1 **On the road to somewhere:**
2 **Brain potentials reflect language effects on motion event perception**
3 *Monique Flecken[12], Panos Athanasopoulos[3], Jan-Rouke Kuipers[4] & Guillaume Thierry[5]*
4
5 [1] Max Planck Institute for Psycholinguistics
6 Wundtlaan 1
7 6525 XD Nijmegen
8 the Netherlands
9
10 [2] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition
11 Radboud University
12 Montessorilaan 3, 6525 HR Nijmegen
13 the Netherlands
14 Email: monique.flecken@mpi.nl
15 Telephone: 0031 24 35 21 347
16
17 [3]Department of Linguistics and English Language
18 Lancaster University
19 County South
20 Bailrigg
21 Lancaster
22 LA1 4YL
23 United Kingdom
24 Email: p.athanasopoulos@lancaster.ac.uk
25
26 [4]Psychology, School of Natural Sciences
27 University of Stirling
28 FK9 4LA Stirling
29 United Kingdom
30 Email: jk28@stir.ac.uk
31
32 [5]School of Psychology
33 Bangor University
34 LL57 2AS Bangor
35 United Kingdom
36 Email: g.thierry@bangor.ac.uk
37
38 **Abstract**
39
40 Recent studies have identified neural correlates of language effects on perception in static domains
41 of experience such as colour and objects. The generalization of such effects to dynamic domains like
42 motion events remains elusive. Here, we focus on grammatical differences between languages
43 relevant for the description of motion events and their impact on visual scene perception. Two
44 groups of native speakers of German or English were presented with animated videos featuring a dot
45 travelling along a trajectory towards a geometrical shape (endpoint). English is a language with
46 grammatical aspect in which attention is drawn to trajectory and endpoint of motion events equally.
47 German, in contrast, is a non-aspect language which highlights endpoints. We tested the
48 comparative perceptual saliency of trajectory and endpoint of motion events by presenting motion
49 event animations (primes) followed by a picture symbolising the event (target): In 75% of trials, the
50 animation was followed by a mismatching picture (both trajectory and endpoint were different); in
51 10% of trials, only the trajectory depicted in the picture matched the prime; in 10% of trials, only the
52 endpoint matched the prime; and in 5% of trials both trajectory and endpoint were matching, which
53 was the condition requiring a response from the participant. In Experiment 1 we recorded event-
54 related brain potentials elicited by the picture in native speakers of German and native speakers of
55 English. German participants exhibited a larger P3 wave in the endpoint match than the trajectory
56 match condition, whereas English speakers showed no P3 amplitude difference between conditions.
57 In Experiment 2 participants performed a behavioural motion matching task using the same stimuli
58 as those used in Experiment 1. German and English participants did not differ in response times

showing that motion event verbalisation cannot readily account for the difference in P3 amplitude found in the first experiment. We argue that, even in a non-verbal context, the grammatical properties of the native language and associated sentence-level patterns of event encoding influence motion event perception, such that attention is automatically drawn towards aspects highlighted by the grammar.

1. Introduction

Research studying language differences in the domain of motion events has relied mainly on behavioural measures obtained during verbal and non-verbal tasks. Various studies of language production, analyzing motion event descriptions elicited with pictures, animations, or storybooks show cross-linguistic differences in information selection and organization. Variation in the lexical and grammatical concepts available for expressing motion influences the extent to which speakers attend to specific elements of motion events, when construing utterances that describe them (cf. the 'thinking for speaking' hypothesis, Slobin, 1996). For example, speakers of languages that provide manner of motion verbs (e.g., English, 'to run/walk/stroll') will pay more attention to the aspects of the event that provide manner information, and they will encode this linguistically (e.g., Papafragou et al., 2008). Speakers of languages that provide verbs expressing path- or direction-information (e.g., French, 'entrer dans' to enter, 'se diriger vers' to direct oneself towards), pay less attention to manner; the linguistic encoding of manner of motion in sentences is optional (cf. Talmy, 1985). This questions the extent to which such language differences in event conceptualization may also show in non-verbal measurements, such as eye movements recorded during motion event scene perception, in the absence of explicit description instructions (Flecken et al., 2014; Papafragou et al., 2008), and in performance on various types of categorization, matching, and recognition tasks (e.g., Genarri et al., 2002; Kersten et al., 2010; Papafragou et al., 2002). Overall, results are mixed with language effects occurring primarily in paradigms involving verbally-mediated working memory (see Athanasopoulos & Bylund, 2013; Papafragou & Selimis, 2010).

Recently, researchers studying relativity effects in the domains of colour perception and object categorization have begun to use neurophysiological measures to gain insights into effects of language on visual processing (Boutonnet et al., 2013; Liu et al., 2010; Mo et al., 2011; Thierry et al., 2009). These studies provide the strongest support to date for the claim that conceptual distinctions expressed linguistically impact cognitive processing, at least in part automatically, and in the absence of explicit linguistic manipulations. Such language effects can be detected by event-related brain potentials known to index attention processing and stimulus evaluation. Here, we test whether previously attested behavioural differences in attention to and saliency of aspects of motion events can be characterised in a non-verbal task.

The cross-linguistic comparison is based on language differences relating to *temporal* properties of motion events. The type of events involves volitional motion of an entity (e.g., vehicle, person) along a trajectory (road, path), in the direction of potential endpoint-objects or locations (house, tunnel). Crucial for the analysis of information selection is the fact that the endpoints are not reached by the entities in motion in motion event stimuli; they are thus optional for selection, and dependent on the viewpoint of the speaker[1]. A growing body of literature studing linguistic encoding patterns shows how the presence or absence of *grammatical aspect* in a language affects the extent to which speakers pay attention to and describe endpoints of such motion events (v. Stutterheim & Nuese, 2003; v. Stutterheim, Carroll & Klein, 2009; v. Stutterheim et al., 2012): Speakers of non-

---

[1] The type of motion events thus differs from 'boundary-crossing' events, showing the reaching/entering of endpoints. Boundary-crossing events have been used frequently in cross-linguistic comparisons of motion event conceptualization in verb-framed versus satellite-framed languages, as this is the domain showing the largest contrast between the two language types (cf. Slobin, 2006).

