COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 37 (2013) 286–309

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12006

Does Grammatical Aspect Affect Motion Event Cognition? A Cross-Linguistic Comparison of English and Swedish Speakers

Panos Athanasopoulos,[a] Emanuel Bylund[b]

[a]*School of Education, Communication, and Language Sciences, Newcastle University*
[b]*Centre for Research on Bilingualism, Stockholm University*

Received 18 February 2011; received in revised form 11 May 2012; accepted 11 May 2012

Abstract

In this article, we explore whether cross-linguistic differences in grammatical aspect encoding may give rise to differences in memory and cognition. We compared native speakers of two languages that encode aspect differently (English and Swedish) in four tasks that examined verbal descriptions of stimuli, online triads matching, and memory-based triads matching with and without verbal interference. Results showed between-group differences in verbal descriptions and in memory-based triads matching. However, no differences were found in online triads matching and in memory-based triads matching with verbal interference. These findings need to be interpreted in the context of the overall pattern of performance, which indicated that both groups based their similarity judgments on common perceptual characteristics of motion events. These results show for the first time a cross-linguistic difference in memory as a function of differences in grammatical aspect encoding, but they also contribute to the emerging view that language fine tunes rather than shapes perceptual processes that are likely to be universal and unchanging.

Keywords: Grammatical aspect; Endpoint encoding; Whorf; Linguistic relativity; Motion events

1. Introduction

The extent to which speakers of different languages think differently has been at the heart of scientific research and debate in the cognitive sciences since Benjamin Lee Whorf

Correspondence should be sent to Dr. Panos Athanasopoulos, School of Education, Communication, and Language Sciences, Newcastle University, King George VI Building, Queen Victoria Road, Newcastle upon Tyne, NE1 7RU, UK. E-mail: panos.athanasopoulos@ncl.ac.uk

formulated what he called the ''linguistic relativity principle.'' According to Whorf, ''users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation'' (Whorf, 1940/1956, p. 221). A great deal of empirical research to date has supported the notion that language acts as an attention-directing mechanism for the purposes of matching, similarity judgments, and categorization, at least in some domains like color (Davidoff, Davies, & Roberson, 1999; Kay & Kempton, 1984; Regier & Kay, 2009), space (Levinson, 2003; Majid, Bowerman, Kita, Haun, & Levinson, 2004), and objects and substances (Athanasopoulos, 2007; Imai & Gentner, 1997; Lucy, 1992).

In this study, we ask whether Whorfian effects may be obtained when people are thinking about motion events. Recent evidence shows that the presence of aspectual marking in a language (e.g., English, Spanish) correlates with a tendency not to mention the goal or endpoint of an event when describing goal-oriented dynamic scenes, while the absence of aspectual marking in a language (e.g., German, Swedish) correlates with the reverse tendency, that is, a bias toward mentioning the goal of actions (e.g., Bylund, 2008; Schmiedtová, von Stutterheim, & Carroll, 2011). Hitherto, these studies have shown that the structure of language affects how people verbalize events, which is a classic demonstration of what Lucy (1997) calls an effect of language on language. Although such an effect is useful for our understanding of conceptualization processes during speaking, it does not address the broader question of whether language affects thinking. Thus, a gap in the literature exists of whether these cross-linguistic differences may affect nonverbal cognitive processing of events in speakers of these languages. Here, we aim to fill that gap, by explicitly testing the relationship between linguistic structure (whether a language formally encodes grammatical aspect) and attention to the goal of an event during speaking, but also crucially, during nonverbal cognitive processing.

In its most generic conception, a motion event may be characterized as an evolvement through space and time, such as the case of a bowling ball rolling on a lane, occupying different spatial positions at different temporal intervals. Time and space are, in other words, fundamental concepts in the definition of a motion event (see Casati & Varzi, 2008). At present, relativistic research on motion events has focused on cross-linguistic differences in the encoding of path and manner of motion (Talmy, 1975, 2000), thus exploring whether linguistic categories that primarily convey *spatial properties* of a motion event affect cognition. There is, however, virtually no research examining how linguistic categories concerning the *temporal properties* of motion events may affect people's thinking. Against the background outlined in this and the previous paragraph, the specific aim of the study at hand is to establish whether the linguistic category of aspect, the basic function of which is precisely to define the temporal distribution of an event (such as the distinction between progressive and nonprogressive aspect in English, see Slobin, 2003; Talmy, 1988), may give rise to cognitive differences between speakers of different languages.

In the following paragraphs, we review the existing research on motion event cognition in adults, focusing on studies that have attempted to investigate the Whorfian question in this domain with regard to the encoding of manner and path of motion. We then focus in on the grammatical aspect framework, presenting a series of recent studies that have examined

this linguistic category from a psycholinguistic perspective. Subsequently, we formulate the specific aims and research questions of the present study.

1.1. Lexicalization patterns of path and manner of motion and the Whorfian question

According to Talmy (1975, 2000), some languages tend to emphasize manner of motion, whereas other languages highlight path of motion. Consider a scene with a person who is dancing out of a house. Speakers of Greek would typically describe this scene in the following way:

O andras	*vgainei*	*apo*	*to*	*spiti*	*(horevontas)*
The man	exits	from	the	house	(dancing)
	PATH				MANNER

Speakers of Swedish, in contrast, would describe the same scene as follows:

Mannen	*dansar*	*ut*	*ur*	*huset*
The man	dances	out	of	the house
	MANNER	PATH		

As these examples show, so-called Manner languages (e.g., Swedish, English) encode manner of motion in the main verb, while information about the path/direction is encoded in verb particles and adverbs. Path languages (e.g., Greek, Spanish, Japanese), on the other hand, encode path of motion in the main verb, and only optionally add information about manner of motion in modifiers such as gerunds. As will be shown, Talmy's (1975, 2000) typology of cross-linguistic differences in motion verb semantics has constituted an important theoretical basis for the exploration of Whorfian effects on motion event cognition.

In what is now considered a seminal study in the field, Naigles and Terrazas (1998) showed that speakers of Manner languages like English interpret novel verbs as referring to manner of motion, whereas speakers of Path languages like Spanish preferred a path interpretation, but only when syntactic frame was manipulated. Results from one experiment showed that English speakers preferred a manner interpretation when the novel verb was presented in a manner syntactic context, whereas Spanish speakers preferred a path interpretation when the novel verb was presented in a path syntactic context. The two groups showed no significant preference for either path or manner in the context that was incongruent to their language-specific preferences (see also Feist, 2010).

A series of groundbreaking studies by Papafragou and colleagues explored whether these differences in lexicalization patterns between Manner and Path languages have any consequences for nonlinguistic cognition, thus directly addressing the Sapir-Whorf hypothesis. In one study, Papafragou, Massey, and Gleitman (2002) looked at speakers of English and Greek in a recognition memory task and in a categorization task utilizing cartoon pictures adapted from the ''Frog where are you?'' story book (Mayer, 1969). Over a two-day testing session, participants were asked to describe the pictures (day 1), and then in day 2

discriminate between pictures they had seen and pictures they had not seen. The new pictures contained an alteration either in the manner or path of motion from the original pictures. As predicted by Talmy's typology, English speakers largely used manner verbs (e.g., *the frog is jumping into a room*) to describe the motion events in the pictures, whereas Greek speakers mainly used path verbs (e.g., *the frog is entering a room*). However, no differences were found between groups in the recognition memory task. That is, their performance did not vary as a function of the manner or path alternate pictures. These results were replicated in a categorization task utilizing real photographs, where participants had to match a path or manner alternate to a target picture. However, interpretation of results from this study is constrained by the fact that the stimuli used depicted static rather than dynamic scenes.

