Grammatical aspect influences motion event perception: findings from a cross- linguistic non-verbal recognition task

MONIQUE FLECKEN, CHRISTIANE VON STUTTERHEIM and MARY CARROLL

Language and Cognition / Volume 6 / Issue 01 / March 2014, pp 45 - 78
DOI: 10.1017/langcog.2013.2, Published online: 29 January 2014

Link to this article: http://journals.cambridge.org/abstract_S1866980813000021

How to cite this article:
MONIQUE FLECKEN, CHRISTIANE VON STUTTERHEIM and MARY CARROLL (2014). Grammatical aspect influences motion event perception: findings from a cross- linguistic non-verbal recognition task . Language and Cognition, 6, pp 45-78 doi:10.1017/langcog.2013.2

Language and Cognition **6** (2014), 45–78. doi:10.1017/langcog.2013.2

Grammatical aspect influences motion event perception: findings from a cross- linguistic non-verbal recognition task*

MONIQUE FLECKEN

Radboud University Nijmegen, Donders Institute for Brain, Cognition, and Behaviour, the Netherlands

CHRISTIANE VON STUTTERHEIM

AND

MARY CARROLL

Heidelberg University, Germany

(*Received* 26 *July* 2012 – *Revised* 18 *February* 2013 – *Accepted* 23 *March* 2013 – *First published online* 29 *January* 2014)

ABSTRACT

Using eye-tracking as a window on cognitive processing, this study investigates language effects on attention to motion events in a non-verbal task. We compare gaze allocation patterns by native speakers of German and Modern Standard Arabic (MSA), two languages that differ with regard to the grammaticalization of temporal concepts. Findings of the non-verbal task, in which speakers watch dynamic event scenes while performing an auditory distracter task, are compared to gaze allocation patterns which were obtained in an event description task, using the same stimuli. We investigate whether differences in the grammatical aspectual systems of German and MSA affect the extent to which endpoints of motion events are linguistically encoded and visually processed in the two tasks. In the linguistic task, we find clear language differences in endpoint encoding and in the eye-tracking data (attention to event endpoints) as well: German speakers attend to and linguistically encode endpoints more frequently than speakers of MSA. The fixation data in the non-verbal task show similar language effects, providing

[*] We would like to thank the DFG (German Research Foundation) and the NWO Veni scheme (Netherlands Organization for Scientific Research) for financial support for this study. Address for correspondence: Monique Flecken, Donders Centre for Cognition, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands. e-mail: m.flecken@donders.ru.nl

relevant insights with regard to the language-and-thought debate. The present study is one of the few studies that focus explicitly on language effects related to grammatical concepts, as opposed to lexical concepts.

KEYWORDS: Linguistic relativity, verbal aspect, motion event cognition, non-verbal task, eye tracking, visual attention, Arabic, German

1. Introduction

The question posed in this paper addresses the role of linguistic knowledge in event cognition, in particular in visual processing of events where no explicit verbal representation is involved. The inter-relation between cognitive processing and linguistic form has been the topic of a centuries-long debate. Starting with the idea of an artificial language which would be best fit to represent thought (Leibniz), the assumption of a tight link between language and thought was later carried on to a fundamentally different level, the concept of *Sprachliche Weltansicht* by von Humboldt, who claimed that a person's mother tongue shapes the way the world is perceived and interpreted. Empirical research on this inter-relation, which was taken up in the field of anthropology only 100 years later, has become a topic of discussion under the terms 'Sapir–Whorf hypothesis' or 'linguistic relativity'. For a long period this position was put forward and defended on speculative grounds mixed with ideological appraisal of a specific language, depending on the historical period. Given recent developments, we are in a position to pursue these questions on a new basis due to the presence of groundbreaking techniques in the recording and analysis of cognitive processes. The old question of the inter-relation between language and thought can now be posed in a far more differentiated way (see, e.g., Casasanto, 2008; Gumperz & Levinson, 1996). We are no longer restricted to language structure as the only systematically analyzable manifestation of cognitive processes. We now have methodologies for the analysis of, for example, visual attention and temporal aspects of processing, down to the level of milliseconds, as well as the measurement of brain activity in real time. Drawing on these new techniques, the language and thought debate has gained momentum over the last years (cf. recent overviews in Cook & Bassetti, 2011; Pavlenko, 2011). However, looking at the current discussion on the relation between language and cognition we have to realize that the picture is rather diverse and limited, given the complexity that language and its use in context entails. To put this more precisely, we should speak about knowledge, which is structured and represented by linguistic form. This knowledge component is categorical, giving form to conceptual units derived from objects, actions, and properties, as well as to the principles of composition/decomposition of complex conceptual structures. Broadly speaking the first type of knowledge relates to lexical forms while the second relates to

morphosyntactic forms. Language-related knowledge in this sense is certainly one form of acquired knowledge – besides mental representations in other modes such as pictorial schemata of actions, events, objects, or motor patterns – which is relevant for interpreting and structuring incoming information. When viewing a scene, for example, a coherent interpretation is spontaneously constructed on the basis of what has previously been experienced as characteristic of the type of scene, for example, an event. Previously acquired knowledge is therefore the background against which selective attention is structured anew. One central and deeply entrenched component of relevant knowledge is thus linguistic knowledge in all its different layers and subsystems.[1]

Although there is empirical evidence that linguistic categories influence the way people organize thought, there is also counter-evidence. Limitations are given by the fact that, so far, investigations have mainly included language-specific lexical structures which have been correlated with specific patterns in non-verbal cognitive processing (*cut* and *break* verbs in Korean and English: Majid, Boster, & Bowerman, 2008; motion verbs in French and English: Soroli & Hickmann, 2010; colour terminology: Athanaspoulos, 2011; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; spatial concepts: Levinson, 2003; space/time metaphors: Casasanto & Boroditsky, 2008). Given that grammaticalized concepts in languages are abstract, highly automatized and obligatory in specific contexts, one can assume that these categories are in the foreground and accessed early when preparing content for speaking. While this effect could be shown for processes of 'thinking for speaking' (Lucy, 1996; Slobin, 1996) and 'seeing for speaking' (Flecken, 2011; von Stutterheim, Andermann, Carroll, Flecken, & Schmiedtová, 2012; von Stutterheim & Carroll, 2006) the question we would like to address in the present paper is how far these effects also show in non-verbal tasks. In the present study, visual attention is used as a window on processes of conceptualization.

2. Background

2.1. VISUAL ATTENTION AND CONCEPTUALIZATION

The method used to investigate preferences in event perception and construal across languages is the measurement of gaze allocation by means of eye-tracking, as mentioned above. This method has been shown to provide insights into the inter-relation between visual and cognitive, and, in particular, linguistic processing (see survey in Huettig, Rommers, & Meyer, 2011). Numerous studies using the so-called visual-world paradigm have provided evidence for a tight temporal link between visual attention and language

[1] Here and in the following we use the term *language*, which is ambiguous in English, in the sense of a *specific* linguistic system.

processing, mainly during language comprehension. Moving from the processing of single objects, or constellations of multiple objects, to more complex linguistic tasks that involve the comprehension and construal of events with agents and actions, studies have shown how visual attention reflects patterns of conceptually motivated selective foci of interest (e.g., Altmann & Kamide 2009). The use of eye-tracking as a window on cognitive processing allows us to disentangle the role of different factors that guide the timecourse of processing and the selection of targets in the allocation of visual attention. While it has been shown that eye-movements are tightly connected to conceptualization in language comprehension (Rayner, 2009), as well as in production (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998; Papafragou, Hulbert, & Trueswell, 2008), research on visual processing in contexts in which no overt use of language is involved has so far focused on explanatory factors other than language (see discussion below). In the study at hand, we exploit the fact that the placement of fixations allows for insights into the ongoing conceptualization of scenes, in this case scenes depicting events. By extending current approaches, we set out to test how far language-related patterns in visual attention, observed during the language production process, are habituated to such an extent that they are also activated in structuring gaze allocation in tasks without overt language use.

