

How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance¹

CHRISTIANE VON STUTTERHEIM, MARTIN ANDERMANN,
MARY CARROLL, MONIQUE FLECKEN, AND BARBARA SCHMIEDTOVÁ

Abstract

The role of grammatical systems in profiling particular conceptual categories is used as a key in exploring questions concerning language specificity during the conceptualization phase in language production. This study focuses on the extent to which crosslinguistic differences in the concepts profiled by grammatical means in the domain of temporality (grammatical aspect) affect event conceptualization and distribution of attention when talking about motion events. The analyses, which cover native speakers of Standard Arabic, Czech, Dutch, English, German, Russian and Spanish, not only involve linguistic evidence, but also data from an eye tracking experiment and a memory test. The findings show that direction of attention to particular parts of motion events varies to some extent with the existence of grammaticized means to express imperfective/progressive aspect. Speakers of languages that do not have grammaticized aspect of this type are more likely to take a holistic view when talking about motion events and attend to as well as refer to endpoints of motion events, in contrast to speakers of aspect languages.

1. Introduction

The way in which events are perceived and conceptualized is shaped in part by factors that are task driven and perspective based, as illustrated by studies on phenomena such as inattentional blindness (Mack and Rock 1998) and change blindness during information intake (Simons and Levin 1997; Simons 2000). If a witness at court is asked to provide information on an event, for example, the response given in solving this task will be defined by the setting as well as the nature of the question: “what did you actually see at that point?”, or “what did the bank robber look like?” The way information is organized in language production will differ in both cases. Before any statement is made, the speaker will have to activate his knowledge of the event as a whole, select the material he wants to put into words and then decide on the order in which it should be presented. These processes relate to the phase of *conceptualization* (Levelt 1999) in language production, i.e., the phase in which speakers prepare information for expression before passing it on to the *formulator*. Clearly, the nature of the verbal task, as defined by the explicit or implicit question, influences these initial processes. But are they also influenced by the structural properties of the particular language used? In other words, is message preparation at the level of the *conceptualizer* to some extent language-specific? One form of specificity results from the particular way in which a language lexicalizes concepts, compared to those that are grammaticized, since both play a role in how events, or parts of events, are described.

An area where this is relevant in the description of events is temporality, in particular the concept of *aspect*, which relates to the perspective under which particular temporal properties of an event are presented. In English, for example, one and the same situation can be described as ongoing (*John was crossing the street*) or as completed (*John has crossed the*

street); in fact, aspect is a core grammatical category in English where use is obligatory in specific contexts: *Jane is in the basement. What is she up to? She *fixes the shelf.* The simple form *fixes* is not acceptable, since speakers are required to make the aspectual distinction given with *she is fixing the shelf*, where the *ongoingness* of the event is made explicit for the relevant time of assertion. In other languages, such as German, for example, the concept of aspect is not grammaticized. If speakers want to differentiate between various ways of presenting one and the same situation temporally, they must choose other means such as temporal adverbials (*Hans überquerte gerade die Straße*, lit. Hans crossed *just now* the street ‘Hans just crossed the street’) or periphrastic constructions such as *Hans war dabei, die Straße zu überqueren* (lit. Hans was *there-at* to cross the street ‘Hans was just crossing the street’), but these devices are optional in German.

The precise meaning of viewpoint aspect (cf. Smith 1991) and the degree to which related concepts are grammaticized varies considerably across languages. Thus, the Russian simplex imperfective form (*on rabotal* he work-IMPERFECTIVE-PAST ‘he was working’) is related to, but not identical with the English progressive. Language-specific differences become manifest when the speaker eventually puts the message into words: in the formulator stage of language production, the speaker must use the constructions his language offers. But do they already influence language processing at the conceptualization stage in language production, the phase in language production for which the well-known *thinking for speaking* hypothesis has been proposed (Slobin 1996)? This hypothesis states that the preparation of information for verbalization will be shaped by specific linguistic categories available in the speaker's linguistic system.

The present study investigates whether linguistic categories focus speakers' visual attention on certain aspects of a given event while preparing to describe it. We formulate this as the *seeing for speaking* hypothesis: If a particular language (A) encodes a particular concept grammatically and native speakers of language (A) use this grammaticized concept frequently and systematically, then speakers of this language are very likely to attend to visual features of a given situation that are linked to this concept. By contrast, if language (B) encodes the same concept lexically, or by phrasal means, and native speakers of language (B) do not encode this concept frequently and systematically, then they may not attend to the respective features of a given situation, or at least not to the same extent.

The current study focuses on event conceptualization and covers seven languages. It shows how structural differences in the temporal-aspectual domain affect patterns in language production, using eye tracking as the main tool in investigating the phase of conceptualization in language production. The data consist of dynamic stimuli showing motion events. The analysis builds on a number of earlier crosslinguistic comparisons, which, together with relevant studies by other research groups, will be outlined in the following section.

2. Earlier research

Initial research in the field of visual attention and language production (eye tracking studies) focused on relatively simple linguistic tasks such as object naming or the production of single-sentence event descriptions, using pictures as stimuli (cf. Meyer et al. 1998; Griffin and Bock 2000; Meyer and van der Meulen 2000). Factors driving attention in the events depicted (agent-action-patient) have been attributed to the degree of agentivity, since speakers direct attention to the participant which is highest on a scale based on a continuum between proto-agent and proto-patient. This participant will be encoded as the first constituent in the clause, if necessary by means of a passive construction: *a man is being chased by a dog* (Griffin and Bock 2000). What happens, however, if the grammatical structure of a language does not show a close relationship between subject role and initial position, in contrast to English?

German, for example, does not have a tight grammatical link between syntactic subject and clause-initial position, since it presents the option of placing constituents other than the subject in initial position (adjuncts, (in-)direct objects, etc.). Crosslinguistic differences of this kind are potentially relevant when investigating the link between patterns of visual attention and language production. Given the rate at which decisions in language planning are executed in language production, concepts that have paved their way into the grammar of a language may serve, on a default basis, in the direction of attention and selection of associated preferences during conceptualization and formulation. We assume that they are highly automatized and facilitate high speed access during language production in relevant contexts (Carroll et al. 2004).

Research so far has focused on language-specific differences related to different *lexicalization patterns* across languages, in particular to the way in which verbs encode information on *manner of motion* versus information on *direction*. Research on these preferences was initiated by Talmy (1985) and has revealed a number of remarkable and often very subtle differences in lexicalization (cf. Slobin 1997a, 1997b). A further question addressed in this research framework is whether language-specific mapping preferences also affect non-linguistic conceptual performance. The experimental methods used in investigating this question cover memory tests, similarity judgments, recognition and categorization tasks. Typical memory investigations within the lexicalization paradigm examine whether English speakers, for example, may be more likely to remember information on manner of motion while path information may be more prominent for Spanish speakers, in keeping with the lexicalization patterns found in these languages. Although there are clear findings showing linguistic preferences in linguistic tasks, results for language-specific differences in non-linguistic tasks are mixed, however (for memory tests relating to a range of different tasks: Gennari et al. 2002; Papafragou et al. 2002; Pavlenko 2003; Marian and Fausey 2006; for recognition tasks: Billman et al. 2000; for categorization tasks: Soroli and Hickmann 2010). Slobin (2000) shows how the different lexicalization patterns in English and Spanish lead to different mental images, when Spanish and English speakers are dealing with the same text. The possible role of grammaticized linguistic means in the direction of attention in language production was explored in an eye tracking study on motion events which compared how speakers of English and German process the relevant visual input (von Stutterheim and Carroll 2006). When viewing a series of everyday events (video clips which include a set of motion events) and telling *what is happening*, speakers of English conceptualize the event as ‘in progression’ and segment the situation into phases (inchoative, intermediate, terminative phase): *a car is driving along a country road* (intermediate phase); *a truck is approaching a village* (terminative phase), thereby focusing on the phase that is prominent in the stimulus. Speakers of German take a holistic view and typically represent the event — whatever phase of the event has actually been depicted — with an endpoint (*ein Auto fährt auf einer Straße zu einem Dorf*, a car drives on a road *to* a village). The linguistic differences are reflected in the degree of visual attention paid to the endpoint, as depicted in the video clip: native speakers of English first direct attention to the phase focused in the video clip (intermediate phase). Fixations on a possible endpoint occur after speech onset and information on this phase is less likely to be mentioned. Speakers of German are more likely to view the event in holistic terms and direct attention to a possible endpoint from the outset, i.e., before speech onset, in contrast to English speakers. The findings of the eye tracking study on the distribution of attention across motion events by speakers of English and German reveal processing differences with respect to direction of attention during information intake, as well as systematic differences in event conceptualization which are language-specific.

An eye tracking study on direction of attention in motion events (Papafragou et al. 2008), which compares English and Greek speakers, also reveals differences in the extent to which speakers attend to specific features of motion events, as hypothesized on the basis of

lexicalization patterns. Greek speakers are more likely than English speakers to attend first to the path-endpoint region, while English speakers are more likely to attend first to manner of motion. Direction of attention at the earliest stages of event conceptualization is affected by linguistic encoding preferences, when information is verbalized in the task. The authors conclude, however, that language-specific differences observed in event conceptualization when verbalizing information on the event do not affect non-linguistic conceptual performance.

In summary, previous findings show that although language-specific categories clearly affect event conceptualization when thinking for speaking, the actual scope of these effects is still unclear.

3. The present study

The present study focuses on temporal concepts and shows how speakers of seven languages, which differ in the way aspectual concepts are grammaticized, may also differ in the direction of attention to temporal features of events while performing a linguistic task.

3.1. *Typological classification*

The languages Standard Arabic, English, Russian, Spanish, Czech, Dutch and German differ with respect to the presence of *imperfective/progressive* aspectual means — the temporal viewpoint that explicitly represents an event as ongoing — in the grammar, and the actual use of these imperfective/progressive aspectual means by native speakers of these languages. Table 1 provides an overview of the relevant differences in the languages studied. It classifies the aspectual verb-morphological systems in broad terms, taking into account both standard descriptions found in typological analyses of tense and aspect systems (cf. Dahl 1985; 2000),² as well as findings based on recent empirical studies (see below).

Table 1. *Tense-Aspect systems*

Temporal categories grammaticized	Arabic	English	Russian	Spanish	Czech	Dutch	German
Tense	no	yes	yes	yes	yes	yes	yes
Imperfective aspect	yes	no	yes	yes	yes*	no	no
Progressive aspect	yes	yes	yes secondary imperfective	yes	yes secondary imperfective	no	no

*this classification has been revised in different respects in recent years (see Section 3.2).