Gennari, Sloman, Malt, and Fitch (2002) studied English and Spanish speakers with respect to linguistic encoding and categorization of clips depicting dynamic motion events. In the first phase of the study (encoding phase), participants were assigned to one of three conditions: one condition where they described the scenes, one condition where they were asked to freely inspect the scenes without explicit verbalization, and one condition where they were asked to inspect the scenes, while repeating a nonsense syllable (verbal shadowing condition). Then after a brief distraction task, participants in all conditions received a recognition task, asking them to look at scenes and indicate whether they had seen the event in the encoding phase. The recognition task was followed by a similarity judgment task, where participants were asked to match a same-manner or same-path alternate video to a target video. As expected, results from verbal encoding showed that Spanish speakers used more path verbs than English speakers, whereas English speakers indicated manner of motion in the main verb more often than Spanish speakers. In the similarity judgment task, cognitive differences emerged between groups, such that Spanish speakers tended to select the same-path alternate as opposed to English speakers, who showed no strong preference for either alternate. However, this pattern was observed only for those groups that had previously verbally encoded the scenes, and not for participants in the other two conditions (free inspection and verbal shadowing). No differences between the groups were found in the recognition memory task. Thus, the authors concluded that cognitive differences between speakers of different languages emerge only in situations that force participants to explicitly verbalize the stimuli before categorization.

Yet in a more recent study (Papafragou & Selimis, 2010), differences between speakers were found even in a matching task that did not involve any overt linguistic labeling. Utilizing dynamic cartoon scenes, Greek and English speakers were asked to give a same-path or same-manner alternate match to a target scene, by indicating whether the actor is ''doing the same thing'' in one of the two alternates as in the target (experiment 1). Results showed that Greek speakers selected the same-path alternate more than English speakers, who in turn tended to select the same-manner alternate. Papafragou and Selimis (2010) attributed these differences to the instructions, which according to them contained linguistic cues that could have potentially affected responses. Indeed, when the instructions were changed, asking participants whether they ''see the same'' (experiment 2), no differences between groups were found. In a third experiment utilizing the same instructions as in experiment 2,

the target and alternate clips were played simultaneously (instead of the target and each alternate appearing sequentially as in the first two experiments), and results were similar to experiment 2.

According to Papafragou and Selimis (2010), these results support what they call the ''linguistic intrusion'' account. Language may affect motion event cognition only in the presence of a biasing linguistic prompt, such as when task instructions prime participants to respond in language-specific ways, or when stimuli are presented sequentially and language is employed strategically to hold events in memory. However, given the fact that experiment 3 of Papafragou and Selimis (2010) confounded type of instruction (containing no linguistic prompts) with presentation style (simultaneous instead of sequential), we cannot know whether the absence of the group differences can really be attributed to the simultaneous nature of the presentation, as the instructions alone could yield the same results (as indeed shown in experiment 2).

Despite these minor methodological shortcomings, Papafragou and colleagues' studies have provided invaluable insights into the potential effects of language on cognition. The studies reveal clear differences in verbalization patterns between populations who speak typologically different languages with respect to manner and path verb semantics. The cross-linguistic differences become less clear or disappear altogether in nonverbal tasks. For example, Papafragou, Hulbert, and Trueswell (2008) found no overall differences in attention allocation between groups, when participants were freely inspecting events as they were unfolding, but they did find differences in latter stages of attention allocation when participants had to commit events to memory. Trueswell and Papafragou (2010) further showed that any between-group differences in attention allocation to motion events are abolished in dual-task paradigms, where participants need to engage language to perform a verbal interference task parallel to the nonlinguistic cognitive task. Thus, according to the authors, effects of language on motion event cognition are of a transient nature.

1.2. Grammatical aspect and event construal

The grammatical category of aspect has as a basic function to denote the internal temporal constituency of an event, in the sense that it allows the speaker to either verbalize the intermediate phase (imperfective aspect) or the phase of completion (perfective aspect) of a given event (cf. Comrie, 1976). Languages across the world differ substantially in the relative prominence they give to the structural encoding of aspect; whereas some languages lack grammatical means to convey aspectual contrasts (e.g., German and Swedish), others present highly complex aspectual systems (e.g., Arabic and Chinese) (for overviews, see Binnick, 1991; Comrie, 1976; Dahl, 2000).

A growing body of psycholinguistic investigations has demonstrated that grammatical aspect is an important category in event construal and situation models (e.g., Madden & Zwaan, 2003; Magliano & Schleich, 2000). For example, using a semantic priming paradigm, Ferretti, Kutas, and McRae (2007) showed that English imperfective verb forms activate ongoing event schemas, where the location in which the event took place is highly relevant and therefore primed. This conclusion was confirmed by Willits, Sussman, and Amato

(2009), who in a series of experimental and corpus-based studies in addition found that verbs alone were more likely to co-occur with their related locations regardless of their aspectual form. Willits et al. (2009) argue that language users seem to be finely tuned to the differences between knowledge about events that is encoded in language structure (in this case grammatical aspect), and knowledge about events gained through experience, and they are very sensitive to which kind of knowledge is likely to be relevant and useful in a given situation.

Grammatical aspect has also been shown to play an important role in the domain of motion event construal. Anderson, Matlock, Fausey, and Spivey (2008) asked participants to listen to sentences containing simple past and past progressive verb forms. Using the computer mouse, participants had to place a character in the scene so that the scene matched the sentence they heard. Results showed that participants dropped the character to an intermediate point in the scene if they heard a sentence containing a verb in the past progressive form, while they tended to drop the character toward the destination described in the sentence if they heard a sentence containing a verb in the past simple. In addition, an analysis of the duration of participants' mouse movements showed that they moved the character into the scene for a longer duration of time when listening to past progressive sentences than when listening to simple past sentences. In a subsequent study using similar methodology, Anderson, Matlock, and Spivey (2010) showed that contextual information influenced interpretation of the sentences containing past progressive, but not simple past, presumably because contextual information pertains to the terrain of the ongoing action (expressed by the past progressive), and not to the completed state of the action (expressed by the simple past).

Another set of psycholinguistic studies have examined the relationship between grammatical aspect and motion event endpoints in language production. Imagine a scene where a person is walking along a road, with a house at a distance, but the clip ends before the person has gone anywhere near the house. When describing this event, the speaker may choose to adopt either a holistic perspective, that is, to interpret the event in its entirety including its endpoint (i.e., ''a person walks to a house''), or an immediate perspective, that is, to only focus on the ongoing phase of the event, excluding its endpoint (''a person is walking'') (see Langacker, 1987, 2008; Radden & Dirven, 2007). Previous studies have found that the choice of event perspective depends to a large extent on the native language of the observer. In languages with grammatical aspect (e.g., Algerian Arabic, English, Russian, Spanish), there is a verb form (e.g., the progressive *–ing* form in English) that relates to the ongoing phase of the event, ''zooming in'' on its progression and excluding its endpoints. Consequently, speakers of these languages have been shown to be more prone to defocus the endpoint of the event and instead direct attention to its ongoingness, describing it simply as *a person is walking*. In nonaspect languages (e.g., German, Norwegian, Swedish), on the other hand, in the absence of aspectual markers of ongoingness, the same event is typically construed under a holistic perspective where the endpoint is included, for example, from Swedish, *en person går till ett hus* (''a person walks to a house'') (Bylund, 2009; Schmiedtová, von Stutterheim, & Carroll, 2011; von Stutterheim & Nüse, 2003). Robust evidence on the link between grammatical aspect and endpoint encoding comes from a study on Swedish–Spanish bilinguals (Bylund & Jarvis, 2011). This study, which focused on the bilinguals' Spanish language skills, found that sensitivity to aspectual errors was negatively correlated

with a predilection for encoding event endpoints, in the sense that the weaker the bilinguals' command of the Spanish perfective–imperfective contrast, the more prone they were to encoding endpoints (thus deviating from the Spanish monolingual preference to encode few endpoints[1]).