2.2. LANGUAGE-SPECIFICITY IN CONCEPTUALIZATION IN NON-VERBAL TASKS: THE CASE OF EVENT COGNITION

Studies based on Talmy's lexical typology of verb-framed versus satellite-framed languages (1985, 2000) have shown that, depending on speakers' language type, they allocated more or less attention to the manner of motion than to specific features of the path of motion, when watching short scenes of motion events. Some studies found this language-related preference in gaze allocation only when subjects were asked to speak about the scenes (Gennari, Sloman, Malt, & Fitch, 2002; Papafragou et al., 2008). The implications of these cross-linguistic contrasts in the encoding of manner and path features of motion events were also tested in the context of a non-verbal memory task, with the analysis of the allocation of visual attention as an indication of cognitive salience. In the non-verbal task, with the same stimuli as in the verbalization task, no language-related preference was found (Papafragou et al., 2008). The discussion of cross-linguistic differences in conceptualizing motion events was extended in a study by Papafragou and Selimis (2010), using an event matching task carried out by five-year-old children and adults with Greek or English as their native language. In developing their hypotheses, they formulated two positions in the language and thought debate, the 'salience

hypothesis' and the 'under-specification hypothesis'. The salience hypothesis claims that event cognition is generally shaped by language-specific principles acquired in the course of language development. This predicts that language-specific differences do not show only in language-related processes of event cognition, but also when language is not explicitly involved. The under-specification hypothesis, on the other hand, takes the position that linguistic and conceptual representations of events are dissociable (p. 225). The experiments designed to test these hypotheses involved categorization tasks for motion events. The first experiment was carried out with a linguistic prompt, which was then removed in the second experiment. This was followed by a third experiment in which events were presented simultaneously. The findings for Experiments 2 and 3 show that speakers of Greek, a path-salient language, and speakers of English, a manner-salient language, behave identically. In conclusion, the authors interpret their results as supporting the under-specification hypothesis. Language-specificity effects which were observed in Experiment 1 and in other (verbal, but partly also non-verbal) studies are interpreted as 'linguistic intrusions' (p. 249) which are transient and task-specific. There are, however, two major problems concerning this interpretation. First, the patterns found in cross-linguistic descriptions of motion events do not differ in black and white terms, in contrast to the initial typological classification based on Talmy (1985). Both language types allow for the expression of path as well as manner by means of a verb. Depending on the type of stimulus and the perspective selected, or induced by the instruction, speakers of different language types may diverge or converge in their choice of lexical means. In a study on the expression of motion events in French (path-language) and in German and English (manner-languages), it was found that, depending on the nature of the stimulus (real-world videos, varying in the length of the trajectory and the degree of goal orientation of the moving entity), French speakers used path verbs, but also manner verbs, clustering with German and English native speakers, despite their typological classification (Carroll, Weimar, Flecken, Lambert, & von Stutterheim, 2012; see Slobin, 2006 on the salience of motion events with 'boundary crossings' as the clearest context for typological differences). Thus, there are no one-dimensional predictions with respect to language-related conceptualization patterns based on a narrow typological distinction.

The other problem relates to the fact that in Papafragou and Selimis (2010), one aspect of the linguistic system, that is, the motion verb lexicon, is taken as the sole linguistic factor underlying language-specific effects on motion event conceptualization. However, if we look at the domain of event construal, conceptualization requires the speaker to compose a complex conceptual representation by integrating a number of concepts of a different nature

(e.g., entity, action, time, and space). It might well be that in the set-up of Experiments 2 and 3 (Papafragou & Selimis, 2010) other aspects, such as the timecourse and temporal properties of the events, rather than the spatial properties, are brought to the foreground. Since both English and Greek are languages that encode viewpoint aspect grammatically (see Smith, 1991), and are thus typologically related in this domain, it may be the case that the parallel results obtained in Experiments 2 and 3 are rooted in linguistic similarity at this level. This would mean that related conceptual patterns are rooted in a domain other than spatial expression. Interpreting the results as evidence for the under-specification hypothesis seems to be premature, given the highly reductionist view on the different aspects and domains that are involved in the cognitive processing of events.

This caveat is in order since a number of recent studies show language-specificity effects on event conceptualization, also in contexts in which language is 'suppressed' as much as possible: Soroli and Hickmann (2010) tested French and English speakers using dynamic stimuli of motion events. The experiments included linguistic encoding (verbalization), non-verbal categorization tasks, and the recording of gaze allocation patterns. The results present further evidence for language-specificity effects on visual attention in language production tasks, but they also show language-specificity effects in non-verbal categorization and gaze allocation patterns, but not on a consistent and systematic basis. The authors also point to the fact that the results for linguistic encoding show options that cut across typological patterns, which could be one major reason for the absence of a consistent pattern. They conclude by saying that, given the complexity of processes of event conceptualization, more empirical research is necessary before we can actually come to any conclusion with respect to evidence for the linguistic relativity hypothesis.

Further insights into the inter-relation between language structure and principles of conceptualization come from studies on bilingualism and second language (L2) acquisition and use. Studies on bilinguals help to sharpen our view on the phenomenon of linguistic relativity in a particular way: if it is the case that our mother tongue equips us with specific 'spectacles' through which we perceive and conceptualize reality, what happens in the case of early bilinguals who develop two linguistic systems simultaneously, and what happens in the course of second language acquisition at later stages in life? There are quite a number of studies to date which have addressed these questions empirically (see Cook & Bassetti, 2011; Pavlenko, 2011, for comprehensive overviews) and the results paint a multi-faceted picture. In studies on event conceptualization in advanced L2 speakers, where languages are learned in succession, the hypothesis was tested that the habitual conceptualization of events, which is shaped and, to some extent,

automatized in the categories of the mother tongue, also underlies use of the second language. The results obtained confirmed this hypothesis (Bylund, 2009; Carroll et al., 2012; Schmiedtová, von Stutterheim, & Carroll, 2011; von Stutterheim and Carroll, 2006). On the other hand, other studies show that, depending on the level of proficiency, the L2 speakers are able to adopt L2 principles in conceptualizing events or categorizing objects (e.g., Cadierno, 2004).

What these studies teach us is that there is certainly no evidence for linguistic determinism concerning principles of cognitive processing in the domains studied. Rather, we have to think of a continuum between tight and loose connectedness, which leaves us with the task of finding out what underlies the different degrees of entanglement. As the experience of learning and using a second language shows, conceptual patterns are undoubtedly all the more difficult to 'take in', the more abstract, complex (requiring the integration of several conceptual dimensions), and perspective-dependent they are. This means that the impact of grammaticalized categories on processes of conceptualization will be of particular interest.

In this context we have to mention those studies which have looked at the effects of grammatical categories on non-verbal conceptualization in conceptual domains other than event cognition. Huettig, Chen, Bowerman, and Majid (2010) investigated the influence of Mandarin classifiers on non-verbal processes of object classification. They interpret their results in line with the findings on motion events by Gennari et al. (2002), by saying that "Mandarin classifiers influence online overt attention only during linguistic processing of these language-specific distinctions" (Huettig et al., 2010, p. 55). In contrast to these findings, Boroditsky, Schmidt, and Phillips (2003) found language-specificity effects in a cross-linguistic study (Spanish, German, English) on the influence of gender systems on the conceptualization of object properties, when performing memory tasks, tasks of similarity assessment, and feature ascription to objects. They conclude that differences in thought can be produced solely on the basis of grammatical differences and in the absence of other cultural factors (p. 77). Other research that looked at cross-linguistic differences in number marking on nouns found that this affects the way speakers of different languages categorize objects based on their shape or material properties (for a comparison of English and Yucatec speakers, see Lucy, 1992; Lucy & Gaskins, 2001, 2003; for Japanese and English, see Imai & Mazuka, 2003; for Japanese–English bilinguals, see Athanasopoulos & Kasai, 2008).

All in all, these empirical studies leave us with an inconsistent picture of the inter-relation of language and cognitive processing and point to the need for greater stringency in the identification of the phenomena under investigation.

3. Previous study: gaze allocation to event endpoints during a language production task (von Stutterheim et al., 2012)

While it is hard to pin down cognitive correlates for, for example, case markers, other grammatical forms maintain some independent meaning that adds to the meaning of a sentence or construction. Grammatical markers of temporal categories are a case in point. As with all grammatical categories, they developed out of elements carrying lexical meaning, but in the process of grammaticalization the original lexical meaning of these elements becomes 'bleached'. The contribution of a grammatical marker to the meaning of a construction can be highly abstract, and languages differ in the extent to which specific types of temporal categories are grammaticalized. Tense and aspect markers profile particular conceptual categories, and carry a major function in temporal anchoring and perspective taking when talking about events.

Given the fact that temporal–aspectual perspective taking, as expressed by aspectual systems, is an essential component of event construal, an empirical study was carried out in order to test the conceptual implications of grammaticalized temporal categories. The question is whether, and to what extent, native speakers of languages that differ in the degree to which aspectual categories are grammaticalized show differences in event conceptualization patterns. We hypothesized that those grammatical categories which are automatized, because use is obligatory, and which are accessed early in language production, should function as a filter for conceptualization in a language production task. In this context 'viewpoint aspect' is a conceptual category which allows for temporal decomposition of a situation into phases. Under a rigid typological perspective, two groups of languages can be distinguished: languages which do or do not have grammaticalized aspect. German, Dutch and Norwegian, for example, do not have grammaticalized verbal aspect,[2] but lexical means or periphrastic forms that select a particular subinterval of a situation do exist (German: *gerade* 'right now', Dutch, for example: *aan het* X *zijn* 'to be at the X', Norwegian, for example: *sitter og* + finite verb 'to sit and' + finite verb). However, these expressions are constrained in use. On the other hand, there are languages which have highly grammaticalized markers of verbal aspect, such as English,

[2] In contrast to Norwegian or German, Dutch may be viewed as on a path towards becoming an aspect language, given the ongoing process of grammaticalization of progressive aspect. The nominalized form *aan het* (*verb*) *zijn* is used by native speakers in specific contexts with high frequency (Flecken, 2011). However, there are still strong semantic constraints connected to the use of this form. Motion events, for instance, cannot be encoded under this temporal perspective. This means that there is currently no grammatical opposition between a simple verb form and the periphrastic construction, and Dutch was categorized as a non-aspect language in the v. Stutterheim et al. (2012) study on motion events.