aspect languages (that do not encode aspect systematically on the verb, such as Afrikaans, German or Swedish) show a linguistic bias towards action goals and motion event endpoints (Schmiedtova, von Stutterheim & Carroll, 2011; Bylund, Athanasopoulos & Oostendorp, 2013; e.g., 'a woman walks towards a building', when describing a stimulus showing a person walking in the direction of, but not reaching, a building). Speakers of aspect languages, i.e., languages that provide grammatical means to provide aspectual contrasts (e.g., the progressive in English, perfective and imperfective aspect in Russian; cf. Comrie, 1976; Dahl, 2000) tend not to mention endpoints of motion events when they are not explicitly focused and not unambiguously represented as part of the event ('a woman is walking [along the road/past the church, etc.]', for the same stimulus). Speakers of these languages thus take a different perspective on the same event: Aspect-language users (e.g., speakers of English) focus on the inner temporal contours of the situation and are thus more sensitive to the specific ongoing phase of the event depicted, thereby taking an 'inside' view of a situation ('a person is walking'). Non-aspect language users (e.g., speakers of German) typically take a holistic perspective (Athanasopoulos & Bylund, 2013; v. Stutterheim & Nuese, 2003) and refer more frequently to the endpoint of a motion event, when boundary-crossing is optional and dependent on the perspective taken by the viewer ('a person walks towards X'). The empirical evidence covers a wide variety of languages and shows the same interrelation between aspect and event endpoints in linguistic encoding, tested for the the non-aspect languages IsiXhosa, Afrikaans, Dutch, German, Swedish, and the aspect languages Arabic, English, Russian, Spanish (v. Stutterheim et al., 2012; Bylund et al. 2013; Bylund & Athanasopoulos, 2014, etc.).

To tap into conceptualization processes and attention-directing mechanisms more directly, recent studies have included measurements of co-verbal processing, such as the registration of eye movements (fixations on endpoints in motion scenes, design cf. Papafragou et al., 2008), in comparison to fixation patterns registered in non-verbal tasks (Flecken et al., 2014). Findings show a higher degree of attention allocated to event endpoints by non-aspect language users (German) compared to speakers of English or Arabic, in verbal as well as non-verbal tasks. Other recent studies have investigated non-verbal motion event categorization preferences using a triads-matching task. Speakers of non-aspect languages (Swedish) were more prone to rely on endpoints as the main criterion for event categorization (Athanasopoulos & Bylund, 2013). Interestingly, in a number of follow-up studies on bi- and multilinguals, this endpoint categorization-criterion was modulated by the specific degree of exposure to the aspect language (English) (see Bylund et al., 2013; Bylund & Athanasopoulus, 2014). Crucially, all the above-mentioned studies on grammatical aspect used naturalistic, dynamic motion event stimuli (live-recorded video clips), which arguably are easily amenable to implicit verbalization, as attested by the fact that the cross-linguistic differences were abolished under a verbal interference condition (Athanasopoulos & Bylund, 2013). At this point, we thus cannot make definitive claims regarding the online involvement of language, i.e., the use of verbal strategies when task conditions allow for it, versus the existence of different representations of events in speakers of different aspect systems, as would be assumed in strong relativistic views (Lucy, 1992, 1997; overviews in Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1996).

We take the aspectual contrast between German and English as our test case for the linguistic relativity hypothesis, and extend it to the neuropshyiological domain. In line with recent studies of language-perception interactions using ERPs (Athanasopoulos et al., 2010; Boutonnet et al., 2013; Thierry et al., 2009) we used a visual oddball paradigm to investigate language effects on perceptual and attentional processing of motion events. To our knowledge, this is the first study investigating language effects on perception beyond the word-level and rooted in a core grammatical category using ERPs.

In Experiment 1, we engaged native speakers of German and English in an event matching task, in which they saw animated schematic motion events followed by target pictures. On each trial, they were asked to pay attention to both the trajectory and the endpoint of the animation (e.g., a dot moving along a straight trajectory towards a square) and then match these features of the animation to those depicted in the target picture. Prime–target associations defined four conditions: Fully incompatible in terms of trajectory and endpoint (mismatch), fully matched for trajectory and

endpoint (full match), matched for endpoint but not trajectory (endpoint match), and matched for trajectory but not endpoint (trajectory match). The frequency of the conditions was manipulated in order to elicit a P3 ERP response in the full match (5% of trials), endpoint match (10%) and trajectory match (10%) conditions. The P3 component is known to reflect attentional processing, stimulus evaluation, and target detection (see Polich, 2007 for a review). Therefore, we expected to see largest P3 amplitudes in the full match condition (response trials), and P3 modulations in the endpoint and trajectory match conditions, since they were presented with a low local frequency and tempted participants to respond. Differences in P3 amplitude between the two partial match conditions would index the participants' attentional bias towards the specific motion element. In the German group, we anticipated a larger P3 in the endpoint than the trajectory match condition, indicating a higher degree of saliency or attention devoted to this feature. On the other hand, English speakers should devote similar levels of attention to endpoint and trajectory, leading to a P3 of similar amplitude in the two conditions.

In Experiment 2, native English and native German participants performed a similar non-verbal motion matching task, without an oddball manipulation (behavioural variant of the ERP paradigm). This time, the picture preceded presentation of the animation, and participants were instructed to indicate on each trial whether or not the animation fully matched the preceding picture. We analyzed the number of responses as well as reaction times. Experiment 2 was conducted to test whether potential attentional biases reflected in ERPs can be measured behaviourally during a higher order similarity judgement task. Performance on motion matching and similarity judgement tasks have shown that the involvement of verbal working memory may well be the cause of language effects (Athanasopoulos & Bylund, 2013). A comparison of results in the two tasks, given their specific conditions, will be informative with respect to the mechanism underlying language effects in non-verbal paradigms. This is especially interesting in light of current questions that relate to whether or not the language system exerts online (i.e., in the form of online verbal encoding processes which are at play also during non-verbal tasks) or structural (i.e., in the sense of representational differences in speakers of different languages) feedback to the perceptual system during processing. In both experiments we used schematic stimuli (basic shapes, repeated numerous times) that show abstract motion, in response to which participants were asked to make prompt judgements. In contrast to previous paradigms using real-world motion video clips, here, online event encoding would be challenging, given the speed of presentation, the absence of naturalistic, volitional motion, and the schematic nature of the stimuli. If language effects are not obtained in the behavioural variant of the non-verbal ERP paradigm, we can with more certainty state that any differences found in ERPs between the two groups of speakers concern purely non-verbal motion event cognition.

If German participants are biased towards motion event endpoints during visual processing, greater P3 amplitudes should be elicited by endpoint match than mismatch pictures, and crucially, than trajectory match pictures. If such attentional biases also affect overt judgements of the same schematic motion stimuli in a behavioural task, reaction times as well as false alarm rates could be affected: German participants could be slower on endpoint match trials, and they may be tempted to judge those as full match trials (resulting in false alarms). If language effects emerge in both experiments, this would imply that, for schematic motion stimuli, verbal event encoding processes are active, affect attention processing, stimulus evaluation, and overt comparison and similarity judgement processes. If, however, language effects are only obtained in the matching task, but are not reflected in the P3 component of ERPs, there are still reasons to assume that online verbal encoding processes take place – results would imply that, in principle, the stimuli allow for verbal encoding regardless of their non-verbal nature. However, the implicit construal of a full-fledged event description may require more time than the time captured by the ERPs recorded here, which may thus capture processing stages prior to sentence generation. These stages could encompass 'purely' visual and / or attentional processing, potentially followed by lexical retrieval processes related to individual objects or elements of motion. Effects of the actual linguistic construal of events are then reflected in matching behaviour. If, as a third option, language effects are only

obtained in the ERPs, but not in its behavioural variant, the cause of the effect is less likely to be due to verbal encoding of the stimuli. Instead, this would render stronger grounds to assume that language drives an attentional bias in the visual domain in an entirely non-verbal context, implying that language can affect certain non-verbal levels of representation.