The cross-linguistic differences in endpoint encoding documented by these studies reflect preferences rather than absolute patterns, in the sense that it is not ungrammatical in aspect languages to refer to endpoints, nor is it obligatory in nonaspect languages to omit endpoints.

1.3. The current study

There is, to the best of our knowledge, no study to date that investigates how cross-linguistic differences in the grammatical category of aspect may influence nonverbal motion event cognition. In this article, we further the study of language effects on motion event cognition by exploring the relationship between the grammatical category of aspect and endpoint encoding in language, memory, and cognition. The majority of previous studies exploring Whorfian effects on motion event cognition has focused on cross-linguistic differences in the encoding of manner and path of motion, thus taking as a starting point linguistic categories that first and foremost relate to the spatial properties of a motion event. However, given that motion events are characterized by a change through both space and time, then it remains known that our knowledge about Whorfian effects of language on motion event cognition will be limited if we only focus on categories concerning space, while leaving aside the possible influence of the linguistic encoding of the temporal distribution of an event. Likewise, the studies conducted to date on the effects of grammatical aspect on conceptualization and sentence priming show that aspect is a psychologically salient category during language production and comprehension, but they leave open the question of whether cross-linguistic differences in aspectual marking may give rise to relativistic effects.

Using stimuli depicting real life everyday motion events performed by different agents in different contexts, we investigate how speakers of Swedish (a nonaspect language) and English (an aspect language) encode and categorize motion events, by studying their linguistic observations while describing an event, their similarity judgments when asked to match different scenes from memory, and their similarity judgments when asked to match different scenes while they are unfolding. Given the recent findings by Papafragou and colleagues that show that language may affect cognition only when language is necessary to perform the task, it is particularly important to compare memory and online behavior in the populations under investigation.

The languages under study differ with regards to aspectual marking. In English, the verb is obligatorily marked for progressive aspect via the periphrasis BE + *-ing*. Swedish, on the other hand, lacks the grammatical means to convey aspectual contrasts. Speakers of Swedish may, however, optionally encode ongoing aspect through lexical circumlocutions, such as serial verbs (e.g., *sitta och äta* (''to sit and eat''), *stå och prata* (''to stand and talk''), *ligga och sova* (''to lie and sleep''), and *ute och gå* (''out and walk'') (Kvist-Darnell, 2005). It should nevertheless be noted that even though there is a possibility to convey aspectual contrasts in Swedish, the Swedish system is different from the English in two important regards:

first, the lexical means used to express ongoingness are not applied on an obligatory scale; and second, these lexical circumlocutions have a limited application, in the sense that they can only be combined with certain types of verb phrases (see Traugott, 1978, for further discussion on the grammaticalization of linguistic categories).

2. Experiment 1: Verbal encoding of goal-oriented motion events

In view of the differences in aspectual marking between Swedish and English, and the already established findings that speakers of nonaspect languages preferentially attend to endpoints more than speakers of aspect languages in their verbal descriptions of events (e.g., Bylund, 2009; von Stutterheim & Nüse, 2003), we predict that Swedish speakers will mention the endpoint of a motion event to a greater extent than English speakers in the verbal description task.

2.1. Participants

Forty participants (20 native speakers of Swedish and 20 native speakers of English) performed the task. All participants fell within the same age-range (19–32 years old), and had similar socioeconomic and education backgrounds (all University students in Sweden or the United Kingdom coming from middle class suburban backgrounds). All participants were given a monetary reward for their participation.

2.2. Materials

Twelve video clips showing goal-oriented motion events were used to study endpoint verbal encoding. The clips had been prepared and compiled by the research team of C. von Stutterheim, B. Schmiedtová, and M. Carroll at the University of Heidelberg, Germany, and were used previously in, for example, Bylund (2009), Bylund and Jarvis (2011) and Schmiedtová, von Stutterheim and Carroll (2011). The video clips depicted scenes with an intermediate degree of goal orientation, showing an entity (e.g., a vehicle or a person) moving along a trajectory at the end of which there was a possible endpoint that was not reached (e.g., a village). Six video clips depicting a simple action that did not involve movement along a trajectory (e.g., a person playing the piano) were used as fillers. Appendix A lists descriptions of all the video clips used. All video clips were 6 s long.

2.3. Procedure

Each participant was tested individually in a quiet room at a University in the United Kingdom or in Sweden. Participants were informed that they would watch a series of video clips depicting some action performed by one or more people. Participants watched the clips on a computer monitor and were asked to describe in their respective native language (English: *what is happening?* or Swedish: *vad händer*?) what was happening in each scene as soon as they recognized the type of situation.[2] The participants' descriptions of the video

clips were audio recorded and transcribed for analysis. Clips were presented one by one, in random order, using PowerPoint. Each clip lasted for 6 s and was followed by a purple screen which appeared for 8 s. Then a star appeared at the center of the screen, indicating that another clip was about to be shown. Participants' descriptions were audio recorded and transcribed by the task administrators for analysis.

2.4. Results and discussion

Participants' event descriptions were quantified in terms of the frequency with which they mentioned event endpoints. Results showed that on average Swedish speakers mentioned event endpoints in 61.7% (SD = 14.2) of the motion event scenes, whereas the English speakers did so for 42.8% of the scenes (SD = 17.2). In other words, when confronted with a scene such as two people in the direction of some houses, English speakers where more prone to describing this scene as *two people are walking*, whereas Swedish speakers were inclined to mention the endpoint, *två personer går till ett hus* (''two persons walk to a house''). An independent samples t test confirmed that the difference in endpoint encoding between the two groups was statistically significant, t (38) = 3.93, $p < 0.01$, d = 1.26. We further considered this behavior within each group, to see to what extent participants showed significant biases toward ongoingness or endpoints of events. One-sample t tests against chance level (50%) confirmed that Swedish-speaking participants mentioned event endpoints significantly above chance, t (19) = 3.69, $p < 0.01$, while English-speaking participants focused on ongoingness, omitting endpoints significantly above chance, t (19) = −2.06, $p < 0.05$. A closer look at the linguistic encodings revealed that the English speakers used the progressive present form of the verb in 100% of their descriptions (i.e., irrespective of whether they encoded an endpoint or not). This finding is not unexpected: the simple present tense in English may be used in generic predications, but rarely to encode an unfolding, single event (cf. *She cycles to work* vs. *She is cycling to work*). In the Swedish event descriptions, a lexical construction to convey ongoingness was on average used in 0.3% of the times. The specific construction used was *ute och* + VERB (''out and''), which is not combinable with endpoints. The default choice in the Swedish group was thus to describe the events using the simple present tense.[3]

Our results confirmed the hypothesis that Swedish speakers will tend to mention event endpoints to a greater extent than English speakers. Having established that Swedish and English speakers verbalize events in different ways with respect to endpoint encoding, we now turn to investigate whether these differences are reflected in two nonverbal cognitive tasks.

3. Experiments 2a and 2b: Memory based and online nonverbal similarity judgments of goal-oriented motion events

The main question of interest in this article concerns performance in the nonverbal tasks. Here, we can plausibly posit three alternative possibilities. One possibility is that grammatical aspect encoding and motion event cognition are completely dissociable, such that differences between groups will be limited only to linguistic descriptions, and no differences will

be observed in the nonlinguistic cognitive tasks. Another possibility is that effects of grammatical aspect on cognition will be observable only in the memory version of the triads matching task, because participants may use language as a tool to encode and retain scenes for later matching (as predicted by Papafragou & Selimis's linguistic intrusion account). The third possibility is that grammatical aspect habitually shapes motion event cognition, such that differences in behavior between Swedish and English speakers will be observable in both of the cognitive tasks.