Spanish, Russian and Modern Standard Arabic. Although there is a large degree of differentiation within these languages with regard to the aspectual categories represented (progressive, imperfective, perfective, aorist, etc.), whereby the value of a category is dependent on other categories that are grammaticalized in the respective language, for the study at hand they were collected to form one group. This gave one group consisting of languages which do not have grammaticalized viewpoint aspect (German and Dutch – the minus-aspect languages), and one group with languages that do have grammaticalized viewpoint aspect (Spanish, Arabic, Russian, and English). In contrast to the first group, speakers of these latter languages are required to make decisions with respect to aspectual distinctions concerning the phasal decomposition of situations, rooted in grammaticalized progressive or imperfective aspect marking on each verb.

In von Stutterheim et al. (2012), speakers of these different languages were shown short video clips of real-world events. Different event types were depicted, but the critical scenes were voluntary goal-oriented motion events which were varied with respect to the degree of goal orientation and the specific phase of the event depicted in the clip: one event type presented ongoing motion events, whereas the other event type showed an event with a point of completion given by a change of state of an entity in motion, that is, from being underway to actually reaching a specific endpoint which was visible in the video clip (entering a house, going in a door, etc.). Speakers were asked to describe the events depicted with "What is happening?" in the video clips, and gaze movement was recorded during this period. Based on the research reviewed above on the inter-relation between language use and visual attention, the hypothesis was formulated as follows: When verbalizing information on scenes showing goal-oriented motion events (i.e., where a figure in motion is underway, but a possible endpoint shown in the video clip is not actually reached during the phase of the event shown), speakers of languages that do not use imperfective/progressive aspect will both attend to endpoints of the event during information intake, as well as refer to endpoints in these scenes, to a high degree. This will contrast with speakers of languages in which the temporal–aspectual concept 'event is ongoing' is grammaticized, and used frequently in this particular context, since this requires speakers to focus on a specific phase of the event as they decide which subinterval of the event is actually in progression at the time of speech. In other words, speakers of languages who use grammaticized imperfective/progressive aspect will be more likely to attend to the phase focused in the video clips when viewing them and preparing to talk about them, and less so to potential endpoints. We predict no difference between the two groups of speakers for the items showing motion events in which an endpoint is actually reached by the moving entity (von Stutterheim et al., 2012).

Twenty subjects (matched for age and balanced for gender) per language took part in the language production task. The stimuli were 60 short, live-recorded video clips of six seconds each: 10 critical items (motion events, endpoint not reached), 10 control items (motion events, endpoint reached) and 40 filler items (different types of events, e.g., a person knitting, or cleaning a table). Subjects were asked to start speaking as soon as they had recognized what was going on in the clip. During stimulus display eye-movement and speech were recorded. Following the production task, a memory test was performed in which subjects were asked to recall objects presented in the clips. These were the possible 'endpoints' of the motion events shown in the critical items (e.g., a house, a car, a garage).

The results for all measurements taken show an effect of the factor 'grammatical aspect'. The production data display a significant difference in endpoint encoding, but only for the critical scenes (those in which potential endpoints were not reached by the entities in motion): speakers of languages in the aspect group (Arabic, Russian, Spanish, English) used progressive/imperfective aspect in all cases, and they mentioned less endpoints than speakers in the no-aspect group. Interestingly, the eye-tracking analyses showed a similar pattern: speakers of the aspect group fixated the endpoints in the critical scenes less and for a shorter duration, in comparison to speakers of the no aspect group. No differences were observed for the control items (see von Stutterheim et al., 2012, for a detailed description of the analyses and results). For the memory task, the hypothesis predicted a better memory of endpoints for the no aspect group. The results confirm this hypothesis, but again only for the critical scenes (endpoint of motion event not reached). The results underline the language effects observed in the production data and in the speakers' allocation of visual attention to specific aspects of the scenes.

The authors summarize that, in accordance with features of the verbal system used, and not – as one might assume – in correlation with cultural differences, speakers of the two aspect groups differ in (i) the selection of a temporal perspective, that is, phasal decomposition by means of viewpoint aspect (progressive, imperfective), as indicated by the segment of the route selected for mention, vs. construal of the event in holistic terms by inclusion and mention of an endpoint; (ii) the allocation of visual attention when processing the event scenes both before and during production (fixation on endpoints); and (iii) memory of specific components (i.e., endpoints) of the motion event.

These results supported the hypothesis since an effect of grammaticalized concepts on cognitive processing was found during the verbal task. This suggests that conceptualization is not only affected by lexicalized, but also by grammaticalized concepts, thereby presenting a new finding.

The previous predominant focus on lexicalized conceptual categories is all the more surprising, since grammatical form may be the basis for carrying out highly automatized routines in language production, given the complexity and speed of delivery involved. Concepts profiled by grammatical systems can be viewed as providing a 'default' scaffold, the most familiar route, when structuring content for speaking. In this sense grammaticalized conceptual categories constitute a major basis for a language effect. In conclusion, grammar can be put forward as providing a cognitive filter for attention allocation and information selection. However, this claim has to be substantiated and elaborated on in different directions. One of these is the question as to how 'deep' this language effect reaches with respect to cognitive processing in general. The present follow-up study takes the analysis a step further by looking at potential language effects on non-verbal cognitive processing.

4. A focused analysis of German and Arabic: verbal and non-verbal experiments

For the present study, the languages showing the clearest contrast in the language production task described above were taken as a starting point for further analysis. These were German (from the no aspect group) and Modern Standard Arabic[3] (part of the aspect group), referred to simply as 'Arabic' from now on. The follow-up study tests potential language effects on visual attention in a task in which explicit language use is not involved, using the same set of stimuli; this will be referred to as the 'non-verbal task', or Experiment 2, in the following sections. To provide the statistical background for the non-verbal task, the data obtained in the first study were reanalyzed for German and Arabic only (Section 4.1, Experiment 1).

4.1. EXPERIMENT 1 – VERBAL TASK: RE-ANALYSIS OF LANGUAGE PRODUCTION AND EYE-TRACKING DATA (VON STUTTERHEIM ET AL., 2012)

4.1.1. *Participants*

Speakers of German ($N = 20$) were students from southern Germany. They were age- and gender-matched and from comparable social backgrounds (university students). Speakers of Arabic were from Tunisia, Algeria, and Morocco ($N = 20$), age- and gender-matched, also from similar educational and social backgrounds (students at universities in the respective countries).

[3] Modern Standard Arabic is a variety of Arabic which is considered the official "high" language in Arabic speaking countries. It is used in academic and professional contexts, as well as in the media and in written and spoken modalities.

4.1.2. *Stimuli*

The stimuli used were a set of 60 dynamic, real-world video clips, with different event types. Each video clip was six seconds in length and was preceded by a black screen with a focus point, which lasted eight seconds. Event types included the set of motion events (critical, endpoint not reached, $N = 10$, and control items, endpoint reached, $N = 10$) and a set of filler items (see list of motion event stimuli in the 'Appendix'). In each motion event scene, an entity in motion (an animal, a vehicle, or a person) was shown moving along a route (road, path, lawn, etc.) towards a specific endpoint object (a house, a gate, a petrol station, a playground, etc.). In the critical items, the video clips end while the moving entity is still on its way and has not reached the possible endpoint (see an example of a screenshot of a critical item in Figure 1).

The filler items consisted of causative events, in which one agent was shown performing an action on one specific object (e.g., a woman seated on a sofa, knitting a scarf, or a man standing in a room, washing a plate at a sink). Each recording was preceded by a training session with six items covering all categories.

4.1.3. *Apparatus*

Gaze movement was recorded by means of a remote Eye Follower II eye-tracker, and run on the software NYAN. The software was specifically developed for use with dynamic stimuli (i.e., recording and analyzing eye-movement on a frame-per-frame basis) and for language production experiments. The tracker's sampling rate was 120 Hz, with a 0.45 degree gaze-point tracking accuracy throughout the operational head range. The TFT monitor was 20" and participants were seated approximately 60 to 70 cm from the screen. Calibration was carried out once for each participant before the experiment (tracking eye gaze on yellow dots on a black screen which appeared in identical order at specific positions on the screen). Automatic recalibration only occurred when necessary in the inter-stimulus interval during the experiment.

4.1.4. *Procedure*

Each session started with the following instruction which participants were asked to read:

> You will see a set of 60 video clips showing everyday events which are not in any way connected to each other. Before each clip starts, a black screen with a white focus point will appear. Please focus on this point.

Fig. 1. Screenshot of a critical item: two people are walking in a park (towards a climbing rack).

> Your task is to tell 'what is happening', and you may begin as soon as you recognize what is happening in the clip. It is not necessary to describe the video clips in detail (e.g., "The sky is blue"). Please focus on the event only.

Instructions were translated into the two languages by native speakers, and the experimenter was also a native speaker of the language tested. This means that all exchanges took place in the participants' native language to ensure that this was fully activated during the experiment. Given the automatic adaptation of the cameras to eye position, no recalibration or validation was necessary during the production task. Cases in which initial calibration was not successful were excluded. Each session lasted approximately 15 minutes with no option of manipulating the presentation pace of the 60 items. Following the eye-tracking experiment, participants spent approximately five minutes filling out a questionnaire on their educational and linguistic background.