It is important to note that the typological classification of these languages with respect to availability of aspectual means in the grammatical system in Table 1 is simplified. In the case of motion events — the focus of the present study — the relevant feature of the linguistic system for event conceptualization concerns the presence or absence of a grammaticized aspectual marker to express an event as ‘ongoing’ (*progressive* or *imperfective* aspect). Recent findings for reference to motion events show that the aspectual systems of Arabic and Slavic languages, as well as English and Spanish, function differently depending on context.

Our hypothesis as to how speakers of these languages conceptualize motion events is based on preliminary studies that relate exclusively to language use. Our line of argumentation is based on the way in which the relevant means available are actively *used* by native speakers in context. In this sense, the table depicted above does not reveal to what extent speakers of English, Arabic or Russian actually use the aspectual distinction ‘event is ongoing’ and the means that encode it. The findings for motion events show how the use of aspect in Czech and Russian, for example, differs in this domain. We will thus take a closer look at differences in the aspectual systems that are relevant for the present study.

3.2. Aspect systems and language use

Starting with the basic aspectual contrast in which the time course of an event can be presented as *imperfective* (event is ongoing) or *perfective* (event is completed), Arabic has a fully grammaticized³ system in which this contrast is marked morphologically on the verb. In Russian, grammatically-based distinctions combine with lexical aspect to encode this basic aspectual contrast, but none of the other languages in the study have a full-fledged aspectual system that covers both distinctions across a range of contexts.

What this means for the conceptualization of motion events will be described briefly for the present study in which participants were asked to view video clips online (see Section 3 below) and tell *what is happening*. With regard, first, to the question of representing an event as completed (perfective aspect), or not (imperfective/progressive aspect), the videos used in the experiment include one set of clips in which the completion of the motion event can be inferred; in this case the entity which is on its way towards a likely goal point (*a boy is going along a path*) does not reach the endpoint (*a playground*) during the time span shown in the clip. This condition allows for use of either one or the other aspectual distinction when talking about the event.

In addition to the selection of the possible aspectual contrasts mentioned above, we also have to take into account the temporal viewing point relevant for aspectual choice. In being asked to tell *what is happening*, the relevant temporal viewing point is given by the deictic *now* of the speaker. So in selecting the aspectual distinction (event is ongoing) and deciding what is actually ongoing at the time of speech (now), this typically leads to a segmentation of the event into one of its sub-phases (inchoative phase, intermediate phase, terminative phase)

Speakers of Arabic, for example, will typically use imperfective aspect, thereby focusing on what is ongoing with regard to what is viewed as the intermediate phase. Although this pattern could apply, roughly speaking, for the other aspect languages, the aspectual systems of these languages differ in certain respects (see extensive descriptions in e.g., Dahl 2000; Dickey 2000; Sasse 2002; Klein and Li 2009). In English, as indicated in the introduction, speakers can view any section of a motion event as in progression, including the phase that relates to the point of completion, using the same morphological form *be + V-ing* (*the vehicle is leaving the parking lot, travelling along the road towards the building*), (cf. von Stutterheim and Carroll 2006).

With regard to use of linguistic means in the given online elicitation task, we find that while all speakers of English select the progressive, German and Dutch speakers do not use lexical or periphrastic forms that encode an event as *ongoing* in explicit terms (see von Stutterheim et al. 2009). The use of linguistic means in the two Slavic languages is less straightforward, however. Although Czech and Russian speakers use what are classified as the same verbal forms (the so-called ‘simplex imperfective’) for the description of goal-oriented motion events, the critical items in this study, the imperfective form is aspectually unmarked in this context in Czech, but not in Russian (see Schmiedtová and Sahonenko 2008: 59; Schmiedtová 2011). The observations on the use of forms, and their function, support a

position which argues that Czech, as a member of the West Slavic group, does not have a full aspectual opposition perfective–imperfective, in the sense “event completed” vs. “event is ongoing” (see Dickey 2000). Russian, as a member of the East Slavic languages, encodes this contrast within its aspectual system as a clear-cut opposition.⁴ In other words, verb forms can be classified as “simplex imperfective” in Czech and Russian, but the aspectual system in which they operate differs in both languages. There is thus no evidence from the linguistic data that motion events in Czech are phasally segmented and represented as ongoing.⁵ Given these findings on language use, Czech is *not* included in the aspect group for the distinction relevant in the present task. Spanish takes up an in-between position, so to speak, within the group of languages, since progressive aspect, although not obligatory, is used in the present tense on a productive basis when talking about motion events (Bylund 2009).

To summarize the findings for language use across the seven languages studied, the following picture arises with respect to temporal perspective-taking: Speakers of English, Spanish, Russian, and Arabic relate event times to viewing time, leading to a phasal decomposition of events. Speakers of Dutch, German and Czech do not typically encode subphases of an event, but represent them holistically.

3.3. Hypotheses

The findings for event descriptions for Arabic, Russian, English, and Spanish in contrast to Czech, German and Dutch lead to the following hypothesis: When verbalizing information on scenes showing goal-oriented motion events (i.e., where the figure in motion is underway, but a possible goal point 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 critical 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. In other words, speakers of languages who use grammaticized *imperfective/progressive* aspect will be more likely, when viewing the clip, to attend to the phase focused in the clip, the intermediate stage of the event; as mentioned above, this does not show the entity in motion actually approaching the possible goal (although a possible goal is visible). We hypothesize that this will lead to different degrees of attention to the endpoints that are visible (but not reached by the moving entity) in the video clips. Based on the description of the linguistic systems, as well as previous studies on language use described above, German, Dutch and Czech speakers will form one cluster, namely the group of languages with a higher degree of attention to endpoints (for the critical items in our study). We predict no difference between the two groups for the items showing motion events in which the endpoint is reached, since, in these cases, the terminative phase of the motion event (entity in motion approaching a goal) is focused in the clip.

3.4. Eye tracking in dynamic scenes

3.4.1. *Participants.* Each language group consisted of 20 native speakers from comparable socio-cultural backgrounds (students and post graduates), aged between 20 and 35. Numbers were balanced for gender and participants had normal or adjusted vision. All participants took part in the language production experiment and the subsequent memory task (see Section 3.6). Note that the linguistic and eye tracking data of one participant in the Spanish group had to be discarded from the analyses, due to a technical problem that occurred during the experiment.

Data collection was carried out in the language laboratory at the Institute for German as a Foreign Language Philology, University of Heidelberg. All native speaker participants

were given a questionnaire with questions on their social and linguistic background. Speakers were excluded from the analyses if it turned out that they had a very advanced knowledge of a second language,⁶ or had been resident for over three months in a country where a language other than their mother tongue was spoken. The speakers of Dutch, English, Czech, Russian, Spanish, and Arabic were participants in a summer school at the University of Heidelberg, all with very little knowledge of German. They were recruited for the experiment during the first 5 days of their stay in Germany, by a native speaker of the respective language. The instructions for the experiment were also given by a native speaker of each language. We ensured that the native speaker participants were selected so as to be as “monolingual” as possible.

3.4.2. Apparatus. The apparatus used in recording eye movement was the remote system *Eye Follower*TM developed by Interactive Minds, Dresden, Germany on the basis of an *LC-Technologies* system. The cameras were attached to the monitor for binocular eye tracking and the eye gaze system accommodated all natural head movements during normal computer operation. The gaze point sampling rate was 120 Hz, with a highly accurate 0.45 degree gaze-point tracking accuracy throughout the operational head range. The TFT monitor was 20” and participants were seated approximately 50 to 80 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 occurred when necessary in the inter-stimulus interval during the experiment.

3.4.3. Material. The stimulus material consisted of 60 video clips of six seconds in length. The inter-stimulus interval, in which a black screen with a white focus point was shown, lasted eight seconds. This allowed participants sufficient time to verbalize relevant information. The clips showed everyday situations that were filmed and cut by the project group (a screen shot of one of the critical items is shown in Figure 1).



Figure 1. *A car driving along a road (to a village)*

Ten critical items were combined with ten control items and embedded in 40 fillers. The critical items consisted of scenes showing a figure in motion (animal, vehicle or person) on its way along a road or track. The scenes finish before reaching an endpoint, but a potential endpoint is visible in all cases (e.g., in the scene showing a car going along a country road, a village can be seen at the end of the road, as in Figure 1; similarly, two people can be seen walking along a path towards a house without actually reaching it. See Appendix 1 for a full list of critical and control items). In the ten control items, the figure in motion reaches a goal (e.g., walking into a house, driving into a garage).⁷ The fillers showed ten static scenes (e.g., a

candle burning, a person sun bathing) and 30 dynamic scenes with causative events (e.g., a person making a necklace; molding a vase). The 60 clips were presented in pseudo-randomized order. Four lists were established in which stimuli were pseudo-randomized and participants within a language group were assigned to these lists on an equal basis, thereby reducing a possible biasing effect related to the order in which the different situation types were presented (e.g., five speakers of language A saw randomized list 1, the next five speakers of language A saw list 2, etc.); critical and control items were embedded between at least two filler items. Each recording was preceded by a training session with six items covering all categories.

3.4.4. *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 blank screen with a white focus point will appear. Please focus on this point, since this allows us to proceed to the next video clip. 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 all languages by a native speaker, and the experimenter was also a native speaker of the language tested. This means that all exchanges took place in the participants’ native language in order to ensure that this was fully activated during the experiment. Given the automatic adaptation of the cameras to eye position, no recalibration was necessary during the production task. Cases in which initial calibration was not fully 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 the questionnaire on their educational and linguistic background. They were then asked to carry out a memory test which took between two to five minutes. This task was used to test memory performance with respect to the (potential) endpoints shown in the video clips (see Section 3.6). This part of the experiment was not announced at the outset so speakers could not prepare for it during information intake.

3.4.5. *Data coding and analysis.* The transcribed data were coded for verbal forms (temporal/aspectual categories) as well as references to endpoints, and both transcripts and codes were checked by a second researcher. Language production and eye tracking data were evaluated per language and compared crosslinguistically.

Gaze movement was recorded during the entire time that the video clip was playing, i.e., 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. Although the AoIs differed slightly in size depending on the area at goal, they always included one specific, identifiable object (e.g., a house, a car, a playground, etc.).⁸ In the clip illustrated in Figure 2, for example, the AoI analyzed (house at the bend in the road) is framed by a dark grey rectangle.