3.1. Participants

Thirty-four Swedish native speakers and 34 English native speakers were randomly allocated to one of two tasks. Seventeen participants from each language group completed experiment 2a (memory version of the triads matching task), and 17 participants from each language group completed experiment 2b (online version of the triads matching task). All participants fell within the same age-range (19–34 years old), and had similar socioeconomic and education backgrounds (all University students in Sweden or the United Kingdom coming from middle class suburban backgrounds). All participants were given a monetary reward for their participation.

3.2. Materials

Thirty-one video clips from the stimulus pool of the research group described in the previous section were used in all permissible combinations to create 19 triads, consisting of a target and two alternates. The target was a scene with an intermediate degree of goal orientation in which there was a visible, possible endpoint for the motion, but the arrival at it was not overtly shown (e.g., a person walking toward a cafe). One type of alternate, which we shall label [−endpoint], was a video with a low degree of goal orientation, depicting the motion of an entity along a trajectory without an immediate endpoint (e.g., a person walking along the pavement). The other type of alternate, which we shall label [+endpoint], was a scene with a high level of goal orientation. These scenes depicted the motion of an entity that actually reaches an endpoint (for example a person walking and entering a building). Fig. 1 shows examples of clips with high, intermediate, and low degree of goal orientation. Triads were created (and clips edited where necessary) in such a way as to control for manner and direction of motion, and number and gender of agents. Within each triad, each stimulus contained a different agent and contextual background. All clips were 6 s long. Appendix B provides a list of all triads used.

The videos were normed for visual similarity with members of both language groups to establish a baseline of visual similarity outside of the experimental situation. This was done by using a still of the initial frame from each video (because we wanted to avoid the factor of the motion event, which was the crucial experimental variable), and arranging a presentation whereby a still of each alternative video was presented alongside a still of the corresponding target video. Thus, two types of pairs of pictures were created, one type that contained a [+endpoint] alternate and its corresponding target, and one type that contained a [−endpoint] alternate and the relevant target. Each pair appeared twice, with the left/right

position counterbalanced. Twenty-six participants (13 native English speakers and 13 native Swedish speakers, with similar backgrounds to the participants in Experiments 1 and 2) who did not take part in any of the subsequent experiments were asked to rate the members of each pair of pictures for visual similarity on a 9-point Likert scale, where 1 represented ''very dissimilar''/''mycket olika'' and 9 represented ''very similar''/''mycket lika.'' Table 1 shows the mean similarity ratings for each pair type for the two groups. A 2 (Group: Swedish vs. English) × 2 (Pair type: [+endpoint and target] vs. [−endpoint and target]) mixed ANOVA showed that the main effects of Group and Pair type were not significant ($p > 0.05$), and, crucially, the interaction was not significant either ($p > 0.05$). This means that the [−endpoint] alternates looked as similar to the target as the [+endpoint] alternates did, and the two groups did not differ in how they perceived the relative visual similarity between each alternate and its corresponding target.

3.3. Procedure

Experiment 2a (Memory-based triads matching): Each participant was tested individually in a quiet room in a University in the United Kingdom or in Sweden. Participants were told

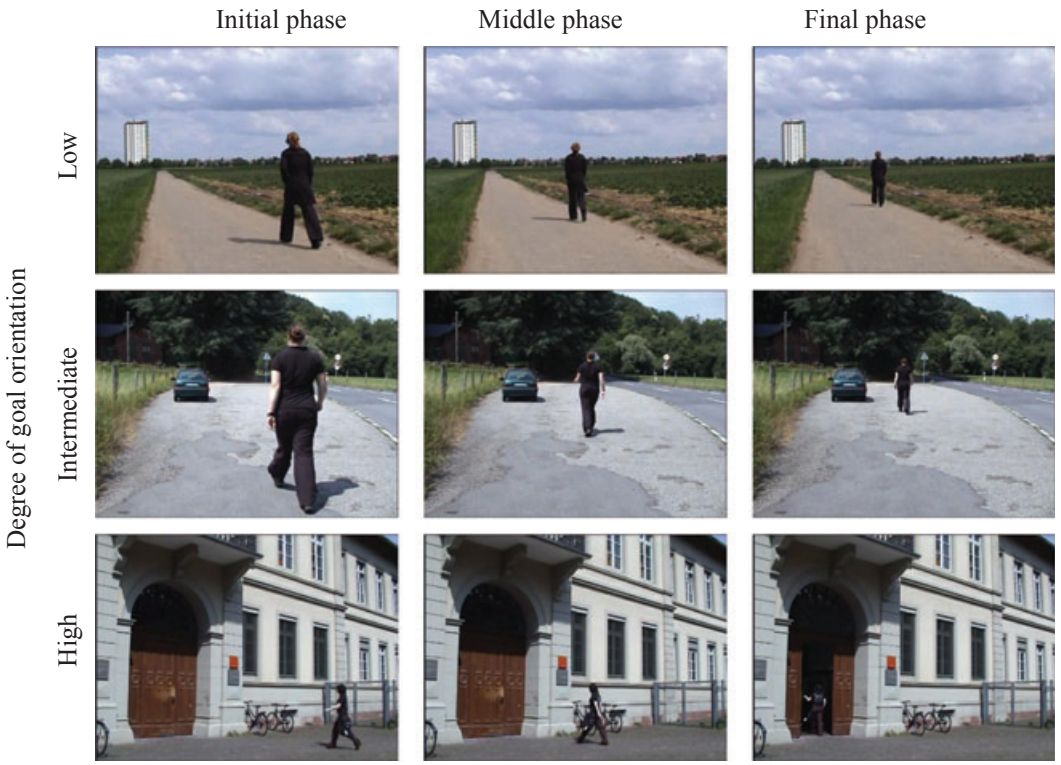

Fig. 1. Examples of clips with high, intermediate, and low degree of goal orientation, representing the initial, middle point, and ending of the clip.

Table 1
Mean similarity ratings (and standard deviations) on a 9-point Likert scale (9 = very similar, 1 = very dissimilar) for each pair type for the two language groups

	Target & [−endpoint]	Target & [+endpoint]
English	4.49 (1.19)	4.29 (1.13)
Swedish	4.48 (1.32)	4.28 (1.43)

that they would see video clips arranged in triads, where clip A would appear first, then clip B, and finally clip X (the target). Participants were asked to indicate whether they thought clip X was more like clip A or more like clip B. Thirty-eight triads were thus presented in an ABX format, where in a counterbalanced design within each participant half of the time the [-endpoint] alternate appeared first (clip A) and half of the time it appeared second (clip B), and vice versa for the [+endpoint] alternate. The precise sequence of the clips in each triad was as follows: Clip A played, followed by clip B, followed by clip X. Participants were instructed to give their responses only after they had watched clip X in its entirety. Clips played immediately after one another (i.e., there was no pause in-between). Clips were presented on a computer screen using the presentation software PowerPoint. The presentation was timed using the relevant function in PowerPoint, and participants had no control over the sequence of presentation (i.e., they could not pause or replay a clip, or pause in-between clips within a triad). Participants were given as much time as they liked to provide their judgment to the task administrator. After their judgment had been recorded by the task administrator, they could proceed to the next triad by pressing the left click of the computer mouse.

Experiment 2b (Online triads matching): Here, the design was the same as in the memory-based triads matching task, except this time all three clips appeared at once, clip X appearing at the center top of the screen, and clips A and B appearing at the bottom left and bottom right of the screen, below clip X. Participants received 38 trials in total, counterbalancing the left/right position of the two alternates in each triad within each participant. Participants were asked to indicate whether they thought clip X was more like clip A or more like clip B. The clips were synchronized and played in a loop, until the participant gave their response. Participants were given as much time as they liked to provide their judgment to the task administrator. After they made their judgment, they could proceed to the next triad by pressing the left click of the computer mouse. The next triad appeared immediately (i.e., there was no pause in-between triads).

For both versions of the triads matching task, the order of presentation was fully randomized for each participant.