4.1.5. *Data coding and analysis*

The transcribed data were coded for the encoding of endpoints, and both transcripts and codes were checked by a second researcher. Language production and eye-tracking data were evaluated per language and compared between the two languages.

Gaze movement was recorded during the entire time the video clip was playing, that is, for six seconds per item. For the analyses of the eye-tracking data, one 'area of interest' (AoI) was defined, which included the endpoint area of the motion event for each critical and control item. This area remained

fixed in the respective clip while the figure moved along a path. AoIs differed slightly in size depending on the area at goal and always included one specific, identifiable object (e.g., a house, a climbing rack).

In order to quantify patterns of eye-movement, we compared the overall number and duration of fixations on the endpoint. Fixations in the AoI (endpoint) were calculated by NYAN using an area-based algorithm where a set of fixations with a maximum deviation of 25 screen pixels and a minimum sample count of six was recognized as a fixation.

For all measures, data were analyzed using linear mixed effect models, using the package lme4 (version 0.999999-0; Bates, Maechler, & Bolker, 2012) in the software R (version 2.15.3). For each measure, one interaction model was set up, taking language (German, Arabic) and condition (control: Endpoint (EP) reached, critical: EP not reached) and, most importantly, their interaction, as fixed effects. Participants and items (stimuli) were included as random effects (random intercepts), to control for unwanted participant and item variability. Random slopes were not included in the model as the present study concerns a between-subject and between-item design (each item and each subject is unique within condition and language). In all analyses, the control condition (EP reached) was coded as the base level, and for language the reference group was Arabic; t-values of +/−1.96 are interpreted as statistically significant ($p < .05$), marked in bold and with an asterisk in all tables; p-values are marked in the tables as well (formula in R for all analyses: lmer (measure ~ Language * Condition + (1|subject) + (1|item)).[4]

4.1.6. *Results*

Language production data: frequency of endpoint encoding. The number of endpoints mentioned was compared between the two types of motion event scenes (critical condition: endpoint not reached; control condition: endpoint reached) and the two languages (German and Arabic). In the critical condition, the scenes depicted motion events in which a potential endpoint was not reached by the moving entity (a car or a person) ($N = 10$). In the control condition, the motion event scenes did show that endpoints were reached by the moving entity in the stimulus, for example, a person walking into a building ($N = 10$). Table 1 and Figure 2 show the relative frequency of endpoint encoding by speakers of the two languages ($N = 20$ per language) for the two conditions.

[4] P-values were added to clarify the findings obtained. They were calculated in R on the basis of the t-values, with the following code: tvalues <- fixef(model) / sqrt(diag(vcov(model))) pvalues <- 2*(1-pnorm(abs(tvalues))).

TABLE 1. *Relative frequency of endpoint encoding*

	Average % of EP encoding	SD
Arabic		
Control condition	83	15.31
Critical condition	33.50	18.57
German		
Control condition	92	9.19
Critical condition	62	18.29

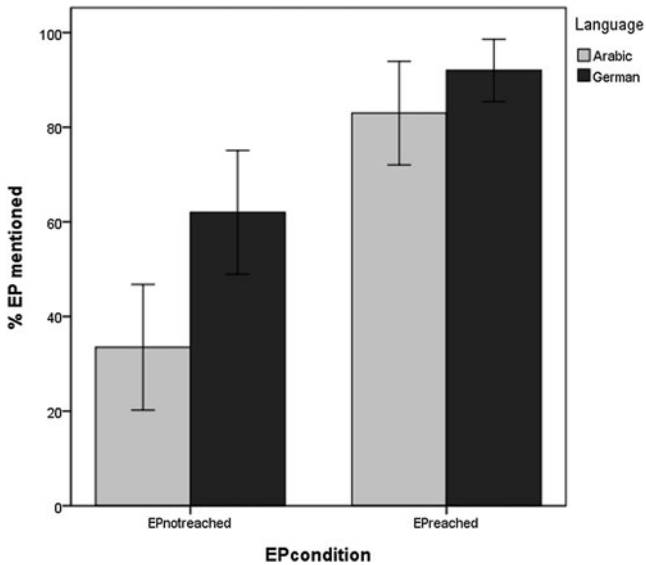

Fig. 2. Bar charts showing the mean relative frequency of endpoint encoding in the verbal task, in the critical and control conditions (EP condition 'EP not reached' and 'EP reached') (error bars indicate 95% confidence intervals).

The relative frequency of endpoint encoding was analyzed by setting up a mixed effect regression model, detailed above. Table 2 shows the results of this analysis.

We find a significant main effect of language and condition, and the critical interaction between language and condition is significant as well. The findings can be interpreted as showing that German speakers mention significantly more endpoints in the verbal task than speakers of Arabic, and mainly so in the critical condition.

Eye tracking data: total number and duration of fixations on endpoints. In this section, we report on the total frequency and duration of fixations in the AoI,

TABLE 2. *Fixed effects in mixed model on data of endpoint encoding (language data)*

Verbal task: endpoint encoding	Factor	Estimate	Std. error	*t*-value	*p*-value
Interaction Model	(Intercept)	33.5	4.995	6.707	< .001
	Language	28.5	5.267	**5.412***	< .001
	Condition	49.5	7.064	**7.007***	< .001
	Language* Condition	−19.5	7.448	**−2.618***	< .05

that is, on the endpoint, during the entire stimulus presentation period (six seconds), for both languages (see Table 3 and Figure 3) and in both conditions.

Total number of fixations on endpoints. Table 4 gives the results of the mixed model set up to analyze the patterns found. We find a significant interaction between our fixed factors, and a marginally significant main effect of condition. In general, we find a higher frequency of fixations on the endpoint in the critical condition. The interaction effect lies in the fact that German speakers fixate endpoints more frequently than Arabic speakers, in the critical condition – they seem to show a stronger differentiation between conditions than the Arabic speakers.

Total duration of fixations on endpoints. Table 5 depicts the results of the mixed model that was set up to analyze the data on the duration of fixations on endpoints, in the verbal task. In this model, we find a significant interaction between language and condition, and a marginally significant main effect of condition. Results may thus be interpreted as indicating a trend in the following direction: on average, the German data show greater differentiation between the two conditions; they tend to fixate endpoints longer in the condition in which they are not reached by the moving entities in the video clips.

4.1.7. *Summary and discussion of findings: verbal task* (*Experiment* 1)

To summarize this section, we find a difference in the linguistic encoding of endpoints for the critical scenes (endpoint not reached), that is, German native speakers mention them more frequently than speakers of Arabic, but there was no difference in the control condition (endpoint reached). This finding confirms our hypothesis with respect to the differing degrees of salience of endpoints in linguistic event encoding, when there is the option of not mentioning an endpoint. An option is given in the sense that a potential endpoint object is present in the visual scene, and an entity is indeed moving in its direction, but within the time span of the video clips, the

TABLE 3. *Mean total number and duration (in ms) of fixations on EPs in the verbal task, in the critical and control conditions*

	Mean number of fixations on EP	SD
Arabic		
Control condition	3.13	1.27
Critical condition	4.78	1.70
German		
Control condition	2.65	0.89
Critical condition	5.44	1.36
	Mean duration of fixations on EP (ms)	SD (ms)
Arabic		
Control condition	535	180
Critical condition	853	274
German		
Control condition	471	140
Critical condition	998	330

TABLE 4. *Fixed effects in mixed model on data of endpoint fixation frequency*

Verbal task: fixation frequency	Factor	Estimate	Std. error	*t*-value	*p*-value
Interaction model	(Intercept)	3.125	0.700	4.462	< .001
	Language	−0.480	0.420	−1.143	.253
	Condition	1.655	0.936	1.769	.077
	Language*Condition	1.135	0.376	**3.016***	< .001

TABLE 5. *Fixed effects in mixed model on data of endpoint fixation duration*

Verbal task: fixation duration	Factor	Estimate	Std. error	*t*-value	*p*-value
Interaction model	(Intercept)	0.535	0.127	4.220	< .001
	Language	−0.064	0.073	−0.874	.382
	Condition	0.317	0.170	1.864	.062
	Language*Condition	0.210	0.067	**3.118***	< .001

endpoints are not reached by the moving entity. Thus, when endpoints are optionally part of the event, and can be inferred as being part of the motion event on the basis of the visual input, German speakers select them as part of their representation of the motion events more frequently than speakers of Arabic.

