Utrecht Institute of Linguistics OTS
Utrecht University

The result is in sight: effects of grammar on event encoding in Estonian and Dutch

Master Thesis

presented by

Maria Sakarias

The study was supervised by

dr. Monique Flecken
Max Planck Institute for Psycholinguistics
Neurobiology of Language Group
Nijmegen, The Netherlands

prof. dr. Frank Wijnen
Utrecht Institute of Linguistics OTS
Utrecht, the Netherlands

Student number: 5531187
Programme: Research Master's in Linguistics
Deadline for submission: August 31, 2016

ACKNOWLEDGEMENTS

The thesis is the result of my internship the Neurobiology of Language Department of the Max Planck Institute for Psycholinguistics. First and foremost, I am extremely grateful to my supervisor Monique Flecken, who has made this whole project possible. She enthusiastically agreed to take on rather greedy experiments, where we ended up eye-tracking over a hundred participants in two different countries. As she closely guided me through every step of the project, I gained an incredible amount of experience and knowledge about research. I could not have asked for a better internship or supervisor, and hope that we can work together again in the future.

This thesis marks also the end of my Research Master programme at Utrecht University. I would like to thank prof. Frank Wijnen for providing valuable feedback on my term paper, internship report and finally, the thesis. I am also grateful to our lecturers who have shared so many inspiring ideas, keeping my passion for linguistics alive, and to our tutors, for putting together a great MA programme. The two years were challenging and involved a serious amount of work – luckily, it was always accompanied by having fun with the linguistics crew at the Janskerkhof basement. Thanks for all the good memories!

Finally, I thank my colleagues and new friends from the MPI, Nicole and Anne, for making rainy days in Nijmegen a lot brighter. Cheers to my closest friends in Utrecht, who have stuck around despite my crazy busy life: Philipp, Lill Eva, Jess, Petra, Hanna, Arthur and Maaike. I also thank my family and friends back home for sending me support and chocolate. And Tijs, for always being there for me.

CONTENTS

ABSTRACT

This study investigated the effects of task demands (verbal or nonverbal encoding) and native language (Estonian or Dutch) of the viewer on visual attention allocation to and memorization of dynamic events. Most previous research has focused on lexical-semantic categories; not much is known about the effects of grammar on cognition. Estonian is an excellent testing ground, where grammatical case marking on the object noun undergoes an alternation depending whether or not the action ends with a result.

Participants were eye-tracked while viewing videos of two types of causative events (actors engaged in resultative/non-resultative events involving an object) which were depicted as either finished or unfinished. In the verbal encoding condition, participants were asked to describe each video clip in one sentence (e.g., "the woman drew a flower"). In the nonverbal encoding condition, they performed a distracter task, which involved detecting sound cues while watching the videos. In a subsequent surprise memory test, they were tested on their recall of the end state (+/- finished) of the events.

For resultative actions, Estonian speakers marked the presence or absence of a result with the appropriate grammatical case-marker on the object noun. In this condition, this group also allocated significantly more attention to the visual area of the action/object during the final frames of the videos than Dutch participants. Estonian participants showed no such action-bias in the nonverbal condition; Dutch participants' attention was generally not influenced by task condition. Looking at memory data, Estonian participants were more accurate at recalling the end state of the events in the verbal than the nonverbal condition. Interestingly, this group showed higher accuracy for resultative compared to non-resultative events in both verbal and nonverbal conditions, while no effect of resultativity of the event could be detected in the accuracy scores for Dutch participants.

The observed visual attention patterns and memory test results for Estonian participants are consistent with the hypothesis that language effects emerge most clearly under overt linguistic encoding instructions (Slobin, 1996). In addition, improved performance for resultative events by this group in the nonverbal condition is taken to reveal a global influence of language on how events are habitually processed and memorized.

1 INTRODUCTION

A core component of perception is the ongoing decomposition of a stream of perceived activities into concrete and meaningful events (Zacks and Swallow, 2007; Newtson, 1973). In broad terms, an event is a segment of time at a certain location that is conceived by the perceiver to have a beginning and an end (Zacks and Tversky, 2001). One longer event can be composed of shorter subevents, e.g., *making a cake* can be viewed as *preparing the dough*, *putting the dough in a tray*, and *baking the cake*. Theories of event cognition postulate that when processing an event, one creates a working mental representation or a model of the event (e.g., Radvansky and Zacks, 2011). In addition to a temporal frame, this model is assumed to include information about the event's location in space, the people and objects that are involved, and the relations between these different elements. Multiple factors based on the viewer's task and perspective, and the visual properties of the scene itself clearly influence how we allocate attention to these various elements, but it has long been a topic of heated debate whether the language that we speak bears an effect on attention allocation as well. Considering that there are striking cross-linguistic differences in how various components of specific types of events can be segmented and mapped onto language (e.g., Talmy, 2000; von Stutterheim et al., 2012), it is an interesting empirical question whether language-specific properties shape how people perceive an event and its components, or whether mental representations and the online encoding of events are uniform across language communities.