3.4. Results and discussion

For the two similarity judgment tasks, the dependent variable was exactly the same, namely the amount of times participants matched the target clip (X) with the [+endpoint] alternate (indicating an endpoint preference). We thus converted all scores to percentages and compared endpoint preference in similarity judgments from memory with endpoint

preference in online similarity judgments in a 2 (Group: Swedish vs. English speakers) × 2 (Task type: SJ memory vs. SJ online) fully between-subjects ANOVA.[4] The main effect of Group was significant, F (1, 64) = 13.20, $p < 0.01$, $\eta_p^2 = 0.17$. There was no significant main effect of Task, F (1, 64) = 1.12, $p > 0.05$, but crucially the interaction was significant F (1, 64) = 7.01, $p < 0.05$, $\eta_p^2 = 0.09$, indicating that there were differences between the groups in their endpoint preference across the two different types of task. Post hoc t tests revealed that the Swedish speakers matched the target clip with the [+endpoint] alternate significantly more than English speakers did ($p < 0.01$, $d = 1.54$) in the SJ memory task, while no difference between the groups was observed in the SJ online task ($p > 0.05$). As Fig. 2 shows, the two language groups differ in their endpoint encoding in their linguistic descriptions, as well as their endpoint preference in their similarity judgments from memory. No between-group differences are observed in online similarity judgments.

We further considered cognitive behavior within each group, to see to what extent participants are showing significant biases toward ongoingness or endpoints of events. One-sample t tests against chance level (50%) confirmed that in both experiments 2a and 2b, both groups selected the [−endpoint] alternate significantly above chance, t (16) = −6.70, $p < 0.01$, and t (16) = −11.61, $p < 0.01$ for the Swedish speakers, and t (16) = −11.85, $p < 0.01$, and t (16) = −9.74, $p < 0.01$ for the English speakers. That is, both language groups preferred overall the [−endpoint] alternate in the similarity judgment tasks, but the Swedish speakers

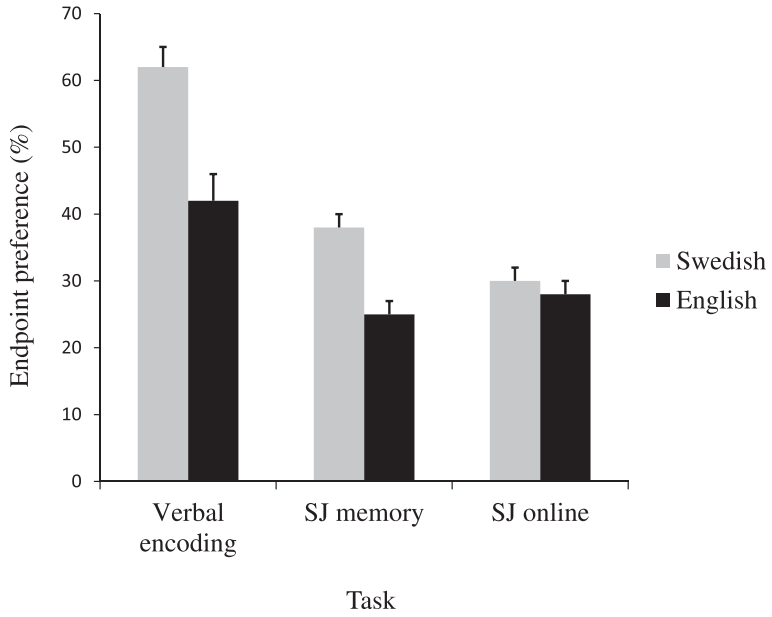

Fig. 2. Endpoint preferences (%) across groups in experiments 2a (SJ memory) and 2b (SJ online), alongside endpoint preferences in experiment 1 (verbal encoding). SJ, similarity judgments. Columns show group means. Error bars show standard error of the mean.

did so significantly less than the English speakers in the SJ memory task, showing a significant increase in their [+endpoint] preferences relative to English speakers.

The results from experiments 2a and 2b clearly demonstrate that in a nonlinguistic matching task that involves memory, differences in event cognition between speakers emerged, demonstrating cross-linguistic differences in memory-based event cognition as a function of grammatical aspect. On the other hand, language effects were not found when stimuli were presented simultaneously, thus alleviating the pressure for linguistic encoding to commit events to memory. To firmly establish the role of language in memory-based similarity judgments, we repeated experiment 2a with a verbal interference component, following Gennari et al. (2002).

4. Experiment 2c: Verbal interference in memory-based similarity judgments

If language is indeed recruited to solve tasks that require temporarily storing information in memory, then we can predict that verbal interference during memory-based cognitive processing will abolish the cognitive difference between the two language groups.

4.1. Participants

Fifteen native speakers of Swedish and 15 native speakers of English who had not taken part in any of the other experiments completed the verbal interference triads matching task. All participants fell within the same age-range (18–36 years old), and had similar socioeconomic and education backgrounds (all University students in Sweden or the United Kingdom coming from middle class suburban backgrounds). All participants were given a monetary reward for their participation.

4.2. Materials

The materials used were exactly the same as in experiment 2a (see relevant section above).

4.3. Procedure

The procedure was identical to experiment 2a, except that at the onset of each triad participants heard a string of three-two digit numbers (e.g., 37, 41, 54) and were asked to repeat the string out loud (in their respective native language) and continue to repeat it until they had watched all three clips in sequence (following the verbal interference technique used in Trueswell & Papafragou, 2010).

When the third (target) clip finished playing, participants were instructed to stop repeating the number sequence and give their response to the task administrator. For each triad, participants heard and had to reproduce a different string of numbers that had been randomly generated for each triad and for each participant using the relevant function in Excel.

4.4. Results and discussion

Similarly to experiments 2a and 2b, the dependent variable was the amount of times participants matched the target clip (X) with the [+endpoint] alternate (indicating an endpoint preference). Scores were converted to percentages and averages calculated for each group. The average [+endpoint] preference for the Swedish speakers was 29.65% (SD = 6.99), whereas for the English speakers it was 32.81% (SD = 9.43). An independent samples t test revealed that the difference between the groups was not statistically significant, t (28) = −1.042, $p > 0.05$.

This finding is in line with a large body of literature that shows that cross-linguistic differences disappear under a verbal interference condition, not only in the domain of motion events (Gennari et al., 2002; Trueswell & Papafragou, 2010) but also in other domains such as categorical perception of color and faces (Roberson & Davidoff, 2000; Winawer et al., 2007). These findings are consistent with accounts that nonlinguistic cognitive processes such as memory are linguistically mediated (Baddeley, 2003), and they show that language is indeed used as a tool for carrying out mental tasks that involve memory. We discuss this finding further later on in the context of the cognitive mechanism that may account for these findings.

5. General discussion

Our results show for the first time cognitive differences between populations as a function of grammatical aspect encoding in their respective languages. In line with previous studies, our results also show that this effect is modulated by how relevant language is in carrying out the task at hand. This finding provides substantial empirical support to Papafragou and Selimis' (2010) linguistic intrusion account, and extends its validity to the category of grammatical aspect. However, our results qualify the linguistic intrusion account in an important way. Contrary to Papafragou and Selimis' (2010) suggestion that effects of language on event cognition are not only superficial but also unnecessary and inconsistent, here we demonstrate that even in the absence of any linguistic prompts in the instructions, cognitive differences between speakers of different languages do emerge. This is arguably because our memory task posed greater cognitive demands on the participants, thus encouraging and increasing reliance on verbal strategies to solve the task. Recall that in this study, participants saw each clip once, and they had no opportunity to make judgments of frozen images, as at the time of making their selection they could not see the alternate images alongside the target.

In addition, our design sought to approximate ''real life'' event cognition as closely as possible where people are exposed to a rich variety of dynamic situations and endpoint-based classification is one of many other linguistic and extra-linguistic possibilities of classifying an event. It is remarkable that even in these varied contexts clear cognitive differences in memory emerged. Our results therefore are also consistent with the idea that ''language invades our thinking because languages are good to think with'' (Levinson, 2001, p. 584).