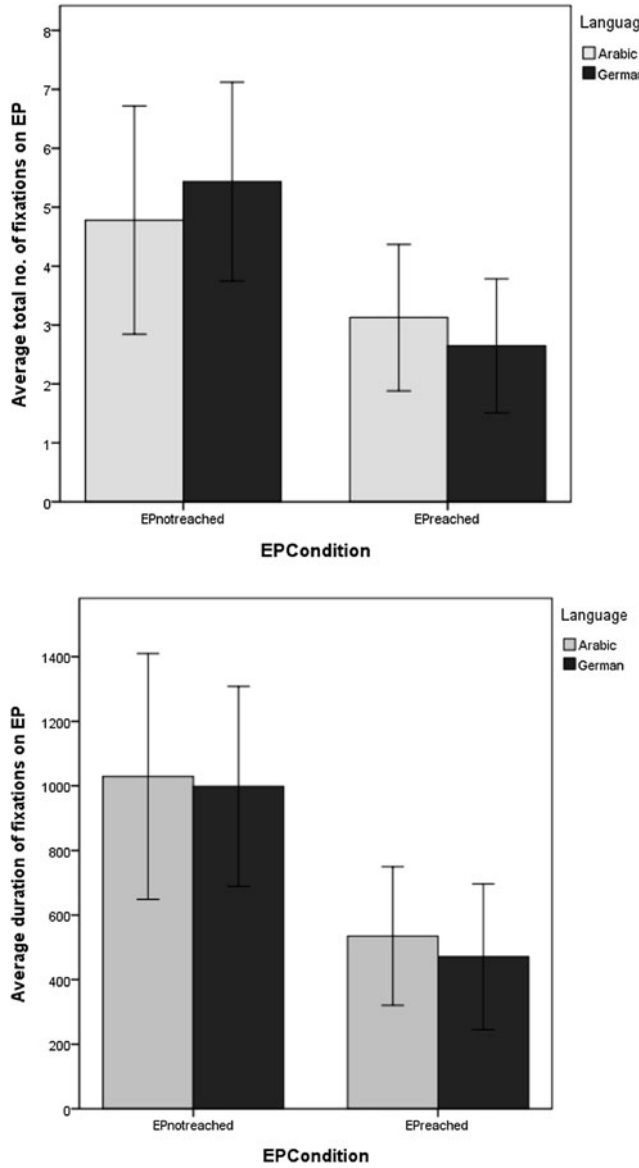

Fig. 3. Mean number (top) and duration (bottom) of fixations on EPs, in the critical and control conditions.

The eye-tracking data show a similar picture: in the critical condition, speakers of both languages gaze more often and for a longer time at the endpoints in the stimuli compared to the control condition. We interpret this as showing that, if an endpoint is immediately evident, and integrally part of the event as depicted in the scene (which is the case in the control scenes as a moving entity reaches or arrives at the endpoint), speakers of all languages will encode this component of the event linguistically. Speakers of all languages will thus have to direct attention to endpoints in order to retrieve the names of the objects when preparing for verbalization. No cross-linguistic differences are to be expected with respect to the encoding of the endpoint in these scenes. In the critical condition, it is not evident from the start of the stimulus whether the entity in motion is 'heading for' a goal, or simply 'going along' a path of some sort. This optionality will thus lead to an increase in fixations on the optional goal, since speakers of all languages will need to process information such as the orientation of the entity towards the goal, and the distance to goal, to decide whether the event in question is a 'reaching a goal' event or not. The data suggest that this is where language comes in as a guiding factor. Although both language groups gaze at endpoints when they are optional and when the reaching of the EPs is ambiguous, there is a strong tendency for German speakers to do so to a greater extent, given their preference for holistic event representation, even after taking into account between- item and between-participant variability in the statistical analysis of the data. Arabic speakers tend to do so to a lesser extent, given a grammar-driven focus on phasal decomposition of events. The focus of their attention lies on the phase depicted in the video clip, that is, the ongoing, intermediate phase, and Arabic speakers allocate less attention to the potential endpoints present in the video clips, when deciding what phase is actually 'ongoing' at the time of speech. In the current analyses of fixation frequency and duration, we find a significant interaction between language and condition, with a higher degree of visual attention to the endpoint region by speakers of German, in the critical condition. The data suggest more pronounced differences for fixation data in the German group between conditions, compared to fixation behaviour in the Arabic group.

4.2. EXPERIMENT 2: NON-VERBAL TASK

4.2.1. *Method*

In order to test non-verbal gaze patterns, a task is needed which does not call for explicit linguistic encoding, but which does require visual inspection of the same motion event scenes as those used in Experiment 1 (verbal task). The same stimulus material was used in Experiment 2, but the number of control items was reduced from ten (verbal task) to five in order to shorten the

overall experimental procedure. The same experimental set-up and apparatus for recording gaze movement was used, and recordings were carried out in the same laboratory.

The task needed to be explicitly non-verbal (i.e., void of the requirement for explicit language use), as our research question relates to potential language-specific gaze patterns in non-verbal event processing. The aim was to design a distracter task that would require focused attention on the whole of the scene, without inducing very cautious scan patterns of all or only specific elements of the scene. We argue that a free-viewing paradigm without an additional task may not be a suitable context to investigate event processing: In the case of free viewing, it is possible that eye-movement patterns are guided by mainly visual principles relating to perceptual saliency, for example, in speakers of all languages. The fact that we use dynamic stimuli might induce an attentional bias to the dynamic aspects of the stimuli (i.e., the entity in motion), and leave only little room for allocating attention to aspects of the background. Remember that, in our critical stimuli, endpoints are not highly salient and clearly backgrounded in relation to the moving entity.

On the other hand, use of a paradigm in which participants are asked to view visual scenes to prepare for an upcoming scene recollection/recognition or memory task may induce careful task-related inspection of the scene, paying attention to all details, thus leaving no scope for potentially language-based patterns associated with event construal. Furthermore, as the eye-tracking results for the verbal task are tentative only, given the nature of the task (spontaneous event description) and stimuli (naturalistic, dynamic stimuli), it was important in the design of the non-verbal task to eliminate other factors that might affect visual attention as much as possible.

A distracter task was designed that serves to relate visual processing of the event scenes to auditory input which is presented simultaneously: While watching the scenes the participants received continuous auditory (non-verbal) input (the sound of ocean waves) and were asked to attend to specific sound cues (loud beeps) that occurred randomly and occasionally in the sound stream. When the sound cue occurred, the task for the participants was to memorize the scene in which the sound cue had been played. Video clips were presented in six blocks of seven trials each. The number of sound cues in each block was either two or three (this was randomized between blocks). After each block, numbered screenshots of the seven scenes were presented on the screen simultaneously (four on the upper part and three on the lower part of the screen). The participants were asked to announce, out loud, the numbers of the screenshots of the video clips in which they had perceived the sound cue (see Section 4.2.5 for more details). The sound cues did not occur during trials in which critical or control event scenes

(motion events) were displayed – this served to avoid a potential attention bias towards the motion event scenes.

Given the present paradigm, we cannot fully exclude the recruitment of linguistic strategies to solve the task. However, the sound cue occurred randomly within blocks, and the number of cues is randomized between two or three times within one block as well. This should prevent habituation, or the development of conscious strategies on the part of the participant, in that participants had to pay attention to the auditory signal, as well as to the visual material, continuously in order to fulfil the task.

4.2.2. *Participants*

Two different groups of native speakers of German (N = 20) and Arabic (N = 20) took part in Experiment 2. The German native speakers were all university students (undergraduates and postgraduates) at the University of Heidelberg in Germany (counterbalanced for gender, mean age 26.5 years). They were all students of non-language-related disciplines and their answers in a language background questionnaire indicated they had no very advanced knowledge of a second language. Participants who indicated a very high level of proficiency in a second language, or who had lived in a non-German-speaking country for more than one semester at a stretch, were excluded from the analyses.

The Arabic-speaking participants were carefully recruited by a native speaker assistant. All communication before and during the experiment was carried out by this native speaker. The Arabic speakers were participants in a German language course and were all enrolled in a beginner-level class. They were recruited and recorded during the first five days of their stay in Germany. Their knowledge of German was low to non-existent, as they indicated having only recently started to learn German at university in their respective countries of origin (first-year students). They were not able to conduct a conversation with a German native speaker, as this was tested after the experiment. The Arabic participants came from Tunisia, Morocco, and Algeria, and had learned and spoken Modern Standard Arabic (MSA) in school. Speakers learn MSA at the latest when they begin to attend school, as this is the language used in education, both at school and university, as well as in the media. There was a slightly higher number of male than female participants (male: N = 13, female: N = 7).

4.2.3. Stimuli and apparatus

The apparatus was the same as in Experiment 1. The stimuli used were also the same as in Experiment 1, though the number was reduced to 42 dynamic

video clips (to create six blocks with seven trials each). Again, each video clip was six seconds in length and was preceded by a black screen with a focus point, which lasted eight seconds. The event types in the set consisted of the same set of motion events (critical scenes: endpoint not reached, $N = 10$; control items: endpoint reached, $N = 5$) and a set of filler items. The 27 filler items again involved agentive causative events.

4.2.4. Procedure

Participants were given written instructions in their native language, while a native speaker experimenter was sitting next to them, in order to clarify possible questions and to guide them through the process of calibration, etc.

The instructions were translated into German and Arabic by native speakers and read as follows:

> You will see 6 sets of 7 video clips showing everyday events that are not in any way connected. Each scene is preceded by a blank screen with a focus point. All of the scenes have a continuous sound in the background (ocean waves).
>
> In some of the scenes, however, you will hear a different additional sound.
>
> The clips will be presented in sets of 7, in other words, the video will be stopped after you have seen 7 clips. Seven screenshots of the clips will then appear on the computer screen (all will appear at once) – numbered 1 to 7.
>
> Your task is to select the screenshots of the clips in which you heard the additional sound. In order to do so, please say the numbers of the relevant screenshots aloud.
>
> You will have 20 seconds to perform this task. During this time it is also possible to revise the selection made by saying the relevant numbers aloud again.
>
> When the 20 seconds are up the next set of 7 video clips will start playing automatically.

