoded in the verb. If the interval encompasses a change of direction or a boundary crossing by the moving figure, then a path verb is typically selected in encoding the motion event. If the interval does not include any change in state at this level, speakers encode the event in aspectual terms as *ongoing* with no change in state. The fact that speakers of Tunisian provide relatively little information on the specific properties of the path can be explained in part by the status of the figure in motion in a "verb-framed" language. But we also assume that temporal aspect is relevant given the concepts *perfective* and the associated concept *concluded*, while the *imperfective* can indicate 'continuing'. A description of the type *a man has entered (PF)* or *a woman is moving (IPF)* would not be formed as reportable unit in the other languages, in contrast to Tunisian. The role of the temporal-aspectual domain thus leads to different patterns in motion event construal across the two verb-framed languages.

English and German privilege the *manner* layer in event unit formation, but differ in other relevant respects. English requires obligatory aspectual marking in specific contexts and speakers use a significant number of path verbs, as shown in the data base. English speakers thus have two options given the presence of means to express both figure- and ground-related spatial concepts. The data show, however, that the different patterns are not equivalent options with respect to one specific situation. The selection of one over the other is driven by the spatial information provided in the external stimulus. If specific information on the trajectory taken can be identified via prominent features of the ground, this will be selected to provide break points when forming reportable events. In this case a manner predicate can form the basic event type, since ground information is "responsible" in forming an $event_2$ representation. Where this is not the case, speakers of English draw on directional verbs which relate to features of the figure

in motion. There is thus no clear dominance of one pattern over the other compared to the other languages.

This contrasts with German, where the *manner* layer is clearly privileged and the *spatial domain* is relevant in forming a reportable unit. Features related to the ground serve in forming bounded units. In contrast to English, temporal-aspectual features are not relevant in German. Patterns in allocation of attention which conjoin the temporal and the spatial domain, given the complex structure of the linguistic means available, lead to the observed preference in representing the complex path observed on the basis of one unit (temporal domain-no phasal resolution) with ground-related references (spatial domain).

In conclusion, the study provides insights into the factors which determine diversity both within and across language types. The three theoretical concepts on which the patterns in allocation of attention are based – the privileged event layer, categories for spatial framing, and viewpoint aspect – present a framework for further typological studies in the field of motion events that go beyond, but also build on, former analyses of patterns of lexicalization (e.g., Talmy 1985, 2000) and saliency (e.g., Slobin 1996, 2006). The findings on event unit formation cannot be explained as *a cline of salience*, based on either *path* or *manner*. The present study points to the relevance of the conceptual domains of time as well as space in motion event construal. The processing-oriented perspective focuses on the role of the speaker who is required to represent an external situation in terms of the formats provided by the available linguistic units and their implications. It reveals the relative weight of the conceptual domains in event unit formation in what can be modeled as language-specific event schemata – *pre-stored attentional templates* – in motion event construal.

Appendix A Instructions for the four language groups:

Instructions: French

Vous allez voir maintenant quelques vidéos représentant des petites scènes. Votre tâche consiste à dire ce qui se passe dans chaque vidéo. Employez seulement des phrases complètes et concentrez-vous sur ce qui est important. Appuyez sur la barre ‘espace’ à la fin de votre enregistrement sonore pour passer à la vidéo suivante.

Instructions: German

Sie sehen jetzt eine Reihe von kurzen Video-Clips. Bitte beschreiben Sie bei jedem Video was passiert. Benutzen Sie bitte nur vollständige Sätze und konzentrieren Sie sich auf das Wesentliche. Drücken Sie die Leertaste, wenn Sie mit Ihrer Beschreibung fertig sind und Sie mit der Beschreibung des nächsten Videos fortfahren möchten.

Instructions: English

You will see a set of video clips, 40 in all, showing everyday events which are not connected with one another. Your task is to tell *what is happening?* It is not necessary to describe the scene in detail (e.g., *sky is blue*). Please use full sentences. You can start as soon as you recognize what is going on. Press the space bar, after you have finished your description. The next video will start automatically.