Figure 2. *A car driving along a road (to a village): AoI*

The analyses were carried out automatically using the software system NYAN. NYAN is adapted to the requirements of analyzing eye gaze in relation to dynamic visual input. In order to quantify patterns of eye movement, the measures adopted in the present analyses were the overall number of fixations in AoI (total fixation count), the total duration of fixations in the AoI (total fixation duration) and the number of first and second passes in AoI. A pass was defined as the period of fixation from the first fixation in the area of interest until the first fixation outside the area of interest.⁹ Fixations within the AoI were calculated by NYAN using an area-based algorithm. With this algorithm a set of fixations with a maximum deviation of 25 screen pixels (corresponding to a gaze movement of less than roughly 0.5° and approximately 68 cm distance from eye to screen) and a minimum sample count of six was recognized as a fixation. Accordingly, all samples with a greater deviation (i.e., gaze movement) were treated as saccades, i.e., movements covering more than 0.5° in scene perception (at the average distance and monitor dimensions given).

3.5. Results¹⁰

3.5.1. *Language production data.* Taking both the critical and control items, “endpoint not reached” and “endpoint reached”, the linguistic data produced by speakers of the different languages were compared across languages. As indicated above, the number of endpoints mentioned in a given group was taken as the basis for comparison in both conditions. No distinction was made with regard to the specific way in which the endpoint was encoded (no distinction was made between endpoint encodings by means of verbal morphology or adjuncts, or whether an endpoint mentioned was represented as reached or not: e.g., *driving into/driving towards X* were both counted as “endpoint mentioned”). The following examples (Table 2) for Arabic, Czech, Dutch, English, German, Russian and Spanish illustrate the *typical* (most frequent) ways in which motion events were verbalized, both in the control as well as the critical condition.

Table 2. *Examples of typical encodings of motion events in the critical and control condition*

	Control condition (endpoint reached)	Critical condition (endpoint not reached)
Arabic	<i>'imra 'a tadxul ilā l-maḥatta</i>	<i>sayyāratun tasīru ʿala tarīqin saḡīr</i>

	woman enters-IMPF to DEF-train station 'a woman enters the train station'	car rides-IMPF on road small 'a car drives on a small road'
Czech	<i>dívka běží na nádraží</i> a girl runs to train station 'a girl runs to the train station'	<i>auto jede do vesnice</i> car rides into village 'a car drives <u>into a village</u> '
Dutch	<i>een meisje rent een station binnen</i> a girl runs a station inside 'a girl runs into a station'	<i>een auto rijdt richting een dorpje</i> a car rides direction a village 'a car drives <u>towards a village</u> '
English	<i>a girl is running-PROG into a train station</i>	<i>a car is driving-PROG along a road</i>
German	<i>eine Frau läuft in ein Bahnhof hinein</i> a woman walks in a train station thither-in 'a girl walks into a train station'	<i>ein Auto fährt zu einem Dorf</i> a car drives to a village 'a car drives <u>to a village</u> '
Russian	<i>devuška bežít na vokzal</i> young girl runs-IMPF into train station 'a girl runs into a train station'	<i>mašina edet po doroge</i> car rides-IMPF on the street 'a car drives on the street'
Spanish	<i>una mujer está entrando a la estación de trenes</i> a woman is entering-PROGR in the train station 'a woman is entering the train station'	<i>una furgoneta circulando por una carretera</i> a truck driving-PROGR on a road 'a truck driving on a road'

Taking the English example as representing the “aspect group”, the encoding options for the critical items (motion events in which the endpoint was *not* reached) in the aspect languages will typically involve mention of the *path* along which the entity in motion is progressing, or the *location* at which the motion event takes place (e.g., *a car is driving along the road* / *a car is driving in the countryside*). For the non-aspect languages, encodings of the critical items typically include an endpoint. It is important to note that all languages offer the option of encoding the endpoint in the critical condition. In other words, there are no linguistic constraints on the possibility of including endpoints as a part of the utterance. The analysis involved a comparison of the absolute frequency of occurrence of these types of verbalizations, i.e., we were looking at *preferences* for the group of speakers as a whole.

Figure 3 depicts boxplots of the average relative frequency of endpoints mentioned in the control and critical condition, for the two language clusters (the “plus aspect” group, and the “minus aspect” group) and for each individual language.

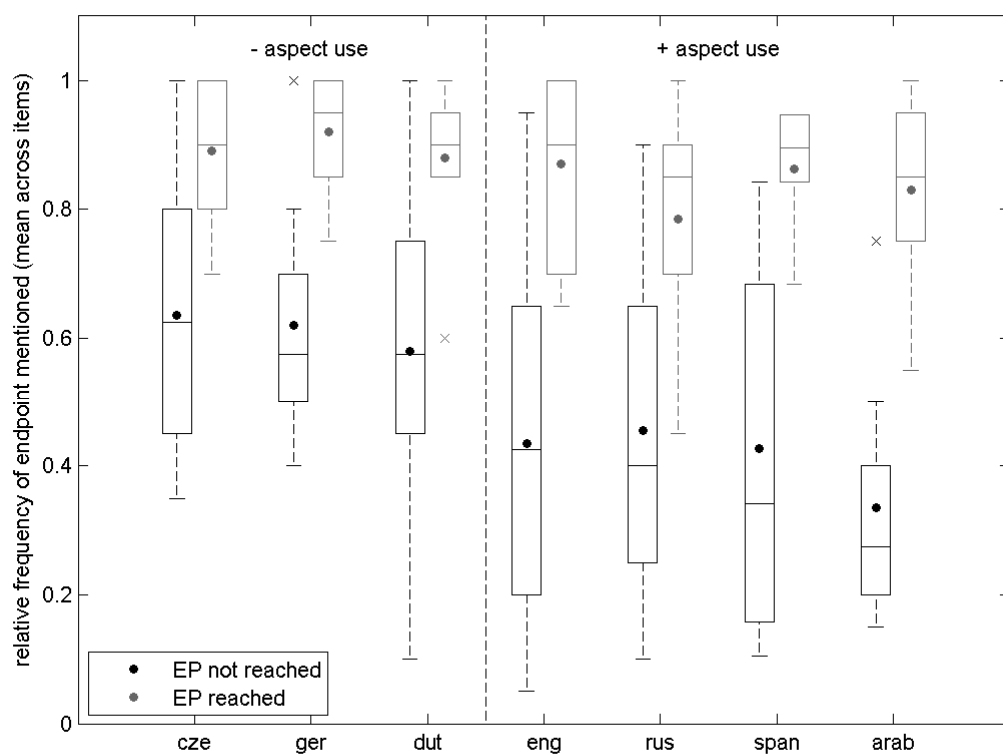
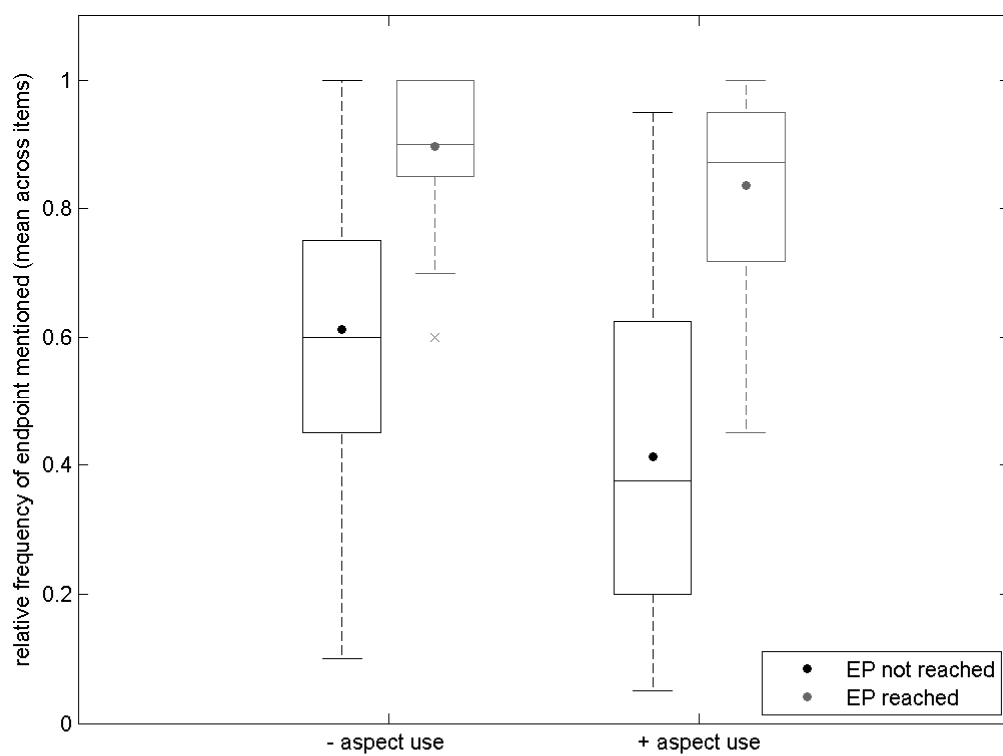


Figure 3. *Relative frequency of endpoints mentioned in the critical and control condition (mean indicated by small dark points) for the two language clusters and for each language*

In order to test our hypothesis with respect to language clusters, we grouped the data for the Czech, Dutch and German speakers together (the “minus aspect” group), as well as the data for the English, Russian, Spanish and Arabic speakers (the “plus aspect” group). An ANOVA testing relative endpoint frequencies *within* the “minus aspect” group (Czech, Dutch, German) revealed no differences between the individual languages ($F(2,57) = 0.187$, $p = 0.830$). The same is true for the “plus aspect” group (English, Russian, Spanish, Arabic) ($F(3,76) = 0.216$, $p = 0.885$), thus justifying our clustering of languages in the two aspect groups.

A one-way ANOVA was conducted on the relative frequency of endpoints mentioned per item, with the factors “condition” (critical vs. control) and “aspect group” (+ aspect use vs. – aspect use). The difference between conditions turned out to be significant ($F(1,136) = 116.762$, $p < .001$), as well as the difference between aspect groups ($F(1,136) = 15.508$, $p < .001$) and the interaction between the factors condition and aspect group ($F(1,136) = 4.499$, $p < .05$).

The number of endpoints mentioned was then compared separately for each condition. In the control condition, no significant difference was found between the two aspect groups ($F(1, 68) = 3.619$, $p = 0.061$). In the critical condition, this comparison was significant ($F(1, 68) = 11.890$, $p < .05$).