After this instruction, the participants were shown an example of a stimulus, and the sound cue (the beep) was played a couple of times to ensure that the participant would recognize it. At this time, the audio track with the sound of ocean waves was started, which lasted in the background until the end of the experiment.

The experiment was preceded by a training block during which participants could get used to the task, and the experimenter was able to give feedback, if necessary. All communication took place in the native language of the participant. Following this training session, calibration was performed and the experiment was started.

When the experiment was finished, participants were asked to fill out a short questionnaire on their linguistic and socio-cultural background, which took about five minutes. After this, there was a short debriefing, during which subjects filled in another short questionnaire, asking them explicitly about strategies they had used to perform the task, and asking whether they remembered using silent speech or inner verbal labels to memorize the scenes with the beeps. The total procedure took about 25 minutes.

4.2.5. Results: non-verbal task (Experiment 2)

In line with the analyses of the eye-tracking data in the verbal task, we report on total frequency and duration of fixations on the endpoint AoIs in the video clips, during the entire stimulus presentation period, for both languages and both conditions (critical: EP not reached; control: EP reached). We analyzed the data by setting up a mixed model for the fixation frequency and duration data, in the same way as described above for the verbal task. Table 6 and Figure 4 give an overview of the mean number and duration of fixations on endpoints.

Total number of fixations on endpoints. Table 7 shows the results for the mixed model set up to analyze the data on fixation frequency on endpoints, in the non-verbal task. The results show a significant interaction between language and condition.

The data show a substantial difference in fixation frequency between languages in the critical condition only, and for the German group there is a larger difference in fixation frequency between conditions than for the Arabic speakers; this group fixates endpoints more frequently when they are not reached by the entities in the video clips (critical condition: EP not reached).

Total duration of fixations on endpoints. Table 8 gives the details of the mixed-effect model on the data of endpoint fixation duration. The interaction between language and condition does not reach significance level, nor are there main effects of the fixed factors, thus indicating that there are no differences in fixation duration on endpoint regions in the stimuli.

5. Comparison between experiments

The findings for the non-verbal task show a significant interaction between language and condition, for the frequency of fixations on endpoint regions only; following the interaction effects in the verbal task experiment, we find that German speakers fixated the endpoints in the critical video clips more frequently than speakers of Arabic. Only in the verbal task, we find a language effect for the duration of fixations on endpoint regions in the video clips. This means that, in the verbal task, German speakers not only fixated endpoints

TABLE 6. *Mean total number and duration (in ms) of fixations on EPs in the non-verbal task, in the critical and control conditions*

	Mean number of fixations on EP	SD
Arabic		
Control condition	1.06	0.55
Critical condition	1.90	1.19
German		
Control condition	1.31	0.58
Critical condition	3.16	1.18
	Mean duration of fixations on EP (ms)	SD (ms)
Arabic		
Control condition	234	127
Critical condition	579	375
German		
Control condition	287	156
Critical condition	819	235

TABLE 7. *Fixed effects in mixed model on data of endpoint fixation frequency (non-verbal task)*

Non-verbal task: fixation frequency	Factor	Estimate	Std. error	*t*-value	*p*-value
Interaction model	(Intercept)	0.774	0.689	1.124	.261
	Language	0.265	0.381	0.696	.487
	Condition	1.270	0.785	1.618	.057
	Language*Condition	0.974	0.382	**2.549***	< .05

more frequently than Arabic speakers in the critical condition, but they also fixated them for a longer period of time. This pattern in the verbal experiment is given by specific task requirements, and by the type of processing that the allocation of visual attention may entail in a language production task: the duration of fixations (dwell time) on specific objects reflects processes of word retrieval (i.e., lexical access during the formulation stage in language production; cf. Griffin, 2004; Griffin & Bock, 2000). German speakers mentioned endpoints more frequently, meaning that they had to fixate endpoints for a longer amount of time, since they were retrieving names for these objects, and they were processing them as part of the verbal structure being generated. In the non-verbal task, on the other hand, the retrieval of words for the objects fixated (the endpoints) was not required, and this is reflected in the absence of the language × condition interaction for fixation duration. This, in fact, supports our interpretation of this task as being

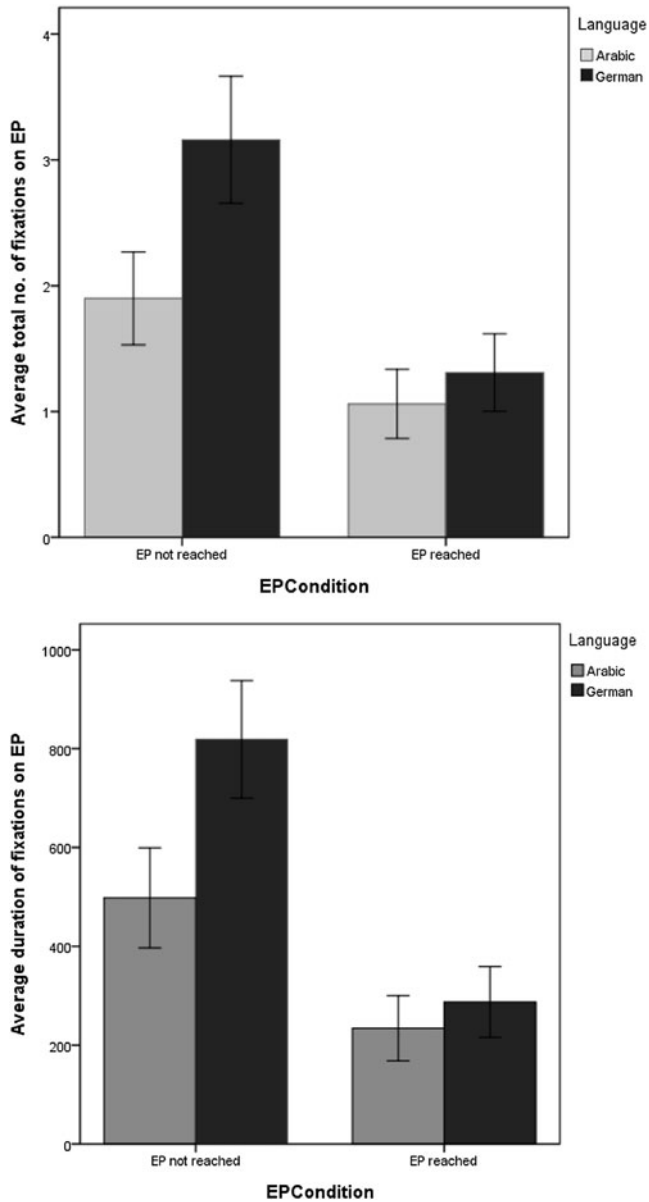

Fig. 4. Mean number (top) and duration (bottom) of fixations on EPs, in the critical and control conditions.

TABLE 8. *Fixed effects in mixed model on data of endpoint fixation duration (non-verbal task)*

Non-verbal task: fixation duration	Factor	Estimate	Std. error	*t*-value	*p*-value
Interaction model	(Intercept)	66.850	13.404	4.987	< .001
	Language	10.540	10.353	1.018	.310
	Condition	−2.948	15.906	−0.185	.853
	Language*Condition	15.213	11.908	**1.278**	.201

non-verbal in nature, and the significant interaction found for fixation frequency thus may actually be based on attentional processing which is not related to language use.

The very different task requirements and different groups of speakers tested in the two experiments do not allow for direct statistical comparisons, thus, we will focus on a brief qualitative comparison of fixation patterns. In overall terms, two main differences can be observed. First of all, we see that, in the non-verbal task, endpoints were fixated less than in the verbal task. This finding indicates task-related differences: when asked to report on an event, one may direct more attention to backgrounded aspects of a scene, in comparison to one's viewing behaviour in a non-verbal event recognition task. Information on objects along the path of motion in the background (which our endpoint objects in all video clips are) may become relevant for participants in the verbal task over time: the event depicted in the video clip actually unfolds as time progresses. Participants cannot immediately tell whether certain backgrounded parts of a scene are going to be relevant, when drawing up the sentence plan while performing the task of describing 'what is happening in the video clip'. In this case they may direct attention to potentially relevant objects along the path of motion. As the scene, and the participant's conceptual representation of the scene, develops over the course of stimulus display, it may be necessary to retrieve information with respect to naming the potential endpoint object in the background. Participants will have to set up a specific conceptual representation or structure of the event, since the task relates specifically to the domain of events and asks for explicit information on a specific event and its structure ("What is happening?"). Since we are dealing with motion events, this means that, besides a source, a path or trajectory and an entity which is moving, a potential endpoint or goal is always part of the abstract structure of a motion event.

On the other hand, in the non-verbal task, we cannot say with certainty how the participants conceived of the task – the domain of events, and all representations associated with it, may or may not have been activated when participants were performing the sound cue recognition task. The fact

that we did not find the interaction between language and condition for the duration of fixations on endpoints may be interpreted as showing that participants were not retrieving names for objects in the scenes, even though German participants fixated the potential endpoint objects more frequently than Arabic participants – reflecting more attention to endpoints, regardless of the necessity to label them, neither explicit nor implicit.