Instructions: Tunisian (Note that Tunisian is basically an oral variety. We provide a written version of the orally given instruction.)

beš tšūf des videos qsār. qolli šnuwwa qa‘ed ysīr fi kol video, rakkebli ğomal kamlīn w rakkez ken ‘ala šnuwwa qa‘ed ysīr. kol ma tkammel inzel ‘ala “espace” beš yeḍhar il video illi ba‘du.

Appendix B Description of the video stimuli: Control items

1. A woman is passing by a fountain in a park.
2. A boat is slowly going up a river.
3. A young man is passing by a fountain while dribbling with a tennis ball.
4. A girl is walking up a hill approaching a cabin.
5. A man is walking down some stairs outdoors, approaching a wooden gate.
6. A woman is walking with a woven basket along a path way.
7. A person on a scooter is slowly driving down a street.
8. A person with crutches is walking up some steps.

Critical items

1. A woman walks past a fountain up some stairs
2. A young woman rushes down some stairs, and runs down the path
3. A tennis ball comes rolling towards some stairs and rolls down the steps
4. A woman pushes a stroller towards a ramp, turns right and pushes it down the ramp
5. A small ball bounces down some stairs and then rolls over to the right
6. A woman on a bike cycles down a cobbled road and goes around a corner towards an open gateway
7. A man passes by a parking lot, turns left and approaches the entrance of an old building
8. A man passes by a parked car, turns left and passes through a gateway
9. A man is walking on a street, turns left and approaches the entrance of a building
10. A man is walking down a street, turns right and walks up some stairs by taking two steps at once
11. An old man on a bike is slowly approaching a lamp post in front of a building and turning.
12. A man on a bike is changing direction and approaching the gateway of a courtyard.

Appendix C Number of segments per item - critical items

Item	Item (Excel tables)	Segment 1	Segment 2	Segment 3	Segment 4
1	2	Past	Up	Away	
2	2n	Out	Down	Change direction	Away
3	3	Along	Down stairs	Away	
4	6	Out	Change direction	Down	
5	7n	Down	Away	Change direction	
6	15n	Along	Change direction	into	
7	17n	Past/along	Change direction	into	
8	20n	Past/along	Change direction	Through	
9	21n	Along	Change direction	Along	into
10	25n	Along	Change direction	Up	into/through
11	14	into	Towards	Change direction	Away
12	16	Across	Towards	Through	

References

Athanasopoulos, Panos & Emanuel Bylund. 2013. Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science* 37(2). 286–309.

Beavers, John, Beth Levin & Shiao Wei Tham. 2009. The typology of motion expressions revisited. *Journal of Linguistics* 46(2). 331–377.

Bennett, Jonathan. 2002. What events are. In Richard M. Gale (ed.), *The Blackwell guide to metaphysics*, 43–65. Oxford: Blackwell.

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*, 93–126. Berlin & New York: Mouton de Gruyter.

Berthele, Raphael. 2013. Disentangling manner and path: Evidence from varieties of German and Romance. In Juliana Goschler & Anatol Stefanowitsch (eds.), *Variation and change in the encoding of motion events,* 55–75. Amsterdam & Philadelphia: John Benjamins.

Bohnemeyer, Jürgen, Nicholas J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke & Felix K. Ameka. 2007. Principles of event segmentation in language: The case of motion events. *Language* 83(3). 495–532.

Bylund, Emanuel, Panos Athanasopoupos & Marcelyn Oostendorp. 2013. Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics* 51(5). 929–955.

Cardini, Fillipo-Enrico. 2008. Manner of motion saliency: An inquiry into Italian. *Cognitive Linguistics* 19(4). 533–570.

Cardini, Filippo-Enrico. 2012. Grammatical constraints and verb-framed languages: The case of Italian. *Language and Cognition* 4(3).167–201.

Carroll, Mary. 2000. Representing path in language production in English and German. In Christopher Habel & Christiane von Stutterheim (eds.), *Räumliche Konzepte und sprachliche Strukturen*, 97–118. Tübingen: Niemeyer.

Carroll, Mary & Christiane von Stutterheim. 2011. Event representation, time event relations, and clause structure: A cross linguistic study of English and German. In Eric Pederson & Jürgen Bohnemeyer (eds.), *Event representation*, 68–83. Cambridge & London: Cambridge University Press.