2. ERP experiment: Method

2.1. Participants

Twenty native speakers of English (students at Bangor University, UK), and twenty native speakers of German (currently studying at Radboud University, the Netherlands) took part in our study. All participants were right-handed and between the ages of 19 and 25. English participants were tested in Wales (13 female; M age 23, range 19 - 25), and German participants were tested in the Netherlands (11 female; M age 21, range 18 - 25).

Language background was evaluated by means of a questionaire. The native speakers of English reported basic to intermediate knowledge of a second language which was never German. The native speakers of German were intermediate to advanced speakers of Dutch, and all subjects reported some knowledge of English. Proficiency in Dutch was estimated at B2 level as rated by the Common European Framework of Reference for languages (Council of Europe, 2001), which is the level required for German and other foreign students to be able to study at Radboud University; they had all taken an intensive Dutch course before the start of the first semester of Psychology, instructed through the medium of Dutch. Knowledge of Dutch is not considered a confound given that Dutch largely overlaps typologically with German and does not encode progressive aspect grammatically on the verb for motion events (Behrens et al., 2013). Knowledge of English in German participants was assessed using two standardized proficiency tests (Oxford Quick Placement Test, QPT, 2001; DIALANG, Alderson et al., 2005); the average score on the QPT was 43.92 (out of 60, corresponding to B2 level), and the level captured by Dialang was similar (17 participants: B2, 3 participants: C1). None of the participants reported daily or frequent exposure to English; at the time of testing their dominant language was German with Dutch their second, late-learned language. All participants received course credits or cash for participation in the experiment, which lasted 1 to 1.5 hours.

2.2. Materials and procedure

In each experimental session, participants watched 492 prime-target pairs. Primes are one-second animated video clips depicting basic motion events, in which a dot moved along a trajectory towards a geometrical shape. The dot never reached the endpoint shape during the video play time. There were four different animations (see example in **Fig. 1**). In all cases, the target pictures provided a symbolic representation of the animated primes.
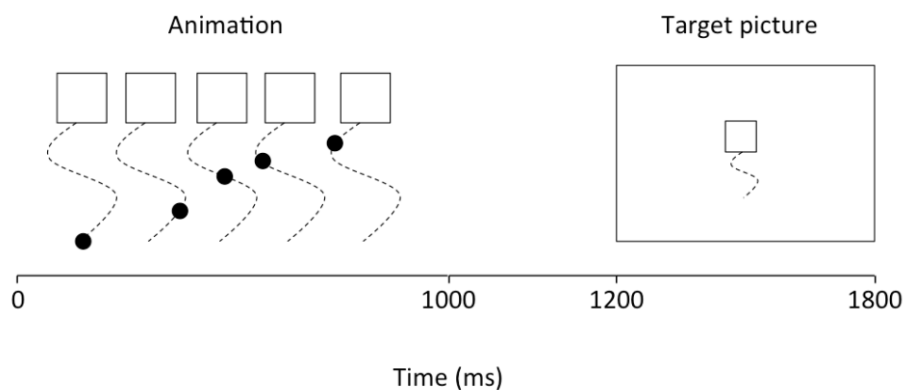
Figure 1: Example of animated prime and target picture used in the experiment (trial depicts a match condition (response required). Cartoon provides a representation of the animated sequence.

The experiment conformed to a classic oddball paradigm in which prime-target pairs belonged to four conditions: Only 5% of trials were full matches (24 trials). Seventy-five percent (372 trials) were full mismatch trials in which neither the trajectory nor the endpoint depicted in the target picture matched the preceding animated prime. In the remaining two conditions either trajectory or endpoint of the animated sequence matched the target picture and these conditions represented 10 % of trials each (see **Fig. 2**). Target pictures were presented on a white background on a 19-inch CRT monitor. Participants were seated about 1 m from the screen and stimuli subtended approximately 8 degrees of visual angle.

The animated prime was displayed for 1000 ms, followed by a blank screen for 200 ms and then the target picture for 600 ms. During the 800 ms inter-trial interval the screen was blank. Experimental conditions were presented in a pseudorandomized order such that rare trial types or deviants (full match and partial matches) were never presented in immediate succession and separated by 3-5 frequent trial types or standards (full mismatch). Participants were instructed to press a button in the full match condition (press the button only if the picture exactly matches the preceding video). Participants were also informed that the picture would look smaller than the video and only symbolise movement. The proportion of correct button presses was high in both groups and not different between groups (~90%).

Full match (response)   Full mismatch   trajectory match   Endpoint match
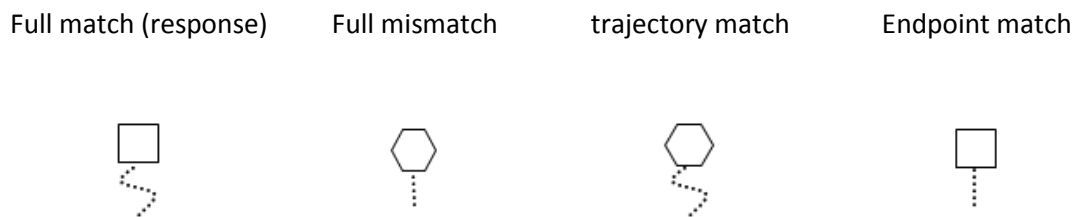
Figure 2: Examples of matching and mismatching target pictures for the animation presented in Figure 1

Both groups of participants were instructed in their native language by the same experimenter, a highly proficient speaker of German and English, assisted by a native speaker assistant in each case. Stimuli were run using the same Presentation (Neurobehavioral systems™) script, ensuring similar task conditions and the activation of a German or English language context in our participants (see for effects of language context on attention to motion elements, Lai et al. 2014).