This means that, whereas the number of endpoints mentioned does not differ between the two aspect groups in the control condition, speakers of the “*minus aspect*” group mention significantly more endpoints than speakers of the “*plus aspect*” group in the critical condition.

3.5.2. Eye tracking analyses. The following measures were used to assess and compare the eye tracking results for allocation of attention to the endpoint region of the motion event (the relevant AoI): the total frequency of fixations within the AoI (*total fixation count*), the duration of fixations in the AoI (*total fixation duration*), and the number of first and second periods of fixation (first and second pass) in the AoI (*fixations in AoI: first and second period of fixation*). All eye tracking analyses were run with average measures across participants (F1) as well as averages over items (F2).

Total number of fixations by participants (F1 analysis)

Figure 4 below shows boxplots for the total number and total duration of fixations in the AoI, in the critical (EP not reached) and the control (EP reached) condition, for the two aspect groups (plus aspect, minus aspect), as well as for the individual languages.

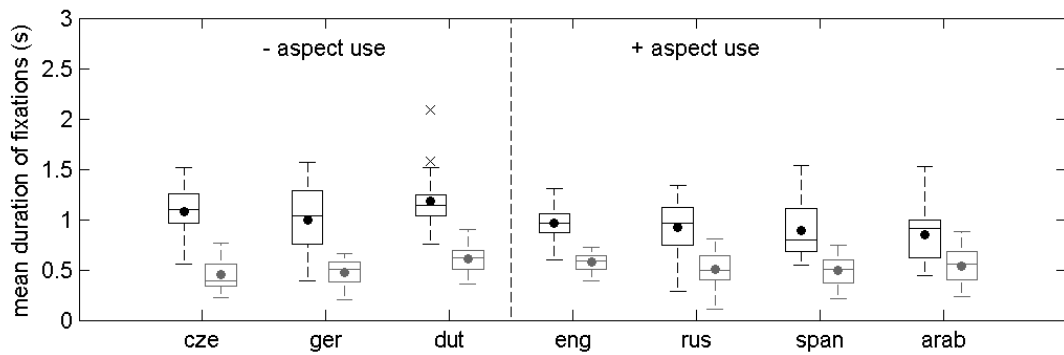
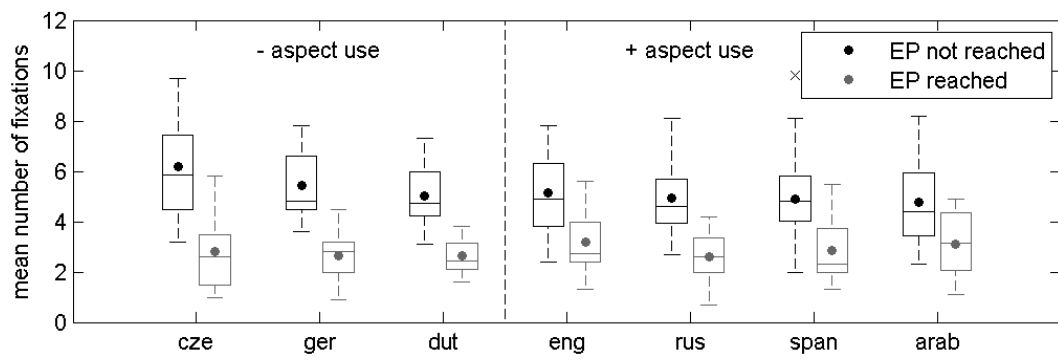
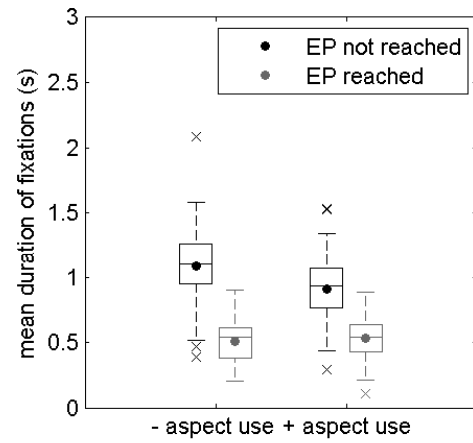
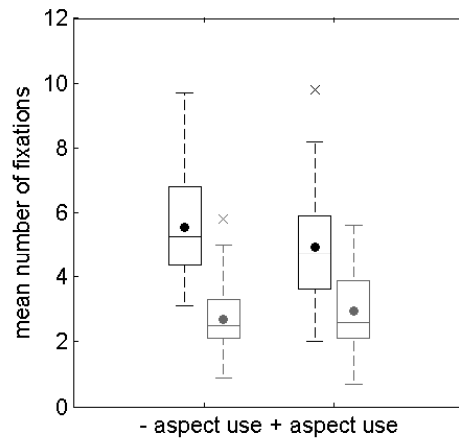


Figure 4. *Boxplots for total number and total duration of fixations in the AoI (critical and control condition)*

We first tested fixation counts *within* the two aspect groups by participants (F1), in order to justify our clustering of languages in the aspect groups, and found a difference between languages within the “minus aspect” group (Czech, German, Dutch) ($F(2,1197) = 3.101$, $p < .05$). Within the “plus aspect” group (English, Russian, Spanish, Arabic), there were no differences between fixation counts for individual languages ($F(3,1576) = 0.850$, $p = 0.466$). An F1 ANOVA was conducted on the frequency of fixations in AoI, including the factors “condition” and “aspect group”. There was a main effect for “condition” ($F(1,2776) = 314.327$, $p < .001$), but not for “aspect group” ($F(1,2776) = 1.724$, $p = 0.189$). The interaction was significant ($F(1,2776) = 9.621$, $p < .01$).

Looking at the control condition only, no significant difference between both aspect groups was found ($F(1,1388) = 2.757$, $p = 0.097$) for the frequency of fixations in the AoI. In contrast, the comparison between the two aspect groups for the critical condition only did reveal a significant main effect ($F(1,1388) = 6.864$, $p < .01$).

Given the fact that we found a significant difference between languages *within* one the groups in the F1 analysis, the “minus aspect” group, we also analyzed fixation counts with regard to individual languages and conditions. No significant main effect of language was found ($F(6,2766) = 1.918$, $p = 0.074$). We did find an effect of condition ($F(1,2766) = 305.091$, $p < .001$) and a significant interaction between language and condition ($F(6,2766) = 2.530$, $p < .05$). Looking at the control condition only, the factor language did not reach significance ($F(6,1383) = 1.565$, $p = 0.154$). However, in the critical condition we did find a significant main effect of language ($F(6,1383) = 2.494$, $p < .05$). Post hoc tests (Tukey-HSD) revealed a significant difference between Czech ($M = 6.17$, $SD = 4.94$) and Arabic ($M = 4.78$, $SD = 4.10$) ($p < .05$). All other post hoc comparisons were not significant.

Total number of fixations by items (F2 analysis)

ANOVAs comparing the average frequency of fixations by individual items (F2) were conducted as well, first checking for within-aspect-group differences. In contrast to the F1 analysis, no differences in fixation counts between the languages within the “minus aspect” group were found ($F(2,57) = 0.077$, $p = 0.926$), nor within the “plus aspect” group ($F(3,76) = 0.066$, $p = 0.978$). The following analyses are thus based on aspect groups, rather than individual languages.

The F2 analysis with the factors “condition” and “aspect group” showed a main effect for condition ($F(1,136) = 24.596$, $p < .001$), but not for aspect group ($F(1,136) = 0.001$, $p = 0.971$). The interaction was significant ($F(1,136) = 5.407$, $p < .05$), similar to the F1 analyses. Testing the control condition separately, no differences between aspect groups were found ($F(1,68) = 3.547$, $p = 0.064$). The comparison between aspect groups in the critical condition only did not reveal a significant difference in the F2 analysis ($F(1,68) = 2.211$, $p = 0.142$). This contrasts with the F1 analyses where averages across participants were analysed.

In summary, in the critical condition we found a difference in the total number of fixations on the endpoint region between the “minus aspect” and the “plus aspect” group in the analysis carried out across participants (F1). Speakers of languages within the “minus aspect” group (Czech, German, Dutch) show a higher number of fixations on the endpoint region, when compared to speakers of the “plus aspect” group.

Total fixation duration by participants (F1 analysis)

First, we tested fixation durations between languages *within* the “minus aspect” group by participants (F1), again to check for the validity of our clustering of languages within aspect groups; this showed a main effect of language ($F(2,1197) = 5.995$, $p < .01$). The second part

of the analyses below thus includes “language” as a factor for the F1 analyses. Within the “plus aspect” group, there were no differences between individual languages ($F(3,1576) = 1.129, p = 0.336$).

In a second step we compared the total duration of all fixations in the AoI between the “plus” and “minus” aspect groups and conditions. An ANOVA (F1) revealed a significant main effect of aspect group ($F(1,2776) = 10.558, p < .01$), condition ($F(1,2776) = 373.271, p < .001$) and a significant interaction ($F(1,2766) = 16.156, p < .001$).

For the control condition only, no effect of aspect group was found ($F(1,1388) = 0.506, p = 0.477$). In the critical condition, however, there was a significant difference between aspect groups ($F(1,1388) = 18.691, p < .001$).

Further, an F1 ANOVA comparing individual languages showed a main effect of language ($F(6,2766) = 4.766, p < .001$), condition ($F(1,2766) = 359.562, p < .001$) as well as a significant interaction effect ($F(6,2766) = 3.102, p < .01$). Post hoc tests (Tukey-HSD) showed a significant difference in fixation duration between Russian ($M = 0.714, SD = 0.764$) and Dutch ($M = 0.899, SD = 0.747$) ($p < .01$), Arabic ($M = 0.693, SD = 0.641$) and Dutch ($p < .001$) and Spanish ($M = 0.695, SD = 0.626$) and Dutch ($p < .001$). An effect of language was also observed within the “minus aspect” group between German ($M = 0.735, SD = 0.661$) and Dutch ($p < .05$).

A further analysis looked at the control condition *only* and again found a significant effect of language ($F(6,1383) = 2.742, p < .05$). Post hoc tests (Tukey-HSD) showed that there was a difference between Dutch ($M = 0.612, SD = 0.531$) and Czech ($M = 0.452, SD = 0.406$) ($p < .05$). In the critical condition, there was a main effect of language ($F(6,1383) = 4.426, p < .001$), and post hoc comparisons (Tukey-HSD) show differences between Dutch ($M = 1.180, SD = 0.822$) and Russian ($M = 0.920, SD = 0.821$) ($p < .05$), Dutch and Arabic ($M = 0.852, SD = 0.753$) ($p < .001$) and Dutch and Spanish ($M = 0.893, SD = 0.717$) ($p < .01$).