This interpretation of the results is further reinforced by the finding that verbal interference in a memory-based task abolished the cognitive differences between Swedish and English speakers found in a memory-based task without such interference, thus showing that language is indeed recruited online to solve mental tasks that involve a memory component.

It is important however to consider the above remarks in the context of the global tendencies in our groups. As it is evident from Fig. 2, and as the one-sample *t* tests confirmed, the two groups of speakers tend to show an overall preference for ongoingness rather than endpoints in their similarity judgments, both in the memory and in the online version of the task. In this larger picture, the cross-linguistic difference found in the memory version of the similarity judgment task points to a very weak Whorfian effect: both groups pay more attention to ongoingness, with the Swedish speakers doing significantly less so than the English speakers when they have to encode events in memory for later matching. This difference between populations is likely rooted in the language they speak, as the linguistic encoding task showed not only a significant between-group difference but also a significant bias toward endpoints within the Swedish-speaking group.

In this way, our results nicely complement emerging evidence that shows that rather than language shaping cognition, language may attenuate already existing and possibly universal perceptual biases. For example, while the finding that English speakers pay more attention to the distinction between countable and noncountable entities than Japanese speakers is well established (Imai & Gentner, 1997; Imai & Mazuka, 2003), the evidence also shows that overall, Japanese speakers also show a significant tendency to differentiate between countable and noncountable entities (Athanasopoulos, 2007; Imai & Mazuka, 2007). The effect is stronger in English speakers because English highlights the universal ontological distinction between shapes and substances by means of explicit grammatical morphology in its nominal system (i.e., the count/mass grammatical distinction, which Japanese lacks, Downing, 1996; Takano, 1994).

We are of course not in a position to claim that the bias reported here for ongoingness is universal because we have evidence from only two languages. However, what is likely to be universal, or at least impervious to linguistic influence, is attention to the perceptual properties inherent to the motion event. In the experimental design of the two nonlinguistic tasks, the basic difference between the scenes is that endpoints are reached in the [+endpoint] alternates, but not reached in the [−endpoint] alternates, and, crucially, also not reached in the targets (where the endpoint was nonetheless visible and highly salient). Thus, the [+endpoint] alternates are of a boundary-crossing nature, where the agent moves from one location into another, so-called ''resultative end-of-path location,'' whereas the targets and [−endpoint] alternates are not (cf. Aske, 1989; Slobin & Hoiting, 1994). Notably, this change of state entails a more general difference between the [+endpoint] alternate vis-à-vis the target scene and the [−endpoint] alternate, whereas the latter two portray the events as unbounded, the former shows a bounded event. The [+endpoint] alternates are, in other words, different from the other two scenes along more than one dimension. These perceptual differences are most likely the driving mechanism underlying the tendency to, regardless of language background, make similarity judgments based on whether the endpoint is reached or not.

Our study thus provides converging evidence in support of the claim that ''linguistic encoding (and the accompanying attention effects) cannot override the perceptual encoding of the events'' (Trueswell & Papafragou, 2010, p. 79). This interpretation is in line with literature that shows that people are guided by perceptual characteristics of motion events (such as the relatively higher perceptual saliency of endpoints compared with beginnings of actions) regardless of their diverse language backgrounds, in adults as well as in children, and in typical and atypical populations (Lakusta & Landau, 2005; Papafragou, 2010; Regier & Zheng, 2007; Zheng & Goldin-Meadow, 2002).[5]

How are we then to explain the differences between populations that we found in the memory variant of the similarity judgment task? What is the cognitive mechanism by which endpoint encoding in language may influence motion event cognition? According to Langacker (2008), frequency of use of a particular linguistic construction makes that construction part of the cognitive routine of the speaker, thus enhancing its salience in the speaker's mind. This view is compatible with a more general account of how language affects thinking. Casasanto (2008) proposes a mechanism for language effects on cognition that posits simple associative learning at its core. Language users habitually activate corresponding conceptual categories when using lexical terms or grammatical constructions. The process of language use then entails strengthening associations between words/constructions and their corresponding referents in the real world. In the context of this article, habitual or recurrent use of certain grammatical features would make the specific perspectives (holistic or immediate) associated with those features more salient in the mind of the language user. Therefore English speakers, whose language has a progressive marker, tend to adopt an immediate perspective defocusing the endpoints of events to a greater extent than Swedish speakers, whose language lacks a progressive marker. Our linguistic data show some support for this hypothesis, as we found reliable differences between groups, as well as significant biases within each group toward the predicted patterns. This finding in itself is consistent with the view that grammatical categories (in this case the category of aspect) influence how speakers organize information for the purposes of verbalization (cf. Slobin, 2003; von Stutterheim & Nüse, 2003).

Obviously, speaking differently hardly entails thinking differently (Pinker, 1994, 2007), and the question of interest to most cognitive scientists would be whether this kind of association strengthening between grammatical aspect and viewing frames of reference has an impermeable grip on cognition, permanently shaping the way observers of different languages perceive and interpret motion events, or whether the effect of language is more dynamic and flexible, becoming apparent only in certain contexts, for example when language is used as a strategy to carry out the task. If the former hypothesis is correct, then we should expect to find differences between speakers of different languages across a range of cognitive tasks, including both tasks involving memory and tasks that involve online processing. Clearly, the results of this study do not provide support for this hypothesis, as we only found differences between populations in the memory version of the triads matching task, and these differences disappeared in a verbal interference task. Rather, our results support the latter hypothesis that language is used strategically to carry out the task.

In this way, our results conform to a fundamental principle of dual coding theory (Paivio, 1971, 1986, 2007), namely the sequential constraints of language versus the synchronous nature of nonverbal cognition in episodic memory. Over the years, several studies have demonstrated that the verbal system is specialized for the sequential organization of information in memory, whereas the nonverbal system is specialized for the synchronous organization of information in memory, in particular visual pictorial information (for overviews see Paivio, 1986, 2007). In this study, a visual pictorial episode was presented in all task conditions. Language was explicitly required in one task (experiment 1), and here, the differences between the groups were maximal. Sequential processing of pictorial information was involved in another task (experiment 2a), and this task likely involved a verbal memory strategy that gave rise to a certain degree of cognitive differences. Synchronous processing of pictorial information was involved in the third task (experiment 2b), and no differences between groups were found as the synchronous nature of the presentation required little or no language to carry out the task. The fourth task (experiment 2c) demonstrated that verbal interference neutralized the cognitive differences between groups, thus establishing the role of language in memory-based cognitive processing. These patterns are quite consistent with the basic tenets of dual coding theory, and this is the likely cognitive mechanism that explains these and previously reported findings from memory, online, and verbal interference tasks in the domain of motion events. Our study then contributes to the emerging picture in the field of event cognition, that linguistic structure matters and can have a big impact on cognition, but that impact is limited to where it is useful and relevant (Willits et al., 2009).

6. Conclusions

We open up the investigation of Whorfian effects of grammatical aspect, motivated on the one hand by recent psycholinguistic evidence showing that aspectual contrasts such as those distinguishing progressive versus nonprogressive events have psychological reality in the minds of speakers (Anderson et al., 2008, 2010), and on the other hand by linguistic research showing a close relationship between linguistic typology (i.e., whether a language has or lacks *grammaticized* means of conveying progressive aspect) and tendencies by speakers to focus on the ongoingness or on the endpoints of motion events in their linguistic descriptions (Bylund & Jarvis, 2011; von Stutterheim & Nüse, 2003).