The second main observation concerns the following aspects: in the verbal task we find that endpoints were fixated more frequently in the critical than in the control condition, regardless of language. In the non-verbal task, this difference is less clear: endpoints were fixated longer in the critical condition, but only German speakers showed more fixations on the endpoint in the critical condition, compared to the control condition. This therefore shows that the language effect is stronger in the non-verbal than in the verbal task: endpoints seem to be specifically relevant for German speakers when they are processing events, in whatever type of task. For German speakers, it may be the case that the default process in scanning a motion event always includes direction of attention to potential endpoints, given their linguistic preference for holistic perspective taking. This way of perceiving events may have become habituated based on linguistic preferences.

The fact that Arabic speakers attended more to endpoints in the verbal than in the non-verbal task can be explained on the basis of decisions that need to be made during language planning: Arabic speakers, with their rich aspectual system, need to activate a specific verbal form from a range of options that are not available to a German speaker (only simple present tense verb forms would be relevant in this context). The activation of a specific verbal form with a specific type of morphological marker of viewpoint aspect depends on the factual unfolding of the event and the degree of goal-orientation of the moving entity, as it proceeds along a path. This means that they will at least have to allocate some attention to this region in the verbal task, when compared to the non-verbal task.

6. Summary and discussion of results

The starting point of the non-verbal experiment (Experiment 2) was given by a previous study on language-specific effects in event construal. The effect was investigated for the grammatical feature 'verbal aspect', in particular grammaticalized markers of progressive aspect, using motion events in video clips as stimulus material. The hypothesis underlying this study was the following: if a language has grammaticalized verbal aspect, then speakers have to make a choice with respect to the phase of a situation they are going to verbalize, the phase that is ongoing at the time of speech. In processing visual input they will therefore pay more attention to the specifics of the

phasal unfolding of a situation, in contrast to speakers of a language that does not require its speakers to make a decision of this kind.

The results of the present non-verbal study show that there is a significant effect which can be interpreted in support of this hypothesis. Speakers of Arabic pay less attention to potential endpoints in scenes depicting motion events (in the sense of frequency of fixation) in comparison to German subjects. This could be due to many reasons: cultural differences, differences in training traditions, differences in handling the experimental task, etc. However, in comparison with the results from a previous language production study (von Stutterheim et al., 2012), in which information selection, visual attention, and memory performance was tested for speakers of seven languages, we find a consistent pattern for other languages that both do and do not share the same feature (groups of aspect and no aspect languages). Speakers of Arabic, as part of the aspect group, and speakers of German, as part of the no aspect group, show a parallel contrast in fixation patterns in the non-verbal task. Given this consistent pattern across a number of conditions, we interpret the differences observed as related to typological diversity with respect to the grammaticalization of temporal–aspectual categories, in this case imperfective or progressive aspect and its function in motion event construal. A language which requires a speaker to make aspectual distinctions predisposes the speaker to constantly discriminate between different phases of a situation. In a task where speakers are asked to verbalize what they currently perceive, they have to focus on the part of a situation that is actually depicted. In deciding on the use of a particular aspectual perspective (e.g., 'ongoing', 'progressive', 'perfective'), the speaker is committed to the factualness of what is asserted in the utterance. A speaker of a language that does not require this distinction (and provides only lexical means to express temporal–aspectual viewpoints which are restricted in use to specifically marked contexts),[5] is led to under-specify the role of phasal decomposition in event construal and to construe events according to a holistic perspective. In other words, the speaker takes a maximal viewing frame (cf. Langacker, 2008). Given this perspective, events are construed on the basis of changes of states, which in the case of motion events can be achieved by the reaching of an endpoint represented by a somehow 'plausible' object (e.g., a house in the direction of which a person is walking, a garage when a car is driving toward it, etc.).

The eye-tracking results of the non-verbal task show similarities and differences which correspond to the implications of the grammatical systems involved. If an endpoint is shown as being reached by an entity in motion, speakers of both languages direct attention to these endpoints in equal terms. If only an ongoing activity (motion along a route) is shown in the clip, with

[5] Adverbs such as *‚gerade‘* ('right now') or periphrastic forms in German which can be used to express ongoingness are not used in the present experimental context.

an object in the distance that could be interpreted as a potential endpoint, the two groups of speakers behave differently. German participants direct significantly more attention to the possible endpoint than Arabic participants.

The allocation of visual attention to potential endpoints in the verbal task may be interpreted as related to the conceptual implications of the respective grammatical system. The fact that these differences also appear in a non-verbal task suggests that these implications are deeply entrenched and operate as a default in processing visual input. This is not to say that these patterns are deterministic; they can always be overruled by specific requirements of a specific task. And this would account for the diversity of findings in this field. What we want to argue for is a moderate (or 'weak') relativist position. Conceptual categories formed and represented through language provide automatically and (pre-)attentively functioning strategies which are brought to bear in cognitive processing on a default basis.

7. General discussion

Turning now to the implications of the results for the key issues in the language and cognition debate, we will address the following three points: (a) the role of general cognitive quasi 'natural' principles in event construal in non-verbal tasks; (b) language effects due to structural properties; and (c) language-on-cognition effects in non-verbal tasks.

(a) *The role of general cognitive quasi 'natural' principles in event construal in non-verbal tasks*. As reviewed above, previous research on event construal in non-verbal tasks advocates the position that language-shaped categories retreat into the background in favour of general cognitive principles guiding segmentation and structuring of perceptual input. The results obtained in the study at hand call this position into question. Our results suggest that there are language-on-cognition effects in attending to dynamic visual input. Given the methodology used, which was designed to suppress the activation of language as much as possible, we still cannot positively say anything about the actual representation of the perceived scene. In how far speakers represent units that correspond to an 'event' is something we do not know. But what we can say is that there is a significant language effect in attending to the scenes, which corresponds with language effects in analogously designed verbalization tasks. This finding leaves us with two possible explanations: the first explanation is that we see an indirect impact of language which could be explained developmentally. In the course of language acquisition, patterns of cognitive processing develop simultaneously, at least to some degree induced by linguistic structure, but in any case independent of explicit linguistic representation. These highly abstract and, in the case of grammatically induced cognitive categories, completely automatized patterns function as a basic tool box for

cognitive processing. The other explanation could be that language is always involuntarily activated in perceptual tasks. This is, for instance, the position put forward by Papafragou et al.'s (2008) and Trueswell and Papafragou's (2010) notion of 'linguistic intrusions'. While we cannot exclude this possibility theoretically, the implications for the interpretation of our results in the context of the language-on-cognition hypothesis do not differ from the first explanation. We find a language effect on gaze allocation when no explicit language use is involved, and this is similar to the effect found in a verbal task.

(b) *Language effects due to structural properties*. When reviewing the literature there is a clear tendency to study language-on-cognition effects with respect to conceptual categories systematically represented in the lexicon. One of the few studies which have addressed the role of grammatical categories is Huettig et al. (2010), which tested the influence of Mandarin classifiers on eye-gaze behaviour. In this study, linguistic influence was only found when subjects used language explicitly, but not in the context of a non-verbal task. The study by Boroditsky et al. (2003) on object categorization obtained different results: they did find effects of gender marking on the assignment of properties to objects. These diverging results point to the fact that in the field of linguistic relativity it would be inadequate to pose the question of language-on-thought effects as an either–or alternative. Rather we have to assume an intricate interplay of general cognitive principles, such as those derived from physical experience, and specific, variable principles formed in the course of language acquisition and socio-cultural development. This means that we have to work on a microscopic empirical level in order to be able to differentiate between the different effects on habituated cognitive processing.

If we find a language-on-cognition effect on the basis of lexicalized categories, it seems even more likely that we will observe an effect on the basis of grammaticalized categories, since grammar forms a component of language which is activated automatically, obligatorily, and pre-attentively. The fact that no structural effects have been found in studies so far (cf. Huettig et al., 2011) seems to be due to the methodological complexity involved in investigating grammatical effects specifically, rather than the different roles grammar and lexicon play in structuring cognition. Lucy's (2011, p. 49) view of the role of structures of meaning (i.e., grammatical structures) is very much in line with our position when he says:

> … but the strategy [of studying grammatical differences] is difficult to implement: Comparing categories across languages requires extensive linguistic work in terms of both local description and typological framing, can be derailed by blindness to categories very different from one's own, and may not easily yield referential entailments suitable for an independent assessment of cognition. Nonetheless, this strategy holds the most potential

for closely respecting the linguistic differences and thus holds the greatest promise for identifying structural differences and directing the search for cognitive influences in appropriate directions.

The study at hand points to the fact that it is worth making first steps in this direction.

(c) *Language-on-cognition effects in non-verbal tasks*. The fact that a language effect has not been reported in previous non-verbal studies on event perception (Gennari et al., 2002; Papafragou et al., 2008) could be due to different reasons. One reason could be that different conceptual domains are affected by language-related factors to a different extent. While this line of argumentation seems justified, at least when looking at lexically versus grammatically induced concepts, it is less convincing for diverging findings with respect to basically the same types of concept. Still, this position cannot be rejected on the basis of the evidence available so far.