Total fixation duration by items (F2 analysis)

Analyses comparing average fixation duration by items (F2) were run as well. An ANOVA (F2) comparing fixation duration between languages *within* the “minus aspect” group was not significant ($F(2,57) = 0.525, p = 0.595$). Within the “plus aspect” group, there were also no differences between individual languages ($F(3,76) = 0.188, p = 0.904$), similar to the F1 analyses above. The subsequent analyses thus compared fixation duration between the two aspect groups, and not between individual languages.

In a next step, we compared the duration of all fixations in the AoI between aspect groups and conditions. The F2 ANOVA showed a significant main effect for condition ($F(2,136) = 29.207, p < .05$), but not for aspect group ($F(2,136) = 1.200, p = 0.275$), in contrast to the F1 analysis. There was also a significant interaction between aspect group and condition ($F(2,136) = 8.925, p < .01$). When analysing the control condition separately, no effect of aspect group was found ($F(2,68) = 2.322, p = 0.132$). In the critical condition, however, there was a significant difference between aspect groups ($F(2,68) = 6.781, p < .05$), as for the F1 analyses.

To summarize, we found a longer fixation duration on the endpoint region for speakers in the “minus aspect” group (Czech, German, Dutch), when compared to speakers in the “plus aspect” group (Spanish, Russian, Arabic, English), in both the F1 and F2 analyses. In the analysis by participants (F1), we found a difference within the “minus aspect” group: Dutch speakers look longer at the endpoint not only when compared to all languages in the “plus aspect” group, but also when compared to Czech and German speakers.

Fixations in the AoI: First and second period of fixation

Figure 5 below shows box plots depicting the mean number of first and second periods of fixation across items, per aspect group and language, for each condition.

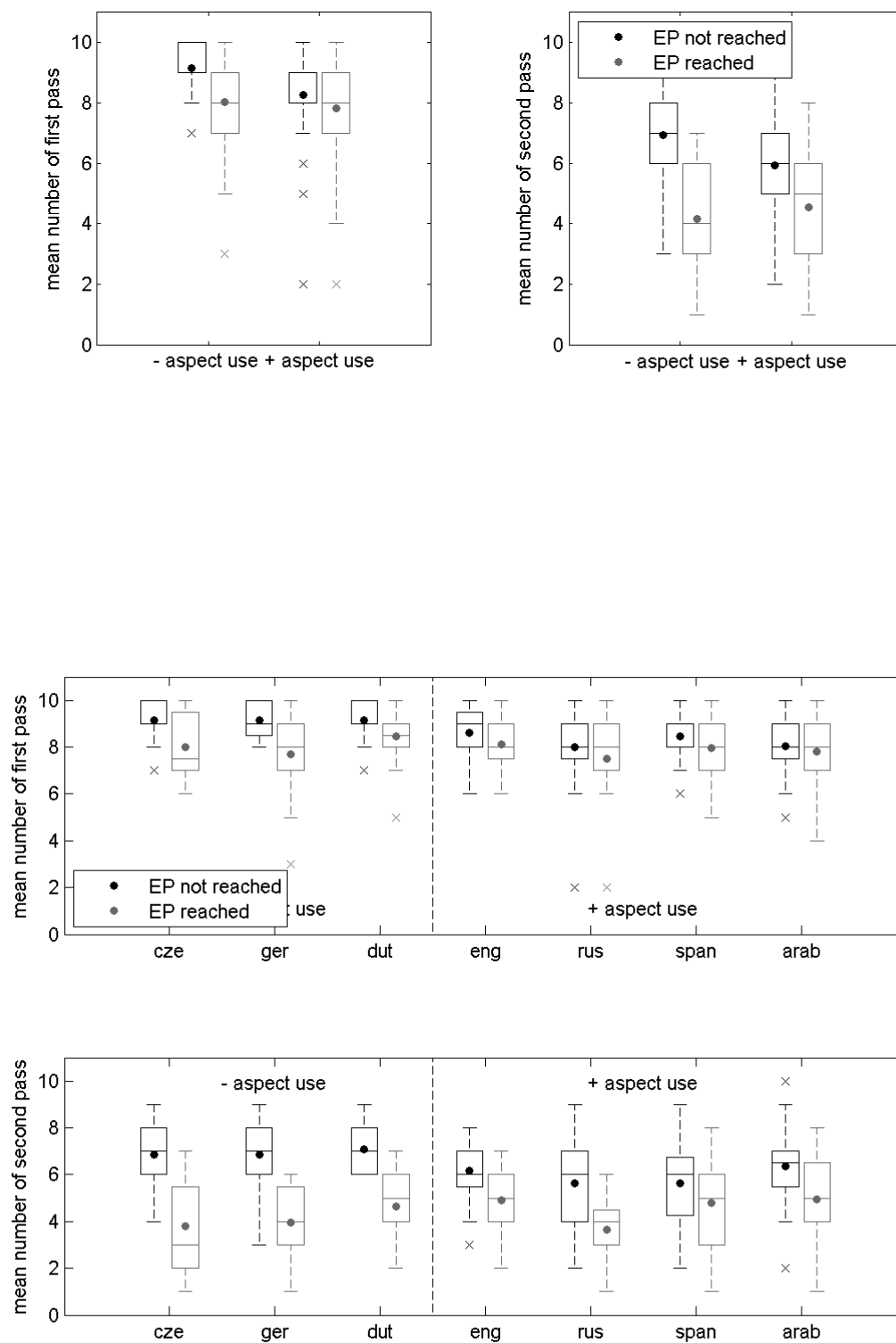


Figure 5. *Box plots with mean number of first and second passes per aspect group/language for each condition*

Number of first periods of fixation by participants (F1)

First, we ran a number of analyses comparing the average number of first periods of fixation (first passes) by participants (F1). We analyzed the average number of first passes *within* the “minus aspect” group (Czech, Dutch, German) and found no effect of language ($F(2,117) = 0.802, p = 0.451$). Similarly, there were no differences between languages in the “plus aspect” group ($F(3,154) = 1.370, p = 0.254$). All subsequent analyses thus compared aspect groups, and not individual languages.

An F1 ANOVA comparing the average number of first passes between aspect groups and conditions was conducted, and a significant main effect of aspect group ($F(1,274) = 10.907, p < .01$), condition ($F(1,274) = 22.016, p < .001$), as well as a significant interaction effect was found ($F(1,274) = 3.991, p < .05$).

For the control condition only, there was no difference between aspect groups ($F(1,137) = 0.718, p = 0.398$). In the critical condition, on the other hand, the mean number of passes did differ between the two aspect groups ($F(1,137) = 17.265, p < .001$).

Number of first periods of fixation by items (F2)

Analyses comparing the average number of first passes by items (F2) first checked for within-aspect-group differences, and found no difference between languages in the “minus aspect” group ($F(2,54) = 0.350, p = 0.706$), nor within the “plus aspect” group ($F(3,72) = 0.361, p = 0.781$).

An ANOVA by items (F2) compared the number of first passes between aspect groups and conditions. A significant main effect of aspect group ($F(1,136) = 4.725, p < .05$) and condition ($F(1,136) = 6.502, p < .05$) was found. The interaction between the two factors was not significant ($F(1,136) = 1.274, p = 0.261$), in contrast to the F1 analysis.

In the separate analyses for the two conditions we found no effect of aspect group for the control condition ($F(1,68) = 0.467, p = 0.497$). There was a significant difference between aspect groups in the critical condition ($F(1,136) = 6.651, p < .05$), similar to the results of the F1 analysis.

Number of second periods of fixation by participants (F1)

The average number of second periods of fixation by participants (F1) were first tested for potential differences *within* the two aspect groups. No effect of language was observed, neither within the “minus aspect” group ($F(2,117) = 0.773, p = 0.464$) nor the “plus aspect” group ($F(3,154) = 2.319, p = 0.078$). The following analysis thus included “aspect group” and not “language” as a factor.

An ANOVA regarding average number of second passes showed a main effect of condition ($F(1,274) = 103.598, p < .001$) and a significant interaction between the factors aspect group and condition ($F(1,274) = 11.961, p < .01$).

In the control condition there were again no differences between aspect groups ($F(1,137) = 2.057, p = 0.154$), whereas the number of second passes did differ between the aspect groups in the critical condition ($F(1,137) = 12.722, p < .001$).

Number of second periods of fixation by items (F2)

Finally, an analysis was run comparing the average number of second passes by items (F2) between languages *within* the “minus aspect” group, showing no difference ($F(2,57) = 0.212, p = 0.810$). The same finding holds for the “plus aspect” group ($F(3,76) = 0.655, p = 0.582$).

A subsequent ANOVA (F2) by items analysed the average number of second passes, for the factors “aspect group” and “condition”. There was a main effect of condition ($F(1,136) = 23.446, p < .001$), but not of aspect group ($F(1,136) = 0.624, p = 0.431$). The

interaction was not significant either ($F(1,136) = 2.796, p = 0.097$), in contrast to the F1 analysis. An analysis of the control condition only revealed no difference between aspect groups ($F(1,68) = 0.409, p = 0.525$). The analysis of the critical condition did not turn out to be significant either ($F(1,68) = 2.894, p = 0.094$), which differs from the F1 analysis.

In summary, we found a higher number of first passes for speakers *within* the “minus aspect” group compared to those in the “plus aspect” group in the critical condition (endpoint not reached), but not in the control condition. Findings for the second pass show a higher number of second passes for the “minus aspect” group in the F1 analysis by participants, but not in the analysis by items.

3.5.3. Summary and discussion of linguistic and eye tracking analyses. The results of the language production and eye tracking analyses across the different aspect groups show a relatively systematic pattern along a continuum between highly endpoint-oriented, on the one hand, and less endpoint-oriented languages on the other. The distribution of the language-specific results is broadly consistent across different measures: Speakers of Czech, Dutch and German (“minus aspect” group) have more and longer fixations on the endpoint region in the critical condition (endpoint not reached), compared to speakers of English, Spanish, Russian and Arabic (“plus aspect” group). Overall, we find a higher degree of attention to endpoints in the critical than in the control condition. This underlines the status of the endpoint as a relevant event segment in this condition: the extraction of information on reaching the endpoint or not in the critical video clips cannot be deduced “at a glance”. In the control condition, the reaching of the endpoint becomes evident relatively early during the time span in which the scene unfolds, which reduces the necessity for further fixations.