Carroll, Mary, Katja Weimar, Monique Flecken, Monique Lambert & Christiane von Stutterheim. 2012. Tracing trajectories: Motion event construal by advanced L2 French-English and L2 French-German speakers. *Language, Interaction and Acquisition* 3(2). 202–230.

Croft, William, Jóhanna Barðdal, Willem Hollmann, Violeta Sotirova & Chiaki Taoka. 2010. Revising Talmy's typological classification of complex event constructions. In Hans C. Boas (ed.), *Contrastive studies in Construction Grammar*, 201–234. Amsterdam & Philadelphia: John Benjamins.

Croft, William & D. Allan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Davidson, Donald. 1980. *Essays on actions and events*. Oxford: Clarendon Press.

Durst-Anderson, Peer, Viktor Smith & Ole Nedergaard Thomson. 2013. Towards a cognitive-semiotic typology of motion verbs. In Carita Paradis, Jean Hudson & Ulf Magnusson (eds.), *The construal of spatial meaning*, 118–143. Oxford: Oxford University Press.

Evans, Jonathan. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 59. 255–278.

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.
Feist, Michele I. 2016. Minding your manners: Linguistic relativity in motion. *Linguagem em (Dis) curso.* SC 16(3). 591–602.
Flecken, Monique, Panos Athanasopoulos, Jan Rouke Kuipers & Guillaume Thierry. 2015a. On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition* 141. 41–51.
Flecken, Monique, Mary Carroll, Katja Weimar & Christiane von Stutterheim. 2015b. Driving along the road or heading for the village? Conceptual differences underlying motion event perception and description in French, German and French-German L2 users. *Modern Language Journal* 99(S1). 100–122.
Flecken, Monique, Christiane von Stutterheim & Mary Carroll. 2014. Grammatical aspect influences motion event perception: Evidence from a cross-linguistic, non-verbal recognition task. *Language and Cognition* 6(1). 45–78.
Gerwien, Johannes & Christiane von Stutterheim. 2018. Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition* 180. 225–237.
Goschler, Juliana & Anatol Stefanowitsch (eds.). 2013. *Variation and change in the encoding of motion events*. Amsterdam & Philadelphia: John Benjamins.
Ibarretxe-Antunano, Iraide. 2010. Path salience in motion events. In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Kei Nakamura & y Seyda Özçalişkan (eds.), *Crosslinguistic approaches to the psychology of language*, 403–414. New York: Psychology Press.
Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.
Kamp, Hans. 1979. Events, instants and temporal reference. In Rainer Bäuerle, Urs Egli & Arnim von Stechow (eds.), *Semantics from different points of view*, 376–417. Berlin: Springer Verlag.
Klein, Wolfgang. 1993. Ellipse. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung/ An International Handbook of Contemporary Researc*h, 1. Halbband/. vol. 1, 763–799. Berlin & New York: Walter de Gruyter.
Klein, Wolfgang. 1995. A time-relational analysis of Russian aspect. *Language* 71(4). 669–695.
Klein, Wolfgang. 2010. On times and arguments. *Linguistics* 48(6). 1221–1253.
Koenig, Jean-Pierre. 2016. Event semantics. In Nick Riemer (ed.), *The Routledge handbook of semantics*, 387–402. London: Routledge.
Kopecka, Anetta. 2009. L'expression du déplacement en français: L'interaction des facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial. *Langages* 173. 54–77.
Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.
Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization*. Cambridge: Cambridge University Press.
Louhichi, Imed. 2015. *The "motionisation" of verbs: A contrastive study of thinking-for-speaking in English and Tunisian Arabic.* Brighton: University of Sussex dissertation.
Lupyan, Garry. 2015. The centrality of language in human cognition. *Language Learning* 66(3). 516–553. (accessed 12 November 2018).

Majid, Asifa, Melissa Bowerman, Sotaro Kita, B. M. Haun Daniel & Stephen C. Levinson. 2004. Can language restructure cognition? The case for space. *Trends in Cognitive Sciences* 8(3). 108–114.