2.2. EEG recording and analysis

For adequate sampling purposes, electrophysiological data were recorded in two different laboratories, ensuring that English participants were as close to monolingual as possible and ensuring homogeneity of the German group. English participants were tested in Bangor, Wales (Bangor University) and data acquisition was implemented with Neuroscan 4.4™. EEG was recorded in reference to Cz. German participants were tested in Nijmegen, the Netherlands (Donders Centre for Cognition). Given the local technical procedures and system parameters, the data were recorded using BrainVision Recorder 1.1™ in reference to the left mastoid. In both laboratories, sampling was set to 1 Khz and data were recorded from 32 electrodes placed according to the 10-20 convention. Impedances were kept below 5k Ohms. EEG activity was filtered offline with a bandpass zero phase-shift filter (0.1 Hz, 12 dB/oct - 30Hz, 48 dB/oct). Eye blinks were mathematically corrected based on the procedure advocated by Gratton, Coles and Donchin (1983). Automatic artefact rejection discarded all epochs with an activity exceeding +/- 75 microvolts. Individual ERPs were computed

from epochs ranging from -100 to 1000 ms after stimulus onset and baseline corrected in reference to 100 ms of prestimulus activity. Individual averages were then re-referenced to the average of the left and right mastoid sites to make the two datasets fully comparable.

P3 analysis was based on individual ERPs elicited in the four conditions. The P3 was maximal over central-parietal scalp areas in both groups and studied at electrodes Cz, CP1/CP3, CP2/CP4, P3, Pz, P4, reflecting a typical P300 scalp topography (see, e.g., Polich & Kok, 1995) (alternative electrodes sites reflect differences between electro-caps used in Wales and the Netherlands). Visual inspection of the maximum P3 effect for the full match response condition in relation to the frequent mismatch condition showed a slightly earlier peak in the English dataset (around 520ms post stimulus onset) compared to the German dataset (around 610ms). To ensure that general task processing and attention devotion (related to spotting full match trials) was similar in each group, a condition analysis was performed on the time window of the maximum P3 effect for the response condition in each group. Furthermore, because of our interest in the positive peaks post 300 ms for the two critical conditions, an additional time window was analyzed. To incorporate for each group the peaks for the full match condition, but especially also the peaks for the two critical conditions (endpoint match and trajectory match), a further condition by group analysis of the P3 component was performed on a broader time window covering 350 to 700ms after stimulus onset.

Mean amplitudes during the time span of the maximum P3 effect for the full match condition compiled from 6 electrode sites (Cz, CP1/CP3, CP2/CP4, P3, Pz, P4) were analyzed using a repeated measures ANOVA with condition (full match, endpoint-match, trajectory-match, mismatch) as a factor for each group separately. Mean P3 amplitudes between 350 and 700ms averaged across the same 6 electrodes were subjected to a mixed 4-by-2 ANOVA, with condition (full mismatch, endpoint-match, trajectory-match, full match) as within-subjects factor and language (German, English) as between-subjects factor. In addition, millisecond-by-millisecond paired samples t-tests comparing mean amplitudes in the endpoint-match and trajectory-match conditions were computed for the entire segment, to determine more precisely the exact time windows of differences between critical conditions in each group.

In the condition analysis in the time spans around the maximum P3 peak for the full match condition in each group, we expected significantly larger P3 amplitudes for the full match condition, compared to all other conditions, reflecting a similarly high degree of attention devoted to the task of detecting full matches between animation and picture in German and English participants. This analysis is important in ensuring similar patterns with respect to task-related attention in our participants. We were also interested to find out whether there were any differences between critical conditions during this window, even though the time windows do not necessarily cover the P3 peaks for the critical conditions.

In the 350 to 700ms time window covering P3 peaks also for the two critical conditions we expected a significant main effect of condition, again given the oddball manipulation (full match condition being the most infrequent and the reponse condition, expected to elicit the largest P3 amplitude), but also a language-by-condition interaction, indicative of differential responses to the trajectory and endpoint match conditions in the two groups.

3. ERP Results: P3

Figure 3 below plots P3 amplitudes, subtracting the frequent (mismatch) condition from the rare response condition (full match) in each group, visualizing the P3 effect for the response condition. In the English dataset, the peak occurs around 520ms and the analysis was performed on a time window covering 100ms preceding and following this peak (420-620ms). In the German dataset, the peak appears to be later (around 610ms) and the analysis was performed on the time window covering 510 to 710ms.
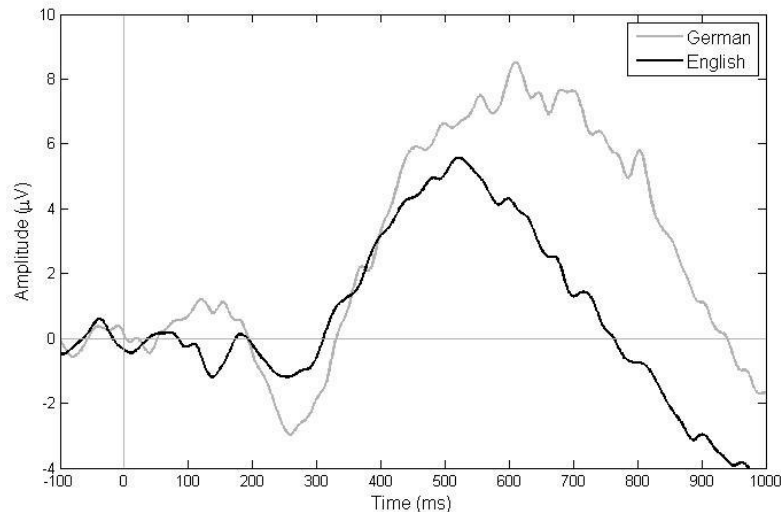
For the English data, a repeated-measures ANOVA showed a significant condition main effect, $F(1.954, 37.133)[2] = 25.211$, $p<.001$, $\eta^2_p = 0.570$, with Bonferroni corrected posthoc tests showing that the P3 elicited in the response condition (full match) was significantly larger than that elicited in all other conditions (full match vs. mismatch: $p<.001$; vs. trajectory match: $p<.05$; vs. endpoint match: $p<.05$). The P3 for the two critical conditions in this window were also significantly more positive than the P3 for the mismatch condition (both comparisons $p<.05$). There was no difference between the trajectory match and endpoint match conditions ($p=1.00$, n.s.). In the German dataset, a similar pattern was found ($F(1.775, 33.731) = 23.842$, $p<.001$, $\eta^2_p = 0.557$): the full match condition elicited most positive P3 amplitudes (full match vs. full mismatch $p<.001$; vs. trajectory match $p<.05$; vs. endpoint match $p<.05$), the two critical conditions were in between the mismatch and full match conditions, with no significant difference between the two ($p=1.00$, n.s.).

Figure 4 below plots P3 amplitudes for all conditions for the entire segment. P3 amplitude analyses in the 350 to 700ms time window rendered a main effect of condition ($F(1.901, 72.228) = 41.946$, $p<.001$, $\eta^2_p = 0.525$) and a main effect of group ($F(1,38) = 17.991$, $p<.001$, $\eta^2_p = 0.321$), such that the P3 for the full match condition was larger than that for the three other conditions, and the German participants showed more positive P3 amplitudes overall. Importantly, the condition by group interaction was significant ($F(1.901, 72.228) = 3.057$, $p<.05$, $\eta^2_p = 0.074$).