As expected, no differences were observed between aspect groups in the control condition, with the exception of Dutch in analyses by participants (F1), where fixations on the endpoint are longer, compared to German (overall) and Czech (in the control condition). Czech, as hypothesized, patterns with the languages that mention endpoints to a large extent (“minus aspect” group).

With regard to differences within the two groups of languages, and the “plus aspect” group in particular, the findings may be attributable both to the complexity of the aspectual systems as well as the extent to which the available grammatical means select the possible sub-phases of the event (inchoative, intermediate, or terminative phase). In the present task speakers may tend to focus mainly on the intermediate phase of the event (Arabic, imperfective aspect) or to sometimes combine the intermediate plus terminative phase (English, progressive aspect), (see Section 4.1 below). The variance observed for measurements across items (F2) as opposed to participants (F1) will have to be tested further with regard to the nature of the stimuli (naturalistic, dynamic video clips).

In summary, the findings correspond on the whole to the central hypothesis: if the endpoint of a motion event is not reached in the stimulus, the direction of attention to endpoints, as well as the extent to which they are encoded, varies depending on the use of grammaticized means to express aspectual distinctions and associated patterns of event conceptualization. The present findings illustrate the importance of comparing actual *language use* across languages, in addition to crosslinguistic categorizations based on the linguistic system.

3.6. Memory tests

3.6.1. Material. Following the production task, all participants were given a memory test covering 15 scenes, ten critical items and five fillers from the stimulus set. They were shown printed colored screen shots in which a particular part was cut out. This was the endpoint area

for the critical items (see Figure 7 below), while in the filler items a specific object was cut out. The latter items were included in order to control for general memory performance. The hypothesis predicts that speakers of endpoint-oriented languages would be better at remembering the objects present in the endpoint region, compared to speakers of the aspect languages.



Figure 7. *Stimulus memory test: screenshot with the village cut out as potential goal of the motion event*

3.6.2. *Procedure.* About five minutes after the production experiments subjects were told that they would be shown 15 screen shots of scenes in which a particular part of the screen shot was cut out. They were asked to write down the object(s) missing in the picture, in one or only a few words. No strict time limitation was given, but speakers were instructed to take only a few seconds per item, and to not reconsider their responses.

3.6.3. *Results.* Figure 8 below depicts box plots with the relative frequency of remembered endpoints for critical items only, plotted per aspect group (“plus aspect” and “minus aspect”), as well as for all seven languages separately.

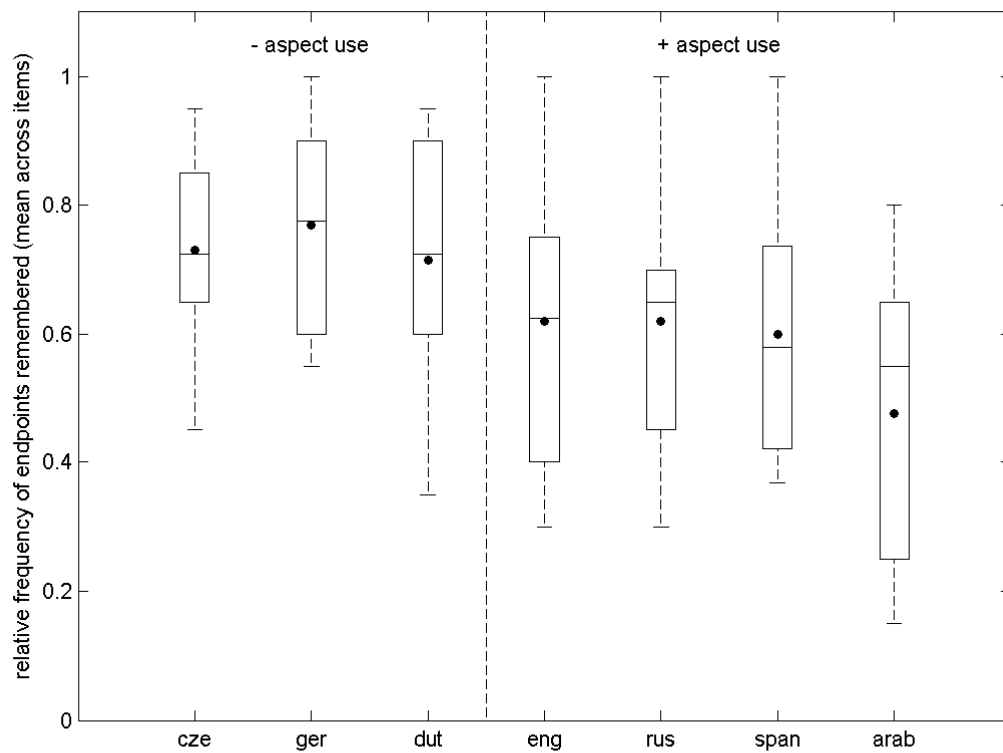
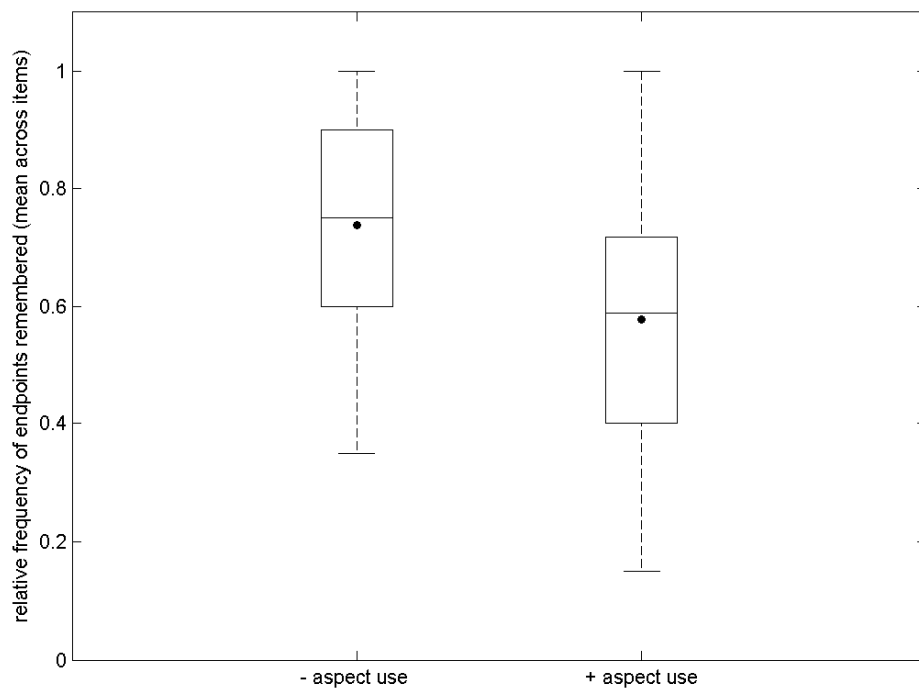


Figure 8.

Relative frequency of endpoints remembered for the critical items

The relative frequency of endpoints remembered for the control items was analysed in order to control for general memory effects. No difference was found between aspect groups ($F(1,33) = 0.327, p = 0.571$), nor between individual languages ($F(6,28) = 0.218, p = 0.968$).

An ANOVA comparing relative endpoint frequencies between aspect groups for the critical condition only did show a significant result ($F(1,68) = 10.587, p < .01$). A significant difference was found for the analysis of all individual languages ($F(6,63) = 2.381, p < .05$). Post hoc tests (Tukey-HSD) showed a difference between Arabic ($M = 0.48, SD = 0.24$) and German ($M = 0.77, SD = 0.16$) ($p < .05$). However, all other pairwise comparisons did not reveal significant differences.

3.6.4. Summary and discussion of the memory test analyses. The results of the memory experiment underline the interrelation between linguistic structure and patterns of attention in visual and cognitive processing. Speakers of languages who look less at endpoints, and who talk less about them, do not store information on endpoints to the same extent as speakers of languages who are more attentive to endpoints during the linguistic task. Speakers of “minus aspect” languages (Czech, Dutch, German), with a higher mention of endpoints than the speakers of the “plus aspect” languages, show a higher level of attention to endpoints in the critical condition. Following the performance of the verbal task, speakers of the first group also remember more endpoints. In this case, the “strongest” representative of the endpoint-oriented group is German, which differs significantly from the other end of the spectrum, Arabic, in the frequency with which endpoints are remembered. It is important to bear in mind the nature of this memory task, which was performed *after* verbal encoding: it remains an open question whether endpoints are remembered better by these speakers because of verbalization, or whether working memory might be influenced by language-specific event conceptualization patterns.

4. Interpretation, discussion, and theoretical implications

4.1. Summary and interpretation of results

The present study set out to test the *thinking for speaking* hypothesis for the domain of event construal. On a general level this hypothesis claims that “each language is a subjective orientation to the world of human experience, and this orientation affects the ways in which we think while we are speaking” (Slobin 1996: 91). In order to test this hypothesis on an empirical basis, specific linguistic domains that differ crosslinguistically have to be identified, so that speakers who use the different systems can be compared during cognitive processing in language production or comprehension.

In contrast to earlier studies in this field, which mainly concentrated on language-specific patterns of lexicalization (e.g., manner versus path verbs), the present study deals with the linguistic domain of *verbal aspect* and its role in event conceptualization. It focuses in particular on the role of *grammaticized* concepts when *thinking for speaking*, an area of investigation which has not been approached so far in this specific form. The underlying hypothesis relates to temporal properties of situations, their selective and perspective-driven linguistic representation, and takes into account the role of *grammatical aspect* when talking about events. The function of grammatical aspect can be described as allowing phasal decomposition of a dynamic situation, with phases which can be described as inchoative, intermediate, terminative, resultative, etc. In other words, languages that provide aspect marking on their verbal means, the core expressions used when talking about events, require the speaker to take aspect into account and to decide which phase of a situation is relevant (with reference to the overall situation being one option among others). Speakers of languages

without grammatical aspect do not have to make a choice of this kind and present the event on a holistic basis. This does not mean that speakers of non-aspect languages cannot conceptualize or express other phases of the event: Phasal decomposition occurs, but only in restricted contexts. Ongoing phases of an event can be explicitly encoded by means of lexical expressions, which are optional.