Our results revealed differences in linguistic and memory-based cognitive behavior between populations that speak typologically different languages with regards to grammatical aspect. We interpreted these cross-linguistic differences in the light of the global patterns of behavior in our groups, which revealed a tendency to rely on common perceptual properties of scenes when making similarity judgments, regardless of the participants' language background. In this way, our findings contribute to the larger picture of Whorfian effects on cognition, suggesting that language may impact upon but not override attentional biases toward perceptual characteristics of events.

Indeed, even in domains like color perception, where strong effects of language that go beyond conscious strategic higher level processing are found (see e.g., Athanasopoulos,

Dering, Wiggett, Kuipers, & Thierry, 2010; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009), studies also show that color systems are not arbitrary and culture driven but are motivated, at least in part, by optimality constraints for encoding the color spectrum as well as universal perceptual foci (Regier, Kay, & Khetarpal, 2007). This observation is all the more interesting due to the fundamental sensory differences between these domains, as contrary to color, motion events are of a transient nature, and can be decomposed into several components, thus giving the observer freedom to interpret the event in many different ways, and not necessarily in a way that is consistent with a particular lexical or grammatical feature of motion event encoding. The way we cut up color spectrum on the other hand is to a large extent dependent on the labels that are made available in our language (Roberson, Davies, & Davidoff, 2000), that is, there is a vis-à-vis relationship between the perceptual construct investigated and the labeling of the construct itself.

More importantly, our study further contributes to the emerging view that Whorfian effects are not an all-or-nothing phenomenon. A common misconception of Whorf's principle of linguistic relativity is to assume that thinking is not possible without language, or that thinking is somehow constrained by the lexical and grammatical categories made available in language. Such misconception may be found in sceptics of Whorf's ideas, who equate linguistic relativity with some form of radical linguistic determinism (see e.g., Pinker's infamous discussion in his 1994 book). In fact, Whorf himself contended that ''language, for all its kingly role, is in some sense a superficial embroidery upon deeper processes of consciousness....[t]he statement that 'thinking is a matter of LANGUAGE' is an incorrect generalization of the more nearly correct idea that 'thinking is a matter of different tongues' '' (Whorf, 1956, p. 239; emphases in original). Thus, in Whorf's view, thinking is not constrained by language, but different languages highlight different patterns of thinking.

In this light, the question of whether cognitive processes are influenced by language or whether they are impervious to linguistic influence becomes pointless. The question might be more usefully rephrased as ''what is the *extent* of the influence of linguistic structure on cognitive processes such as similarity judgments and categorization, and *what conditions* suppress or promote this influence?'' The picture that is beginning to emerge is one that points to an attenuating effect of language on perceptual processes that are likely to be universal and unchanging. Triangulating evidence from different domains of inquiry and from different experimental paradigms that investigate behavior across a range of cognitive processes will nuance further the intricate relationship between language, memory, and thought.

Acknowledgments

The authors were supported by a Research Development Fund from the ESRC Centre for Research on Bilingualism in Theory and Practice. We are grateful to M. Carroll, B. Schmiedtová, and C. von Stutterheim, for making their stimuli available to us, to V. Gathercole and G. Thierry for useful discussions during the second author's research visit at Bangor University, to L. Damjanovic for invaluable help with participant recruitment, and to three reviewers for their constructive and helpful comments.

Notes

1. At first glance, the findings on Spanish speakers not encoding event endpoints (Bylund, 2008; Bylund & Jarvis, 2011; Schmiedtová, Carroll, & von Stutterheim, 2011; von Stutterheim, Nüse, & Murcia-Serra, 2002) might seem at variance with Slobin's (1996) observation that Spanish speakers are prone to mention resultative endstates (unfortunately, Slobin does not bolster his claim with any quantitative data). It is however important to keep in mind that whereas the findings from the first set of studies concern the description of unfolding, goal-oriented—but not goal-reached—dynamic scenes, Slobin's observation stems from static pictures depicting goal-reached scenes, for instance a boy sitting in a tree. Research has shown that cross-linguistic differences in the encoding of endpoints are most likely to occur when there is movement *toward* a goal. However, if the goal is reached cross-linguistic differences disappear (Carroll & von Stutterheim, 2006; Schmiedtová & Sahonenko, 2008; von Stutterheim, 2003). The possible discrepancy between Slobin's observation and the findings by von Stutterheim and associates could then be explained by taking into account the different methodologies and foci that underlie these studies.
2. Previous studies have demonstrated that phrasing the question as ''what happens?'' elicits the same kind of responses in terms of endpoint encoding from English-speaking participants as using the present progressive form (''what is happening''), (von Stutterheim, Carroll, & Klein, 2003; von Stutterheim & Nüse, 2003).
3. Based on Aske's (1989) analysis, a reviewer raises the possibility that verb type (or lexical aspect) and grammatical aspect may interact, such that manner verbs would be more likely to appear with an imperfective/progressive marker, whereas path verbs would be more likely to occur without such marker. In the English event descriptions, it was found that 87.2% of the verbs used were manner verbs, and 12.8% were path verbs. Given that these descriptions without exception contained only the progressive form, no interplay between aspectual marking and verb type can be observed. A very similar distribution of verb type was found in the Swedish data, where 86.8% of the verbs used were manner verbs and 13.2% were path verbs. Those rare instances of lexical markers of ongoingness that were found in the Swedish descriptions were indeed combined with manner verbs only (e.g., *gå* ''walk''; *cykla* ''cycle''). However, the low number of occurrences seriously reduces the generalizability of this finding.
4. Levene's test of equality of error variances confirmed that all assumptions necessary for conducting this ANOVA were met, $F\ (3, 64) = 0.18$, $p = 0.91$.
5. Of course our research is not directly related to this literature as our experiments were not designed to explicitly test source-goal asymmetries in attention to motion events.

References

Anderson, S., Matlock, T., Fausey, C., & Spivey, M. (2008). On the path to understanding on-line processing of grammatical aspect. In B. C. Love, K. McRae, V. M. Sloutsky (Ed.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 143–148). Mahwah, NJ: Lawrence Erlbaum.

Anderson, S. E., Matlock, T., & Spivey, M. J. (2010). On-line interactions of context and grammatical aspect. In S. Ohlsson, R. Catrambone (Ed.), *Proceedings of the 32th Annual Conference of the Cognitive Science Society* (pp. 1198–1203). Mahwah, NJ: Lawrence Erlbaum.

Aske, J. (1989). Path predicates in English and Spanish: A closer look. *Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society*, *15*, 1–14.

Athanasopoulos, P. (2007). Interaction between grammatical categories and cognition in bilinguals: The role of proficiency, cultural immersion, and language of instruction. *Language and Cognitive Processes*, *22*, 689–699.

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J., & Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, *116*, 437–443.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.

Binnick, R. (1991). *Time and the verb: A guide to tense and aspect*. Oxford: Oxford University Press.

Bylund, E. (2008). Procesos de conceptualización de eventos en español y en sueco: Diferencias transglingüísticas. *Revue Romane*, *43*, 1–24.

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

Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition*, *14*, 47–59.

Carroll, M., & von Stutterheim, C. (2006). The impact of grammaticalized temporal categories on ultimate attainment in advanced L2 acquisition. In H. Byrnes, H. Weger-Guntharp, & K. Sprang (Eds.), *Educating for advanced foreign language capacities* (pp. 40–53). Washington, DC: Georgetown University Press.

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

Casati, R., & Varzi, A. (2008). Event concepts. In T. Shipley, & J. Zacks (Eds.), *Understanding events* (pp. 31–55). Oxford: Oxford University Press.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Dahl, Ö. (Ed.) (2000). *Tense and aspect in the languages of Europe*. Berlin: Mouton de Gruyter.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories of a stone-age tribe. *Nature*, *398*, 203–204.

Downing, P. A. (1996). *Numeral classifier systems: The case of Japanese*. Amsterdam: John Benjamins.

Feist, M. (2010). Motion through syntactic frames. *Cognition*, *115*, 192–196.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 182–196.