To explore the interaction further, a separate condition analysis was conducted for each group in the 350 to 700 ms time window. In the German group, there was a condition main effect ($F(1.851, 35.165) = 23.209$, $p<.001$, $\eta^2_p = 0.550$): Posthoc tests showed that the full match condition elicited more positive P3 amplitudes than the other three conditions (all comparisons $p<.001$). P3 amplitudes for the endpoint match condition were more positive than the mismatch condition ($p<.05$), but there was no difference between the trajectory match and the mismatch condition ($p=.073$, n.s.). Crucially, the endpoint match condition differed from the trajectory match condition ($p<.05$). In the English group, we also found a main effect of condition ($F(1.900, 36.093) = 20.701$, $p<.001$, $\eta^2_p = 0.521$). The P3 for the full match condition was significantly more positive than the full mismatch P3, as well as the P3 elicited by the two critical conditions (all comparisons $p<.001$). The mismatch condition furthermore differed from the two critical conditions (all comparisons $p<.05$). Critically, the endpoint match and trajectory match conditions were not significantly different from one another ($p=.766$, n.s.).

---

[2] In some analyses, the degrees of freedom and p-values were Greenhouse-Geisser corrected, as Mauchly's test of sphericity reached significance, indicating a violating of the sphericity assumption.

Millisecond-by-millisecond t-tests support and specify the analyses on average amplitudes for the 350-700ms time window: Whereas in the English group only a small time window in which the endpoint match condition displays higher amplitudes than the trajectory match condition could be identified (350 - 396 ms), this pattern is significant throughout most of the analyzed P3 time range for the German participants, and even for a short time window (from 278 – 302 ms) before the actual P3 peak. In English participants, endpoint match trials thus drew more attention than trajectory match trials only for a brief interval early in the P3 range, whereas the German participants differentiated between the two critical conditions throughout the entire time span of stimulus evaluation.
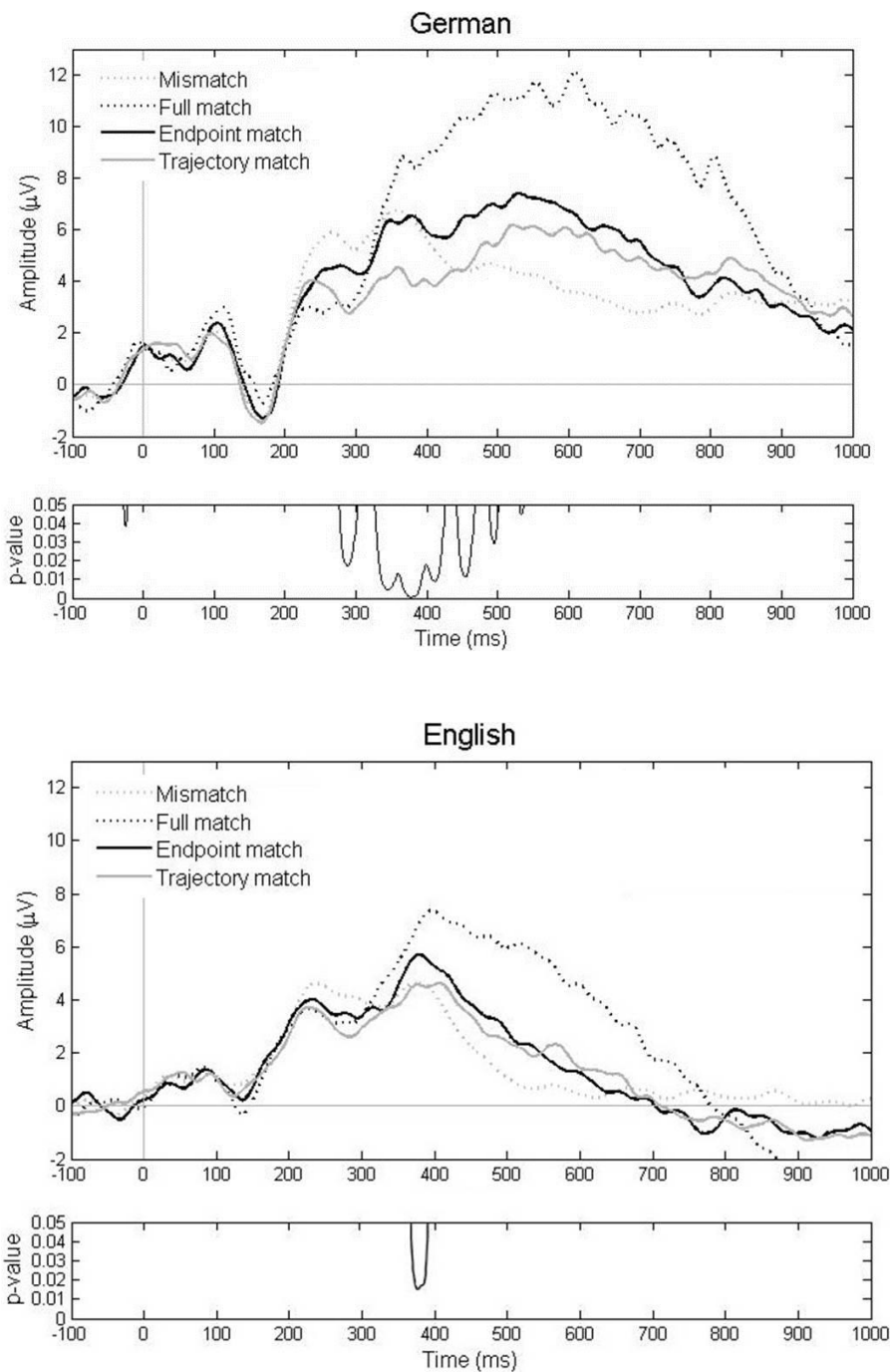
Figure 4 Top panel (German); Bottom panel (English): (Top) ERP peak of the P3 component (linear derivation of electrodes Cz, CP1/CP3, CP2/CP4, P3, Pz, P4) in the English and the German group. (Bottom) p-values of millisecond by millisecond t-tests, comparing amplitudes of trajectory match and endpoint match conditions for the entire segment (with the current degrees of freedom, a t-statistic of 2.093 was considered statistically reliable at $p=.05$).