In order to test the possible conceptual implications of this formal difference while thinking for speaking, the study was designed with speakers of languages that differ with respect to the relevant formal criterion (+/- grammatical aspect). Speakers were exposed to the same visual input in which motion events were presented in the form of short video clips and varied with respect to their goal orientation: In the critical items, the phase of the event in which the endpoint was reached was not shown, but could be inferred from the video on the basis of the trajectory drawn by the entity in motion, going towards a potential endpoint-object. In the control items, by contrast, the endpoint-object was actually reached.

The present study of the role of language specificity for cognitive processing was carried out across seven languages and consisted of a threefold experiment, encompassing language production, visual attention (eye tracking) and a memory test. The experiment was based on the hypothesis that speakers of an aspect language would provide information on the trajectory shown in the clip, when asked to tell *what is happening*, thus relating to the phase focused in the clip, the intermediate phase of the event at issue. Since the endpoint will not be relevant in planning what to say, given the nature of the task, speakers should direct less visual attention to the endpoint-object in the stimulus. The speakers of the non-aspect languages, on the other hand, will represent the motion event holistically, thereby including the object that functions as a possible endpoint, although it was not actually reached by the entity in motion in the clip. The object-endpoints should therefore attract more attention while processing the clip, compared to speakers of aspect languages.

The results show that in accordance with the features of the verbal system at their disposal, speakers differ both in the choice of temporal perspective (ongoing or not explicitly ongoing, as marked on the verb), the segment of the route selected for verbalization, as well as in the distribution of attention when processing the stimuli and preparing for verbalization. This was revealed by eye movement patterns, by the information verbalized with respect to phases and endpoints, as well as performance on memory tests following the linguistic task. The two language groups (+/-aspect) correlate across all three components with regard to the cognitive relevance of endpoints. However, the data do not present a black-and-white picture of overwhelming significance. This can be attributed on the whole to two main factors.

- a) As outlined in Section 3.2, aspect languages are by no means identical. Although they all feature grammaticized distinctions allowing specific temporal view points on a situation, the semantic oppositions, however, are formed on the basis of categories which differ in number and type. In the present case there are differences with regard to the possible integration of endpoints for situations which are aspectually marked. English, for example, allows for the integration of different phases — *situation in progress* and *endpoint* — in the representation of a motion event, as reflected in the data. In cases where the endpoint is not reached in the clips (critical items), the number of endpoints mentioned is relatively low (lower than the “minus aspect” group, see Figure 3), but visual attention, however, is directed to the potential endpoint to some extent (see Figure 4). As results of an earlier eye tracking study on German and English have shown, English speakers look at the endpoints at a *later* point, compared to speakers of German (cf. von Stutterheim and Carroll 2006). In other words, it is necessary to also take into account the time course of visual attention as an additional window on conceptual processing. While German speakers tend to direct attention to the endpoint region before speech onset, English speakers do so in

the course of utterance articulation. One phase can be conjoined with another so speakers do not have to first scan the scene for an endpoint in order to arrive at an adequate conceptual representation for an event marked aspectually as in progression (progressive *-ing*). Significantly, endpoints are more likely to be mentioned, compared to Arabic, where the aspectual contrast given with a full fledged imperfective-perfective system lowers the scope for including an endpoint for events that are marked with the imperfective. This brief example illustrates how specific features of the respective aspectual system have to be taken into account in explaining the complex results of the study at hand.

- b) A second important factor which has to be considered when relating linguistic structure to conceptual processing lies in the potential discrepancy between a linguistic system as an inventory of forms, and the way these forms are actually used in context. In line with the view in cognitive linguistics (cf. Talmy 1988; Bybee et al. 1994; Slobin 1997b; Bowerman and Choi 2003), the hypotheses outlined above are based on the assumption that concepts marked by grammatical means play a fundamental role in structuring information for expression. Different grammatical systems profile different concepts, so that structural features of this kind will serve in facilitating planning processes at the level of the conceptualizer in language production. Therefore, a study on factors that drive *thinking for speaking*, and takes typological differences as a starting point for hypothesis generation, must relate to analyses on the actual use of grammatical forms, as well as their function in context. The present study profiles differences in aspectual systems within the Slavic language group, as a typological unit, showing how Czech and Russian speakers belong to two different groups with regard to *thinking for speaking* as well as *seeing for speaking*, since use of verbal means differs markedly.

4.2. Discussion of determining factors

The central claim in this paper relates to language-driven patterns of event construal, in particular the salience of goal points in motion events for processes of visual and cognitive attention. Over the last 20 years, a large number of studies have addressed questions of attention allocation in visual processing (see for example the overviews in Huettig et al., 2011; Henderson and Ferreira 2004) showing how perception processes are selective. In investigating the nature and function of the filters involved, certain features of a percept have been shown to attract attention, irrespective of individual factors. Such factors relate to color, size, shape, motion, curvature, luster (cf. e.g., Wolfe and Horowitz 2004). Although these studies do not systematically test a broad range of different cultures and languages, there is ample evidence that human beings react alike when confronted with stimuli showing certain types of features (cf. overview in Duchowski 2007). The present study revealed parallel responses across speakers of all languages for the control items. These scenes show the highest level of goal orientation possible – an entity in motion that actually reaches an endpoint. Since speakers are asked to ‘verbalize what is happening in the clip’ speakers of all languages can be expected to refer to the endpoint reached. This explains the homogeneous results for the control items across the different languages. However, the fixation count and duration analyses reveal lower levels in directing attention to endpoints in the control condition, when compared to the critical condition. This can be interpreted as an artifact of the stimuli: during the time in which the scene unfolds, the relevance of the endpoint becomes clear, i.e., it is relatively easy to identify that one is dealing with a case where ‘an endpoint will be reached’, and the type of object that represents the endpoint can also be identified at a glance. This, we speculate, does not require subsequent fixations. Future studies should

include an analysis of the time course of fixations in order to provide detailed explanations of the fixation patterns found. The findings for the control items contrast with those for the critical items, since they show higher range of variation in the distribution of attention.

In addition to the complexity given with clips depicting real-life motion events (most studies use pictures of objects as stimuli, cf. Meyer et al. 1998),¹¹ cultural differences may also lead to variation in the responses given (Boduroglu et al. 2009). The languages included in the set were selected so as to account for this factor as a possible variable. English and German, for instance, differ structurally in the domain of grammatical aspect, but belong to a similar set with regard to cultural affiliation. English and Arabic show a certain degree of overlap with regard to grammatical aspect. Their speakers, however, have different cultural traditions. Languages that converge for the linguistic phenomenon at issue form a cluster in the present study (+/- active use of progressive/imperfective aspectual markers), but they do not share similar cultural backgrounds (Arabic, Spanish, English and Russian belong to one group, while Russian and Czech do not, for example).

In language production conceptual categories which are grammaticized in a language play a constitutive role in shaping the filter for selective attention and conceptualization. We can draw the conclusion that principles of attention allocation during information intake are not reducible to one type of factor. Rather, they result from an interplay of factors, the relative weight of which depends on the respective context. The hypothesis that linguistic structure, deeply rooted in the knowledge system of every human being, profiles cognitive processes can be supported by the fact that language production makes high demands on cognitive processing. Given the large range of possible options in representing a given state of affairs, the processing options profiled by the system may support highly automatized routines that then serve, on a default basis, in language production. In this sense profiled concepts help deal with the complexity and speed of delivery given in language processing. This is not to say that they cannot be overruled where required by specific conditions of the task. Profiled options and associated perspectives are not insurmountable pathways in decision making in language production, but they are very convenient ones. As the results presented indicate, systematic studies can shed light on the extent to which language-driven patterns of attention may affect performance when carrying out different tasks. In the present analysis we have shown that when dealing with the construal of complex situations such as events, there are certain systematic patterns in *thinking and seeing for speaking* that can be identified at different levels: attention and gaze movement, information selection in deciding what to say, and memory after performing the linguistic task.

4.3. Theoretical implications

No one would seriously question the statement that cognitive processing is based, on the one hand, on human abilities that are available to all humans and have played a part in the evolution of our species, and on the other hand on human abilities based on cultural factors, such as the development of a specific language. One of the main issues in research on language and cognition centers on the methods required in investigating the interplay of these different factors in cognitive and language processing. We need fine-grained experimental studies, embedded in an interdisciplinary approach that takes into account relevant findings in language processing, as well as visual and auditory processing in neuroscience and cognitive psychology. Progress in the technical development of research tools over the last few decades provides the basis in taking a course where each step can contribute to our understanding of the interrelation between thinking and speaking in cognitive processing.

So where do the present results obtained in our study fit into current theoretical discussions on cognitive processing? The data reveal certain language effects on how speakers view a visual scene when thinking for speaking. We have claimed that conceptual

categories encoded in a grammatical system play an active role in the cognitive filter set up in processes of attention allocation and event construal when talking about events. The processes involved in deciding what to say can be modeled as hierarchically structured, where the higher level is given by the representation of linguistic knowledge and related abstract representations of event types, and a lower level given by information extracted from the percept and visual input. In order to explain the effect of the higher conceptual level on processes of attention and the way they are directed to different aspects of the percept, top-down and bottom-up processes have to be interconnected, leading where necessary to a number of top-down, bottom-up loops until the final identification of an event type is formed. We can draw here on the *Reversed Hierarchy* model, for example, (Ahissar et al. 2009) which sets out to explain interrelations of this kind.

The reversed hierarchy theory is a concept that attempts to spell out necessary links between the hierarchies involved in the processes and dynamics of perception. It was initially developed for the visual modality and later extended to the auditory modality. It proposes that, by default, rapid perception is based on high-level representations alone (Ahissar et al. 2009: 286). However, if the initial identification of an object, or an event, is not adequate and further information is needed for a successful outcome, then a backward search is initiated, which Ahissar et al. have termed ‘perception with scrutiny’ (2009: 287).

These notions, which have been developed on the basis of object identification, can be investigated with respect to the extent to which they also hold for event identification. At the higher level, types of events are represented as motion events such as ‘a vehicle drive — on a road’ or ‘a person walk — into a house’. In order to produce a sentence encoding the event — depending on the language used- more detailed information is needed. The speaker will focus on details of the stimulus depending on, for example, the way the event is typically represented on a temporal-aspectual basis in the language used. Relevant processes may be uncovered by looking at the time course of attention allocation in correlation with language-specific requirements when building up a higher-level conceptual representation in language production. This is just one of the many steps required in unraveling the complex factors that drive the processes we seek to understand.