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

Imai, M., & Gentner, D. (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

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: MIT Press.

Imai, M., & Mazuka, R. (2007). Revisiting language universals and linguistic relativity: Language-relative construal of individuation constrained by universal ontology. *Cognitive Science*, *31*, 385–414.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65–79.

Kvist-Darnell, U. (2005). Utvidgade och andra användningar av de svenska verben *sitta*, *stå*, och *ligga*. In G. Byrman, J. Einarsson, S. Hammarbäck, M. Lindgren, & P. Stille (Eds.), *Svenskans beskrivning 27* (pp. 148–160). Växjö: Växjö University Press.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*, 1–33.

Langacker, R. (1987). *Foundations of cognitive grammar*, Vol. 1. Stanford: Stanford University Press.

Langacker, R. (2008). *Cognitive grammar: A basic introduction*. New York: Oxford University Press.
Levinson, S. C. (2001). Covariation between spatial language and cognition, and its implications for language learning. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 566–588). Cambridge: Cambridge University Press.
Levinson, S. (2003). *Space in language and cognition: Explorations in cognitive diversity*. 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.A. (1997). Linguistic relativity *Annual Review of Anthropology*, *26*, 291–312.
Madden, C., & Zwaan, R. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, *31*, 663–672.
Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, *29*, 83–112.
Majid, A., Bowerman, M., Kita, S., Haun, D., & Levinson, S. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*, 108–114.
Mayer, M. (1969). *Frog, where are you?* New York: Dial.
Naigles, L., & Terrazas, P. (1998). Motion verb generalizations in English and Spanish: Influences of language and syntax. *Psychological Science*, *9*, 363–369.
Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.
Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, *34*, 1064–1092.
Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 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 categorisation and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, *25*, 224–260.
Pinker, S. (1994). *The language instinct*. London: Penguin.
Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York: Viking.
Radden, G., & Dirven, R. (2007). *Cognitive English grammar*. Amsterdam: John Benjamins.
Regier, T., & Kay, P. (2009). Language, thought and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*, 439–446.
Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436–1441.
Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science*, *31*, 705–719.
Roberson, D., & Davidoff, J. (2000). The categorical perception of colours and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*, 977–986.
Roberson, D., Davies, I., & Davidoff, J. (2000). Colour categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369–398.
Schmiedtová, B., Carroll, M., & von Stutterheim, C. (2011). Implications of language-specific patterns of event construal of advanced L2 speakers. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 66–107). Clevedon: Multilingual Matters.
Schmiedtová, B., v. Stutterheim, Ch., & 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). Bristol: Multilingual Matters.
Schmiedtová, B., & Sahonenko, N. (2008). Die Rolle des grammatischen Aspekts in Ereignis Enkodierung: Ein Vergleich zwischen tschechischen und russischen Lernern des Deutschen. In P. Gommes, & M. Walter

(Eds.), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbforschung* (pp. 45–71). Tübingen: Max Niemeyer Verlag.

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

Slobin, D. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in mind. Advances in the study of language and thought* (pp. 157–191). Cambridge: MIT Press.

Slobin, D., & Hoiting, N. (1994). Reference to movement in spoken and signed language: Typological considerations. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, *20*, 487–505.

von Stutterheim, C. (2003). Linguistic structure and information organisation. The case of very advanced learners. In S. Foster-Cohen, & S. Pekarek-Doehler (Eds), *EuroSLA yearbook* (pp. 183–206). Amsterdam: Benjamins.

von Stutterheim, C., Carroll, M., & Klein, W. (2003). Two ways of construing complex temporal structures. In F. Lenz (Ed.), *Deictic conceptualization of space, time and person* (pp. 97–133). Berlin: De Gruyter.

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

von Stutterheim, C., Nüse, R., & Murcia-Serra, J. (2002). Crosslinguistic differences in the conceptualization of events. In H. Hasselgård, S. Johansson, B. Behrens, & C. Fabricius-Hansen (Eds.), *Information structure in a crosslinguistic perspective* (pp. 179–198). Amsterdam: Ropodi.

Takano, H. (1994). Japanese common nouns and their unquantificational nature. In N. Akatsuka (Ed.), *Japanese/Korean linguistics 4* (pp. 379–394). Leland Stanford Junior University: Center for the Study of Language and Information.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics* (pp. 181–238). New York: Academic Press.

Talmy, L. (1988). The relation of grammar to cognition. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 165–205). Amsterdam: John Benjamins.

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

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

Traugott, E. C. (1978). On the expression of spatio-temporal relations in language. In J. H. Greenberg (Ed.), *Universals of human language: Vol. 3. Word structure* (pp. 369–400). Stanford: Stanford University Press.

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

Whorf, B. L. (1940/1956). Linguistics as an exact science. In J. B. Carroll (Ed.), *Language, thought, and reality. Selected writings of Benjamin Lee Whorf* (pp. 220–232). Cambridge, MA: MIT Press.

Willits, J., Sussman, R., & Amato, M. (2009). Event knowledge vs. verb knowledge. In N. Taatgen, H. van Rijn (Ed.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2227–2232). Mahwah, NJ: Lawrence Erlbaum.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on colour discrimination. *Proceedings of the National Academy of Sciences*, *104*, 7780–7785.

Zheng, M., & Goldin-Meadow, S. (2002). Thought before language: How deaf and hearing children express motion events across cultures. *Cognition*, *85*, 145–175.

Appendix A: List of stimuli used in the verbal encoding task

Action	Possible endpoints
Experimental stimuli	
Person walking	A car, a house, some trees
Person riding bicycle	Houses/a village, some trees
Person walking	A cafe, some chairs outside
Car driving on the road	Some trees, buildings, a petrol station
Person walking	A nearby outdoor market
Person walking	cars parked nearby
Two people walking	Some houses in the distance
Person running	The bank of a river, the river, some trees
Two people walking	A house, some trees
Two people walking	A playground
Car driving on the road	A petrol station, a railway crossing, some trees
Person riding a bicycle	Several shops
Fillers	
Person drinking coffee	N/A
Person waking up	N/A
Person shaving	N/A
Person playing the piano	N/A
Person peeling potatoes	N/A
Person typing	N/A

Appendix B: Triads of stimuli used in the similarity judgment task (memory and online)

Target (intermediate degree of goal orientation)	[-endpoint] alternate (low degree of goal orientation)	[+endpoint] alternate (high degree of goal orientation)
Person walking toward a cafe/outdoor seating area	Person walking on a country road	Person walking and entering a building
Person walking toward a car/trees	Person walking on a country road	Person walking and entering a building
Person walking toward a cafe/outdoor seating area	Person walking on a pavement	Person walking and entering a building
Person walking toward a car/trees	Person walking on a pavement	Person walking and entering a building
Person walking toward an outdoor market	Person walking in a parking lot	Person walking and entering a shop
Person walking toward a car	Person walking in a parking lot	Person walking and entering a shop
Two people walking toward a house	Two people walking on a green	Two people walking and entering a shop
Two people walking toward a house	Two people walking on a green	Two people walking and entering a shop
Two people walking toward a playground	Two people walking in a park	Two people walking and entering a mill
Two people walking toward a house	Two people walking in a park	Two people walking and entering a mill
Car driving toward a gas station	Car driving along the road	Car entering a garage
Car driving toward village	Car driving along the road	Car entering a garage
Car driving toward village	Car driving in the desert	Car entering courtyard gates
Car driving toward village	Car driving on a country road	Car entering a garage
Person running toward a river	Person running in the woods	Person running, entering a building
Person cycling toward shops	Person cycling on the road	Person cycling, entering an outdoor gate
Person cycling toward a village	Person cycling on the road	Person cycling, entering an outdoor gate
Person cycling toward shops	Person cycling in the woods	Person cycling, entering an outdoor gate
Person cycling toward a village	Person cycling in the woods	Person cycling, entering an outdoor gate