9

1
2 In addition, average amplitudes in the P1-N1 range were also analyzed in a similar fashion (in line
3 with Thierry et al. 2009; Boutonnet et al. 2013), rendering no main, nor interaction effects.
4
5 4. Behavioural motion matching task: Method
6
7 4.1 Participants
8
9 Native English-speaking participants were students at Reading University (UK) and tested there
10 (N=15). German participants were again recruited and recorded at Radboud University, Nijmegen,
11 the Netherlands (N=19; both groups balanced for gender). Participants in this experiment had similar
12 educational and socio-economical backgrounds as the participants tested in Experiment 1. They were
13 recruited from the same population of German students of Psychology at Radboud University.
14
15 4.2 Materials and Procedure
16
17 The same materials were used as in Experiment 1. This time, participants watched sequences of a still
18 picture implying a motion event (600ms), followed by a short motion event animation (1000ms) with
19 a blank screen (200ms) in between. The inter-trial interval was a 800ms blank screen. Each animation
20 was preceded by each picture (thus giving 16 combinations in total), and each condition was
21 repeated 10 times, rendering 160 trials in total. There were 40 trials per condition (full match,
22 mismatch, endpoint match, trajectory match). Experimental conditions were presented in a fully
23 randomized order and run using a Psychopy script (Peirce, 2007).
24 Participants were instructed to press one of two buttons on each trial to indicate whether the
25 animation fully matched the preceding picture, yes or no. The order of picture and animation was
26 reversed when compared to Experiment 1, giving participants sufficient time to respond by pressing
27 a button[3]. Reaction times and response types ('yes' or 'no') were registered.
28
29 4.3. Results[45]
30
31 We analyzed accuracy, overall reaction times, as well as reaction times on correct responses to
32 mismatch (rejection) trials ('no' responses for the two partial match and the mismatch conditions),
33 and false alarm rates, with mixed ANOVAs of group by condition.
34 Table 1 below lists the average proportion of correct responses for each condition in each group.
35
36 Table 1: Accuracy (average proportion of correct responses for each condition)

Condition	German	English
Full match	84.07% (11.15)	88.18% (8.85)
Mismatch	91.56% (13.80)	96.18% (5.58)
Endpoint match	90.53% (11.23)	93% (6.70)
Trajectory match	71.05% (37.05)	86% (24.83)

37
38 A mixed ANOVA of group by condition rendered a significant effect of condition ($F(1.130, 36.160) =$
39 5.711, $p<.05$, $\eta^2_p = 0.151$), no main effect of Group ($F(1,32) = 2.703$, $p=.110$, $\eta^2_p = 0.078$, n.s.) and no
40 interaction between the two factors ($F(1.130, 36.160) = .981$, $p=.339$, $\eta^2_p = 0.030$, n.s.). Overall, there
41 was a higher correct performance for the mismatch condition than the full match condition ($p<.001$,
42 all p-values Bonferroni corrected), no difference between the mismatch and the endpoint match
43 conditions ($p=.605$, n.s.), and a marginal difference between the mismatch and trajectory match

---

[3] The time of picture presentation following the design of Experiment 1 would be too short for planning and
executing manual responses.
[4] 18 trials showed missing responses and were removed from the analyses.
[5] Degrees of freedom and p-values were Greenhouse-Geisser corrected when necessary.

conditions, with fewer correct responses in the latter condition (p=.054). There was no difference between the critical conditions (p=.140, n.s.).

Table 2 below gives mean reaction times for all responses.

Table 2: Mean reaction times on all responses (ms), mean (SD)

Condition	German	English
Full match	621 (105)	676 (133)
Mismatch	581 (103)	627(123)
Endpoint match	629 (110)	668 (139)
Trajectory match	631 (115)	666 (131)

The mixed ANOVA showed a main effect of condition (F(3, 96) = 14.043, p <.001, η^2_p = 0.305), no main effect of group (F(1, 32) = 1.199, p=.282, η^2_p = 0.036, n.s.) and no interaction between the two factors (F(3, 96)= .558, p=.644, η^2_p = 0.017, n.s.). Overall, reaction times were fastest for mismatch trials; there were no differences between the critical conditions.

Table 3 shows mean reaction times for correct responses to motion mismatch trials ('no' responses for mismatch and partial match conditions).

Table 3: Mean reaction times for correct motion mismatch trials ('*no*' responses for mismatch, endpoint match and trajectory match conditions) (ms), mean (SD)

Condition	German	English
Mismatch	581 (109)	627 (124)
Endpoint match	627 (119)	664 (141)
Trajectory match	661 (125)	652 (142)

A mixed ANOVA on reaction times for correctly identified mismatches rendered a significant main effect of condition (F(1.581,50.599) = 12.636, p<.001, η^2_p = 0.283), no main effect of group (F(1,32) = .353, p=.557, η^2_p = 0.011, n.s.), but a significant condition by group interaction (F(1.581,50.599) = 3.603, p<.05, η^2_p = 0.101). Overall, participants were faster to decide that mismatch trials were not full match trials, compared to partial match trials (all comparisons p<.001).

In the German group, there was a significant condition effect (F(1.433, 25.796) = 14.327, p<.001, η^2_p = 0.443). Posthoc comparisons showed that participants were faster on mismatch trials, than on endpoint match (p<.001) and trajectory match trials (p<.05). There was no difference between the endpoint and trajectory match conditions (p=.196, n.s.). In the English group, there was only a trend for an effect of condition (F(2,28)=2.814, p=.077, η^2_p = 0.167).

Below, the average proportion of false alarms for the endpoint match and trajectory match, and the mismatch conditions are listed (Table 4). Note that yes responses on the full match trials are the only correct responses; positive responses registered for the other conditions as listed below are thus false alarms.

Table 4: False alarm rates (average proportion of *yes* responses for mismatch and partial match conditions)

Condition	German	English
Mismatch	6.58% (13.50)	2% (1.95)
Endpoint match	9.08% (11.15)	5.33% (5.18)
Trajectory match	28.03% (36.55)	12.68% (24.70)

A mixed ANOVA on the number of false alarm button presses revealed a significant condition effect (F(1.057,33.812) = 7.227, p<.05, η^2_p = 0.184) and a non-significant trend for a group effect (F(1,32) = 2.987, p=.094, η^2_p = 0.085). The interaction of condition by group did not reach significance (F(2,64) = 1.036, p=.320, η^2_p = 0.031, n.s.). Overall, there were more false alarms for endpoint and trajectory match trials than mismatch trials (mismatch vs. endpoint match: p<.05, mismatch vs. trajectory

match: *p*<.05). There was a non-significant trend for more false alarms in German participants than in English participants, for the trajectory match condition.

The behavioural results show no significant group differences or interactions of group with condition with respect to overall performance on the task, as reflected in accuracy rates and speed of processing. This shows that both groups perform roughly the same when making overt judgements related to the degree of match between motion pictures and animations: Performance was fastest and most accurate on the mismatch condition, and worse on the trajectory match condition. In German participants reaction times on rejection (motion mismatch) trials were slower for the two partial match conditions, compared to the complete mismatch condition.