Received 11 November 2010

Revised version received

9 December 2011

University of Heidelberg

University of Heidelberg Hospital

Appendix. Stimuli used for analyses

Critical condition: *Endpoint not reached* 10 items

Video clip	Description
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 gas station)

Control condition: *Endpoint reached* 10 items

Video clip	Description
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 in 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

References

- Ahissar, Merav, Mor Nahum, Israel Nelken & Saul Hochstein. 2009. Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B* 364. 285–299.
- Billman, Dorrit, Angela Swilley & Meredyth Krych. 2000. Path and manner priming: verb production and event recognition. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.
- Boduroglu, Aysecan, Priti Shah & Richard Nisbett. 2009. Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology* 40(3). 349–360.
- Bowerman, Melissa & Soonja Choi. 2003. Space under construction: Language-specific spatial categorization in first language acquisition. In Dedre Gentner & Susan Goldin-Meadow (eds.), *Language in Mind: Advances in the study of Language and Cognition*, 387–428. Cambridge: MIT Press.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Bylund, Emanuel. 2009. Effects of age of L2 acquisition on L1 event conceptualization patterns. *Bilingualism: Language and Cognition* 12(3). 305–322.
- Carroll, Mary, Christiane von Stutterheim & Ralf Nüse. 2004. The language and thought debate: A psycholinguistic approach. In Christopher Habel & Thomas Pechmann (eds.), *Approaches to language production*, 183–218. Berlin & New York: Mouton de Gruyter.
- Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.
- Dahl, Östen (ed.). 2000. *Tense and aspect in the languages of Europe*. Berlin & New York: Mouton de Gruyter.
- Dickey, Stephen. 2000. *Parameters of Slavic aspect: A cognitive approach*. Chicago: CSLI Press.
- Dickey, Stephen. 2007. A prototype account of the development of delimitative PO- in Russian. In Dagmar Divjak & Agata Kochanska (eds.), *Cognitive paths into the Slavic domain*, 326–371. Berlin & New York: Mouton de Gruyter.
- Duchowski, Andrew. 2007. *Eye tracking methodology: Theory and practice*. London: Springer.
- Gennari, Silvia, Steven Sloman, Barbara Malt & Tecumseh Fitch. 2002. Motion events in language and cognition, *Cognition* 83. 49–79.
- Griffin, Zenzi & Kathryn Bock. 2000. What the eyes say about speaking. *Psychological Science* 11. 274–279.
- Henderson, John & Andrew Hollingworth. 1998. Eye movements during scene viewing: an overview. In Geoffrey Underwood (ed.), *Eye guidance in reading and scene perception*, 269–293. Oxford: Elsevier Science.
- Henderson, John & Fernanda Ferreira. 2004. *The interface of language, vision and action: eye movements and the visual world*. New York: Psychology Press.
- Huettig, Falk, Joost Rommers & Antje Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137. 151–171.
- Klein, Wolfgang & Ping Li (eds.). 2009. *The expression of time*. Berlin & New York: Mouton de Gruyter.
- Levelt, Willem. 1999. Producing spoken language: a blueprint of the speaker. In Peter Hagoort & Colin Brown (eds.), *The neurocognition of language*, 94–122. Oxford: Oxford University Press.

- Marian, Viorica & Caitlin Fausey. 2006. Language-dependent memory in bilingual learning. *Applied Cognitive Psychology* 20(8). 1025–1047.
- Mack, Arien & Irvin Rock. 1998. *Inattention blindness*. Cambridge, MA: MIT Press.
- Meyer, Antje, Astrid Sleiderink & Willem Levelt. 1998. Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66. 25–33.
- Meyer, Antje & Femke van der Meulen. 2000. Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review* 7. 314–319.
- Papafragou, Anna, Justin Hulbert & John Trueswell. 2008. Does language guide event perception? Evidence from eye movements. *Cognition* 108(1). 155–184.
- Papafragou, Anna, Christine Massey & Lila Gleitman. 2002. Shake, rattle, ‘n’ roll: The representation of motion in language and cognition. *Cognition* 84. 189–219.
- Pavlenko, Aneta. 2003. Eyewitness memory in late bilinguals: Evidence for discursive relativity. *The International Journal of Bilingualism* 7(3). 257–281.
- Sasse, Hans-Jürgen. 2002. Recent activity in the theory of aspect: Accomplishment, achievement or simply non-progressive state? *Linguistic Typology* 6(2). 199–271.
- Schmiedtová, Barbara. 2011. Do L2 speakers think in the L1 when speaking in the L2? *VIAL, International Journal of Applied Linguistics* 8. 138–179.
- Schmiedtová, Barbara & Natascha Sahonenko. 2008. Die Rolle des grammatischen Aspekts in Ereignis-Enkodierung: Ein Vergleich zwischen Tschechischen und Russischen Lernern des Deutschen. In Patrick Gommès & Maik Walter (eds.), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbforschung*, 45–71. Tübingen: Max-Niemeyer Verlag.
- Simons, Daniel & Daniel Levin. 1997. Change Blindness. *Trends in Cognitive Science* 1. 261–267.
- Simons, Daniel. 2000. Current approaches to change blindness. *Visual Cognition* 7. 1–16.
- Slobin, Dan. 1996. From “thought to language” to “thinking for speaking”. In John Gumperz & Stephen Levinson (eds.), *Rethinking linguistic relativity*, 70–96. Cambridge: Cambridge University Press.
- Slobin, Dan. 1997a. Mind, code, and text. In Joan Bybee, John Haiman & Sandra Thompson (eds.), *Essays on language function and language type: Dedicated to T. Givón*, 437–467. Amsterdam & Philadelphia: John Benjamins.
- Slobin, Dan. 1997b. The origins of grammaticizable notions: Beyond the individual mind. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition: Vol. 5. Expanding the contexts*, 265–323. Mahwah, NJ: Lawrence Erlbaum.
- Slobin, Dan. 2000. Verbalized events: A dynamic approach to linguistic relativity and determinism. In Susanne Niemeier & Rene Dirven (eds.), *Evidence for linguistic relativity*, 107–138. Amsterdam & Philadelphia: John Benjamins.
- Smith, Carlota. 1991. *The Parameter of aspect*. Dordrecht: Kluwer.
- Soroli, Efstathia, & Maya Hickmann. 2010. Language and spatial representations in French and in English: Evidence from eye-movements. In Giovanna Marotta, Alessandro Lenci, Linda Meini & Francesco Rovai (eds.), *Space in language*, 581–597. Pisa: Editrice Testi Scientifici.
- von Stutterheim, Christiane, Mary Carroll & Wolfgang Klein. 2009. New perspectives in analyzing aspectual distinctions across languages. In Wolfgang Klein & Ping Li (eds.), *The expression of time*, 195–216. Berlin & New York: Mouton de Gruyter.

- von Stutterheim, Christiane & Mary Carroll. 2006. The impact of grammaticalised temporal categories on ultimate attainment in advanced L2-acquisition. In Heidi Byrnes (ed.), *Educating for advanced foreign language capacities: constructs, curriculum, instruction, assessment*, 40–53. Washington D.C.: Georgetown University Press.
- Talmy, Leonard. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, vol. III, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1988. The relation of grammar to cognition. In Brygida Rudzka-Ostyn (ed.), *Topics in cognitive linguistics*, 165–205. Amsterdam & New York: John Benjamins.
- Wolfe, Jeremy & Todd Horowitz. 2004. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience* 5. 1–7.

Notes

-
- ¹ We would like to thank the DFG (German Research Foundation) for financial support. We would also like to thank two anonymous reviewers for helpful comments on earlier drafts of this paper. Correspondence address: Christiane von Stutterheim, Institut für Deutsch als Fremdsprachenphilologie, Universität Heidelberg, Plöck 55, 69117 Heidelberg, Germany. E-mail: stutterheim@idf.uni-heidelberg.de.
- ² Although use of the terms *imperfective/ progressive* is inconsistent in denoting the viewpoint ‘event is ongoing’ (cf. Dahl 1985), we draw here on Dahl’s classification as a point of reference since this is the most comprehensive crosslinguistic study on this topic.
- ³ By *fully grammaticized* we mean grammatical markers which are not constrained in their use by lexical (e.g., verb type) or compositional (e.g., other morphological categories) restrictions.
- ⁴ Dickey classifies the perfective in Czech as a semantic subclass of verbs which express situations under a holistic perspective: the perfective represents an event in its totality (*eat up something*) but use does not entail that it is viewed as actually completed (*eaten up*); we find *retention by these languages (west Slavic languages cvs) of totality as the meaning of the perfective aspect* (cf. Dickey 2000; 2007). He argues that West and East Slavic languages differ markedly in their aspectual systems.
- ⁵ Cf. detailed discussion in Schmiedtová (2011).
- ⁶ Corresponding to level C1 and beyond in the Common European Framework (CEFR).
- ⁷ In all cases, the critical and control items involve different types of motion events, a necessary pre-condition for ensuring naturalistic and spontaneous event descriptions.
- ⁸ We are aware of the fact that the slight variance in AoI size across different individual items represents a potential cause for discrepancy in attention allocation patterns between items. However, this is a necessary concession when dealing with dynamic, live-recorded stimuli. The motivation underlying the choice for this type of stimuli is the fact that only this type of stimuli allows for crosslinguistic variation in linguistic encoding. Line-drawn or animated stimuli are less suited for obtaining language-specific preferences in event encoding. In *both* conditions the AoI always involves one specific object (e.g., building, car, door of a building) which may differ slightly in size between individual items. In our view, a comparison between the two conditions is justified because of this precondition.
- ⁹ The first and second pass are defined as the first or second time *after stimulus onset* that speakers spent a period fixating points within the AoI. Use of the measure ‘pass’ stems from analyses of gaze patterns when reading (as in, for example, Henderson and Hollingworth 1998).
- ¹⁰ Trials with total track loss amounted to approximately 20% and were excluded from the statistical analyses of the eye tracking data.
- ¹¹ As the following quote from a study on visual perception and cognitive processing illustrates, we have to expect substantially different results if we move from analyzing highly restricted cognitive processes using simple stimuli (used in most of the relevant studies) to cognitive processing elicited by complex stimuli. “Traditionally, psychophysics embraced the assessment procedures that used consistent (simple cvs) stimuli across trials (‘blocks’), partly due to the resulting better thresholds. In RHT terminology, this protocol allows the evaluation of low-level thresholds. The complementary aspect of this approach is that these fine thresholds,

measured under ‘blocked’ protocols, do not characterize the information available for perception in natural contexts. These, according to RHT, depend on high-level resolution and are therefore limited, for both simple and complex stimuli, as described above” (Ahissar et al. 2009: 289).