Another explanation could lie in the specific biases introduced by the experimental designs. If, for instance, participants are requested to remember the stimulus material on a global basis, as was the case in Papafragou, Massey, and Gleitman (2002), one would expect participants to scan the entire visual input as accurately as possible. In this case, participants will probably focus on objects and features of objects depicted, rather than interpret the scene as an 'event' (i.e., something that is happening). This would mean that language-specific processing patterns in event construal do not come into play, just because there is no process of event conceptualization taking place: the relevant categories and patterns associated with event construal, which include patterns influenced by habituation based on language use, are not activated in these tasks. As far as the results in Papafragou et al. (2008) are concerned, we suggest that it may be too simplistic to consider language-specificity effects only on the basis of one conceptual alternative, in this case the encoding of manner versus path information in verbs or other linguistic means. Spatial as well as temporal categories come into play as soon as the encoding of entire events is investigated, in contrast to single or multiple object naming. The way in which these different components are weighted in conceptualization, or the way in which they interact depending on the specific task and the actual type of content, is something we do not know. The present results show that a claim such as "conceptual organization is independent of language-specific encoding" (Papafragou et al. 2008) is premature. What we can say, given the state of the art, is that we are only at the start of the process of obtaining insights into the complex relationship between language and cognition. This calls for caution, as well as fine-grained analyses, with respect to the questions posed and the methodology used when aiming at an overarching model of the inter-relation between language and cognition.

REFERENCES

Altmann, G. & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: eye-movements and mental representation. *Cognition*, **111**, 55–71.

Athanasopoulos, P. (2011). Colour and bilingual cognition. In V. Cook & B. Bassetti (eds.), *Language and bilingual cognition* (pp. 241–262). New York: Psychology Press.

Athanasopoulos, P. & Kasai, C. (2008). Language and thought in bilinguals: the case of grammatical number and nonverbal classification preferences. *Applied Psycholinguistics*, **29** (1), 105–123.

Bates, D., Maechler, M. & Bolker, B. (2012). Package 'lme4'. Linear mixed effects models using S4 classes [Computer software]. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html.

Boroditsky, L., Schmidt, L.m & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (eds.), *Language in mind: advances in the study of language and cognition* (pp. 59–80). Cambridge, MA: MIT Press.

Bylund, E. (2009). Effects of age of L2 acquisition on L1 event conceptualization patterns. *Bilingualism: Language and Cognition* **12** (3), 305–322.

Cadierno, T. (2004). Expressing motion events in a second language: a cognitive typological approach. In M. Achard & S. Neimeier (eds.), *Cognitive linguistics, second language acquisition and foreign language pedagogy* (pp. 13–49). Berlin: Mouton de Gruyter.

Carroll, M., Weimar, K., Flecken, M., Lambert, M. & von Stutterheim, C. (2012). Tracing trajectories: motion event construal by advanced L2 French–English and L2 French–German speakers. *Language in Interaction and Acquisition*, **3** (2), 202–230.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Cross-linguistic differences in temporal language and thought. *Language Learning*, **58** (1), 63–79.

Casasanto, D. & Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition*, **106** (2), 579–593.

Cook, V. & Bassetti, B. (eds.) (2011). *Language and bilingual cognition*. New York: Psychology Press.

Flecken, M. (2011). Event conceptualization by early bilinguals: insights from linguistic and eye tracking data. *Bilingualism: Language & Cognition*, **14** (1), 61–77.

Gennari, S., Sloman, S., Malt, B. & Fitch, T. (2002). Motion events in language and cognition. *Cognition*, **83**, 49–79.

Griffin, Z. (2004). Why look? Reasons for eye movements related to language production. In J. Henderson & F. Ferreira (eds.), *The integration of language, vision, and action: eye movements and the visual world* (pp. 213–247). New York: Taylor & Francis.

Griffin, Z. & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, **11**, 274–279.

Gumperz, J. & Levinson, S. (eds.) (1996). *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Huettig, F., Chen, J., Bowerman, M. & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, **10**, 39–58.

Huettig, F., Rommers, J. & Meyer, A. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, **137**, 151–171.

Imai, M. & Mazuka, R. (2003). Re-evaluation of linguistic relativity: language-specific categories and the role of universal ontological knowledge in the construal of individuation. In D. Gentner & S. Goldin-Meadow (eds.), *Language in mind: advances in the issues of language and thought* (pp. 430–464). Cambridge, MA: MIT Press.

Langacker, R. (2008). *Cognitive Grammar: a basic introduction*. New York: Oxford University Press.

Levinson, S. (2003). *Space in language and cognition*. Cambridge: Cambridge University Press.

Lucy, J. (1992). *Grammatical categories and cognition: a case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. (1996). The scope of linguistic relativity: an analysis and review of empirical research. In J. Gumperz & S. Levinson (eds.), *Rethinking linguistic relativity* (pp. 37–69). Cambridge: Cambridge University Press.

Lucy, J. (2011). Language and cognition: the view of anthropology. In V. Cook & B. Bassetti (eds.), *Language and bilingual cognition* (pp. 43–68). New York: Psychology Press.

Lucy, J. & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: a comparative approach. In S. Levinson & M. Bowerman (eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge: Cambridge University Press.

Lucy, J. & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (eds.), *Language in mind: advances in the study of language and thought* (pp. 465–492). Cambridge, MA: MIT Press.

Majid, A., Boster, J. S. & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: a study of cutting and breaking. *Cognition*, **109** (2), 235–250.

Meyer, A., Sleiderink, A. & Levelt, W. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, **66**, 25–33.

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

Papafragou, A., Massey, C. & Gleitman, L. (2002). Shake, rattle, 'n' roll: the representation of motion in language and cognition. *Cognition*, **84**, 189–219.

Papafragou, A. & Selimis, S. (2010). Event categorization and language: a cross-linguistic study of motion. *Language and Cognitive Processes*, **25** (2), 224–260.

Pavlenko, A. (ed.) (2011) *Thinking and speaking in two languages*. Clevedon: Multilingual Matters.

Rayner, K. (2009). Eye movements and attention in reading, scene perception and visual search. *Quarterly Journal of Experimental Psychology*, **62** (8), 1457–1506.

Schmiedtová, B. (2011). Do L2 speakers think in the L1 when speaking in the L2? *VIAL International Journal of Applied Linguistics*, **8**, 138–179.

Schmiedtová, B., von Stutterheim, C. & Carroll, M. (2011). Implications of language-specific patterns in event construal of advanced L2 speakers. In A. Pavlenko (ed.), *Thinking and speaking in two languages* (pp. 66–107). Clevedon: Multilingual Matters.

Slobin, D. (1996). From thought and language to thinking for speaking. In J. Gumperz & S. Levinson (eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Slobin, D. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (eds.), *Space in languages: linguistic systems and cognitive categories* (pp. 59–81). Amsterdam: John Benjamins.

Smith, C. (1991). *The parameter of aspect*. Dordrecht: Kluwer Academic Press.

Soroli, E. & Hickmann, M. (2010). Language and spatial representations in French and in English: evidence from eye-movements. In G. Marotta, , A. Lenci, , L. Meini & F. Rovai (eds.), *Space in language* (pp. 581–597).Pisa: Editrice Testi Scientifici.

von Stutterheim, C. & Carroll, M. (2006). The impact of grammaticalised temporal categories on ultimate attainment in advanced L2-acquisition. In H. Byrnes (ed.), *Educating for advanced foreign language capacities: constructs, curriculum, instruction, assessment* (pp. 40–53). Georgetown: Georgetown University Press.

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M. & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, **50** (4), 833–867.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (ed.), *Language typology and syntactic description (Vol. 3, Grammatical categories and the lexicon*, pp. 57–149). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics (Vol. II, Typology and process in concept structuring)*. Cambridge, MA: MIT Press.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B. & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on pre-attentive color perception. *Proceedings of the National Academy of Sciences*, **106** (11), 4567–4570.

Trueswell, J. & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: evidence from dual-task paradigms. *Journal of Memory and Language*, **63**, 64–82.

Appendix

Motion event stimuli used for analyses

Critical condition: *Endpoint not reached* 10 items

Video clip	Motion event
1	a van is driving down a country lane (towards a village/houses)
2	a woman is walking across the parking lot (towards a car)
3	a woman is walking down an alley (towards a barrier)
4	a little boy is walking along a path (towards a playground)
5	a man is climbing up a ladder (to a balcony)
6	a man is crossing a street (towards a car)
7	two girls are walking along a path (towards a house)
8	a girl on a horse is riding (towards an entrance)
9	a mother and a child are walking through a park (towards a slide)
10	a car is driving down a road (towards a petrol station)

Control condition: *Endpoint reached* 10 items (verbal task), 5 items (items 1–5) (non-verbal task)

Video clip	Motion event
1	a car is driving into a garage
2	a girl is entering the station
3	a van is turning into a driveway
4	a man on a bicycle is turning into a gateway
5	a woman is entering a supermarket
6	a dog is running through the door of a building
7	a cat is walking into the kitchen
8	a child is going through a gate into a playground
9	a man is walking into a church
10	a girl on a horse is riding into a barn/stable