5. Discussion

English and German differ in a domain of grammar, relevant for motion event encoding: In English, an aspect language, temporal contours and phases of events are highlighted, whereas in German, a non-aspect language, endpoints of motion events are focused in event encoding. We investigated whether these grammatical differences affect visual attention allocation to elements of motion events in a non-verbal context using brain and behavioural measures.

The visual oddball paradigm used elicited typical P3 brain responses to the rare response condition (full match) in both groups. The maximum difference between the P3 for the full match and the mismatch (frequent control) conditions, i.e., the P3 effect for the response condition, occurred slightly earlier in the English dataset. Overall, the P3 peak in the English data seems steeper and more short-lived than the P3 peak in the German dataset, which lasts longer and is more positive in general. Findings from the behavioural task show that English participants were not overall faster in performing the task of detecting full match trials (bearing in mind that a different participant sample was tested); however, we did find German participants to be slower at rejecting partial mismatch (in endpoint or trajectory) trials. These results could tentatively indicate general lower abilities in inhibiting responses to distracter trials in German participants. This could, at least in part, explain the slightly differing shapes of the P3 waves. Nevertheless, the important finding here is that during the task-relevant peak (full match peak time window) all conditions pattern similarly in each group. This shows that task processing and the attention devoted to the execution of the task was very similar in both German and English participants, a necessary precondition for our interpretation of differences related to the two critical conditions. The next set of analyses aimed at unravelling potential language effects in relation to the critical conditions: In the 350-700ms time window which covers the P3 peak for the task-relevant condition (full match), as well as the P3 peaks for the critical endpoint match and trajectory match conditions, we find that, in German participants, the endpoint match condition elicited a significantly more positive P3 than the trajectory match condition. This indicates that in this group, the endpoint was processed with more attention, and perceived as more relevant or salient than the trajectory when matching animations with pictures. In addition, we found that the differentation between the endpoint and trajectory match conditions already started before the analyzed P3 time windows in German participants.

In English participants, there were no sustained differences in P3 amplitude between endpoint and trajectory match conditions, suggesting that there was no attentional bias and that both elements were similarly attended to. By comparing each oddball condition to the mismatch (control) condition, we can evaluate actual P3 condition effects. P3 effects were obtained for the full match and endpoint match conditions in both groups, but only in English participants a P3 effect for the trajectory match condition was found. The P3 effect obtained for the full match condition again underlines equal task processing and attention as this is the only condition which is relevant for the task and the condition which participants were explicitly instructed to attend to. The P3 effect for the endpoint match condition in both groups may reflect a potential language-independent bias towards motion event endpoints, which has been reported previously in verbal as well as non-verbal task paradigms on motion event processing (see e.g., Zacks & Tversky 2001; Slobin 2006). However, the attentional bias towards endpoints over trajectories of motion lasts for a longer interval in German

participants. English participants show an endpoint-over-trajectory bias for a brief interval early during the P3 window analyzed. In addition, we also find a P3 effect for the trajectory match condition in English, meaning that this effect is present for *both* partial match conditions, underlining that no differentiation is made at the level of attention processing between the two elements of motion events in this group. The different patterns within each group are thus crucial in interpreting these findings as a language-derived effect on attention processing.

These P3 findings suggest that individuals transfer grammatically driven ways of speaking about event scenes to non-verbal visual processing. Importantly, this language effect concerns complex grammatical structures which are relevant for perspective-taking and sentence structure, rather than basic categories such as colour, related to terminological differences between languages, and which have previously been shown to affect perception (Thierry et al, 2009). Here we show that such language patterns affect online visual attention allocation. Importantly, there were no differences in participants' judgements of the similarity of the same abstract depictions of motion, and no differences in the speed with which judgements were made, providing evidence for a conception of the stimuli as non-verbal. This makes the potential online construal of sentences describing the events (verbal encoding strategies) unlikely. We argue that performance on an overt stimulus matching task without concurrent linguistic interference can be prone to verbalization strategies and reliance on verbal working memory to aid the matching of pictures to preceding animations, as previously proposed (Athanasopoulos & Bylund, 2013; Trueswell & Papafragou, 2010); the present null result on the behavioural task thus indicates that the abstract schematic stimuli are non-verbal, i.e., they are not amenable to verbalization strategies related to event encoding. This is hypothetically caused by a lower real-life value of the animations, a necessary concession when recording ERPs to avoid signal contamination by eye movements. The processing time normally required for the conceptualization and formulation of full-fledged event descriptions supports our assumption regarding the absence of verbal event encoding strategies (cf. Griffin & Bock, 2000).

Taken together, these results have implications for existing theoretical accounts of the mechanisms underlying language effects on perceptual and attentional processes, which assume online feedback from the language system (the 'label-feedback hypothesis', cf. Lupyan, 2012, theorizing online modulation of perception caused by linguistic labels), and accounts that assume structurally different representations in participants of different cultural and linguistic backgrounds (a strong relativistic view, see Gumperz & Levinson, 1996; Lucy, 1992). We hypothesize that the specific demands of the non-verbal ERP paradigm, requiring rapid inspection and explicit evaluation of the animation sequences, might enhance the likelihood of involvement of automatized routines of visual processing. These routines draw on motion representations, focusing more attention to endpoints in German participants, arguably because endpoint match trials represent more compatible or typical trials, i.e., they share typical values on dimensions learned to be important when speaking about motion. In this interpretation, the biasing role of language in attention processing would be durable rather than transient (not to say that it cannot be overriden given specific task conditions, this is something for future research to explore): Linguistically-entrained processing routines are recruited when task demands allow for it / require it in an automatic fashion; they are based on structurally different (non-verbal) motion representations (in line with views put forward for grammar specifically, Lucy 1992; 1997).