Mano, Miho, Yuko Yoshinari & Kiyoko Eguchi. 2019. Interlingual versus intralingual tendencies in second language acquisition: Expressing motion events in English, Hungarian, and Japanese. In Ragnar Arntzen, Gisela Håkansson, Arnstein Hjelde & Jörg-U. Keßler (eds.), *Teachability and learnability across Languages*, 183–204. Amsterdam & Philadelphia: John Benjamins.

Matsumoto, Yo. 2003. Typologies of lexicalization patterns and event integration: Clarifications and reformulations. In Shuji Chiba (ed.), *Empirical and theoretical investigations into language: A Festschrift for Masaru Kajita*, 403–418. Tokyo: Kaitakusha.

Naigles, Letitia R., Ann. R. Eisenberg, Edward T. Kako, Melissa Highter & Nancy McGraw. 1998. Speaking of motion: Verb use in English and Spanish. *Language & Cognitive Processes* 13(5). 521–549.

Newtson, Darren & Gretchen Engquist. 1976. The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology* 12(5). 436–450.

Papafragou, Anna, Justin Conor Hulbert & John Trueswell. 2008. Does language guide event perception? Evidence from eye movements. *Cognition* 108(1). 155–184.

Pourcel, Stephanie. 2005. What makes path of motion salient?. In *Annual Meeting of the Berkeley Linguistic Society (BLS 30)*. 505–516.

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

Radvansky, Gabriel A. & Jeffrey M. Zacks. 2014. *Event cognition*. Oxford: Oxford University Press.

Schmiedtová, Barbara & Natascha Sahonenko. 2008. Die Rolle des grammatischen Aspekts in der Ereignis-Enkodierung: Ein Vergleich zwischen tschechischen und russischen Lernern des Deutschen. In Patrick Gommes & Maik Walter (eds.), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbforschung*, 45–71. Tübingen: Niemeyer.

Slobin, Dan I. 1996. Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibatani & Sandra Thompson (eds.), *Grammatical constructions: Their form and meaning*, 195–220. Oxford: Clarendon Press.

Slobin, Dan I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömqvist & Luso Verhoeven (eds.), *Relating events in narrative, vol. 2: Typological and contextual perspectives*, 219–257. Mahwah, NJ: Lawrence Erlbaum.

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*, 59–81. Amsterdam & Philadelphia: John Benjamins.

Talmy, Leonard. 1975. Semantics and syntax of motion. In John Kimball (ed.), *Syntax and semantics* 4, 181–238. New York: Academic Press.

Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description: vol. 3*: Grammatical categories and the lexicon, 57–149. Cambridge & London: Cambridge University Press.

Talmy, Leonard. 2000. *Toward a cognitive semantics: Typology and process in concept structuring*, vol. 2. Cambridge, MA: MIT Press.

Thierry, Guillaume. 2016. Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning* 66(3). 690–713.

Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2). 143–160.

van der Zee, Emile (ed.). 2003–2013. *Language and Space*. 10 volumes. Oxford: Oxford University Press.

von Stutterheim, Christiane, Martin Andermann, Mary Carroll, Monique Flecken & Barbara Schmiedtová. 2012. How grammaticized concepts shape event conceptualization in the early phases of language production. Insights from linguistic analysis, eye tracking data and memory performance. *Linguistics* 50(4). 833–869.

von Stutterheim, Christiane, Abbassia Bouhaous & Mary Carroll. 2017. From time to space: The impact of aspectual categories on the construal of motion events: The case of Tunisian Arabic and modern standard Arabic. *Linguistics* 55(1). 207–249.

von Stutterheim, Christiane, Johannes Gerwien, Abbassia Bouhaous, Mary Carroll, Monique Lambert. 2019. What makes up a reportable event in a language? Motion events as an important test domain in linguistic typology [Dataset]. heiDATA, V1, https://doi.org/10.11588/data/CLNRAU.

von Stutterheim, Christiane & Ralf Nüse. 2003. Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics* 41(5). 851–881.

Zacks, Jeffrey M. & Barbara Tversky 2001. Event structure in perception and conception. *Psychological Bulletin* 127(1). 3–21.