There is a possibility that the activation of language-dependent motion representations resulted in the retrieval of linguistic forms (labels) encoding specific motion elements, i.e., the endpoint- and/or trajectory-shapes, in German but also English participants. The low degree of variability in shape-types, as well as the one second time window of animation viewing, would leave room for lexical retrieval of labels (e.g., 'square', 'hexagon'; 'straight', 'curved'). Such a verbal strategy could have been employed either automatically, or strategically, to aid task fulfillment: Memorizing labels for the endpoints in the animations, for example, would allow for efficient and fast comparison with the objects in the pictures and indeed, effects of labels on perception and categorization of objects have been reported previously (Lupyan & Ward, 2013; Lupyan & Casasanto, in press). Also, studies in the field of perception show that lexical information can be retrieved very

rapidly and can affect perceptual processing (review in Hermann et al. 2010). The observation that this process may take place specifically in relation to endpoints in German participants, but that English participants make no difference between endpoints and trajectories is thus crucial. Indeed, if the participants resorted to labelling trajectory and/or endpoint, such implicit verbalisation would still be oriented by the grammatical characteristics of the native language. An interesting follow-up study could make use of motion stimuli consisting of unsual shapes, which cannot readily be labelled in the specific languages tested, thus blocking the fast retrieval of labels from the mental lexicon; this would allow a clearer view on the potential role of label feedback. Another line of follow-up research could aim at conducting the same experiment, this time using realistic motion event stimuli (pictures of events in the real world), carefully controlling for potential EEG artefacts. Nevertheless, the present absence of language effects in the behavioural variant of the task is strongly in support of the idea that our results involve non-verbal motion event cognition, with no room for online verbal encoding of a motion event. Moreover, our neurophysiological evidence related to attention processing does not cover the time span nor the polarity of components typically associated with semantic processing (N400 time window, cf. Kutas & Federmeier, 2011).

6. Conclusions

All in all, we found electrophysiological evidence that participants attended to motion non-verbally in a way that reflects their habitual encoding of motion events for verbalization, thus providing evidence for language modulation of non-verbal perceptual processes, and extending the case made for objects and labels (for colour and object terminology, Boutonnet et al., 2013; Thierry et al., 2009) to grammatical and sentence-level information, i.e., the structural properties of language. Viewing an object has been shown to immediately activate a linguistic label; here, the viewing of a motion event rapidly activated a representation which involved goal-orientation in German ('where to?') and which appeared to be less holistic and more 'immediate' in English ('what's going on now?'). In general, the present study is in line with work showing how semantic knowledge can be accessed fairly rapidly during time spans associated with attention processing (see, for effects of semantic coherence on scene perception reflected in evoked gamma band responses, Oppermann et al., 2012; review in Herrmann et al., 2010), especially when the system is biased towards specific knowledge types, given priors and prior expectations (see for perspectives from neuroimaging, e.g., Kok et al., 2013; Rahnev et al., 2011). Functional imaging research (Kok et al., 2013), for example, showed that participants' prior knowledge regarding the characteristics of an upcoming stimulus (in the present case, a trajectory and endpoint of an abstract motion event) resulted, upon viewing, in the immediate integration of this prior information with the available sensory input in the visual cortex of the brain, rather than in other areas that are associated with perceptual decision-making, activated later down the processing stream. The fact that we find a highly specific attentional bias (i.e., only for endpoints in the German speakers) suggests a role for a 'long-term' prior, i.e., a long-term memory representation for the domain of events associated with and entrained by language. Our evidence suggests that such long-term prior knowledge may also be integrated fairly rapidly into the perceptual processing stream. For the first time, we show effects related to complex linguistic knowledge, in the domain of grammar and its perspectivizing function for motion events, a well-researched area for the investigation of cross-linguistic differences.

References

Alderson, C., & Huhta, A. (2005). The development of a suite of computer-based diagnosis tests based on the Common European Framework. *Language Testing, 22 (3),* 301.

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

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

Behrens, B., Flecken, M., & Carroll, M. (2013). Progressive attraction: on the use and grammaticalization of progressive aspect in German, Norwegian and Dutch. *Journal of Germanic Linguistics, 25 (2),* 95-136.

Boutonnet, B., Dering, B., Vinas-Guasch, N., & Thierry. G. (2013). Seeing objects through the language glass. *Journal of Cognitive Neuroscience, 25 (10),* 1702-1710.

Bylund, E., Athanasopoulos, P. & Oostendorp, M. (2013). Motion event cognition and grammatical aspect. Evidence from Afrikaans. *Linguistics, 51,* 929-955.

Bylund, E. & Athanasopoulos, P. (2014). Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition, 17,* 431-441.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems.* Cambridge: CUP.

Council of Europe. (2001). Common European Framework of Reference for languages: Learning, teaching and assessment. Cambridge: Cambridge University Press.

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

Flecken, M., v. Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language & Cognition 6,* 45-78.

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

Gentner, D. & S. Goldin-Meadow (Eds.) (2003). *Language in mind: advances in the study of language and cognition*. Cambridge, MA: MIT Press .

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

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

Herrmann, C., Fründ, I., & Lenz, D. (2010). Human gamma-band activity: A review on cognitive and behavioral correlates and network models. *Neuroscience and Biobehavioral Reviews, 34,* 981–992.

Kersten, A., Meissner, C., Lechuga, J., Schwartz, B., Albrechtsen, J., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General, 139(4),* 638-653.

Kok P., Brouwer G., van Gerven M., & de Lange F. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience, 33,* 16275-16284.

Kutas, M. & Federmeier, K. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psycholog*y, *62,* 621-647.

Lai, V., Rodriguez & Narasimhan, B. (2014). *Bilingualism: Language and Cognition, 17,* 139-152.

Liu, Q., Li, H., Campos, J., Teeter, C., Tao, W., Zhang, Q., & Sun, H. (2010). Language suppression effects on the categorical perception of colour as evidenced through ERPs. *Biological Psychology, 85,* 45–52.

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

Lucy, J. (1997). Linguistic relativity. *Annual Review of Anthropology, 26,* 291-312*.*

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3,* 54.

Lupyan, G. & Ward, E. (2013). Language can boost otherwise unseen objects into visual awareness. *PNAS, 110(35)*, 1419-201.

Lupyan, G. & Casasanto, D. (in press). Meaningless words promote meaningful categorization *Language & Cognition.*

Mo, L., Xu, G., Kay, P., & Tan, L. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *PNAS, 108,* 14026-14030.

Oppermann, F., Hassler, U., Jescheniak, J., & Gruber, T. (2012). The rapid extraction of gist – Early neural correlates of high-level visual processing. *Journal of Cognitive Neuroscience, 24,* 521-529.

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

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

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

Polich, J. (2007). Updating P300: an integrative account of P3a and P3b. *Clinical Neurophysiology, 118,* 2128-2148.

Quick Placement Test (2001). Oxford: OUP.

Rahnev D., Lau H., & de Lange F. (2011). Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *Journal of Neuroscience, 31,* 10741-10748.

Schmiedtová, B., v. Stutterheim, C., & Carroll, M. (2011). Implications of language-specific patterns in event construal of advanced L2 speakers. In A. Pavlenko (Ed.), *Thinking and Speaking in Two Languages* (pp. 66-107). Clevedon: Multilingual Matters.

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

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

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

v. Stutterheim, C., & Nüse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics, 41 (5),* 831-881.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Vol.3: Grammatical Categories and the Lexicon* (pp. 57-149). Cambridge: CUP.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, JR. (2009). Unconscious effects of language-specific terminology on pre-attentive color perception. *PNAS, 106 (11),* 4567–4570.

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3-21.