Regarding the potential relationship between language and cognition (see reviews in Ünal and Papafragou, 2016; Wolff and Holmes, 2011), there are three main hypotheses that have received the most attention. According to the first one, often referred to as the *universalist approach*, human perceptual processing and linguistic representations are conceived of as distinct and dissociable; hence, the core aspects of perception are assumed to be largely similar across individuals and not to undergo influence by one's native language (e.g., Gleitman and Papafragou, 2005; Jackendoff, 1996; Pinker, 1995). This point of view is consistent with the aforementioned theoretical work on event cognition by Zacks and colleagues (Zacks and Tversky, 2001): differences in attention to and memory of specific event components are expected to arise due to potential general perceptual biases in human cognition, but not due to the linguistic background. Thus, speakers of different languages should attend to the components in the same order and to the same extent (see also Gleitman and Papafragou, 2005; Gleitman et al., 2007; Klemfuss et al., 2012; Lakusta and Landau, 2012).

The second approach, known as *thinking for speaking* (Slobin, 1996, 2003) assumes an interplay between language and cognition, but postulates that the relation is highly context-dependent: the specific lexical and grammatical structures of a language are expected to have an effect on event parsing and encoding only in the context of speaking, but not when engaged in non-linguistic cognitive processing. The theory assumes that certain domains are more "codable" and thus more "accessible" in some languages than others, or within the same language. He defines a more codable expression as one that is "short, and/or high frequency, and generally part of a small set of options in a paradigm or small small set of items" (Slobin, 2003, p. 161): for example, a concept expressed in a

single noun or verb (e.g., to run into the house) is assumed to be more codable than one expressed via a phrase or clause (e.g., to enter the house while running) The habituation over time to the more codable framing options in one's native language is expected to lead to automatized attention to the relevant event components in verbal communication.

The third approach, referred to as *linguistic relativity*, or the "Sapir-Whorf" hypothesis, states that the categories and distinctions that are available and frequently used in a language shape the speaker's experience of the world affecting both high and low-level cognitive processes such as perception, categorization, attention and memory (Lucy, 1992, 2011; Levinson, 1996; Boroditsky, 2011). In other words, it is predicted that an effect of language prevails not only when language is being used, but in non-linguistic contexts as well. The hypothesis is notoriously difficult to test (how to study language without language?), but after decades of stark stigmatization, it is now resurfacing in experiments where the physiological processes underlying perception are studied by measuring electrical responses from the nervous system (see Thierry, 2016 for a discussion of most recent developments). From the perspective of neuroscience, it does not make sense to assume that language and perception are modular and totally independent from each other, as the brain itself is organized as a highly interactive and dynamic system with continuous feedforward and feedback loops between different parts of the network (Thierry, 2016; Lupyan, 2012).

Interestingly, most of the studies on language-perception interaction have focused on semantic or lexical categories (i.e., the presence or absence of words for a specific concept). As Thierry (2016) points out, following (Lucy, 2011, p. 49) (see also Pavlenko and Volynsky, 2015), differences in *grammar* may in fact hold the most potential for uncovering effects of language on cognition: grammatical categories are obligatory in language use, meaning that the activation of the related concepts and viewpoints are necessarily highly automatized and highly frequent. Therefore, effects of grammar on non-linguistic cognition may be stronger than effects in relation to semantic distinctions. Moreover, grammar exerts influence beyond the level of single referents, as the information conveyed is often relational and concerns predicate- and sentence-level units, such as entire actions, situations and events. This means that effects of grammar have a potentially broader scope on perceptual processing compared to semantic categories.

The next section will list the topics and methods that have been the focus of research specifically in the domain of event perception and cognition, and elaborate on the studies that are relevant for the current investigation of cross-linguistic causative event perception.

2 PREVIOUS RESEARCH

In order to tap into the interplay between language and event perception, empirical studies have contrasted several typologically different languages and employed various experimental paradigms. Early studies looked at the effects of language on performance in offline behavioural tasks such as categorization or memory tests, which are completed after either verbal or nonverbal event encoding (e.g., Gennari et al., 2002; Papafragou

et al., 2002; Finkbeiner et al., 2002; Papafragou and Selimis, 2010; Filipović, 2011; Fausey and Boroditsky, 2011; Athanasopoulos and Bylund, 2013). Several recent experiments have complemented such offline behavioural tasks with online measures such as eye-tracking and EEG, which allow the investigation of cognitive processes as they are unfolding at a temporally fine-grained level (e.g., Papafragou et al., 2008; Soroli and Hickmann, 2010; Trueswell and Papafragou, 2010; von Stutterheim et al., 2012; Flecken et al., 2015a, Flecken et al. 2015b). The types of stimuli that have been used in both offline and online studies range from static pictures (Slobin, 1996) and cartoon clips (Papafragou et al., 2008) to videos of naturally occurring settings (Flecken et al., 2015b). The results that have been obtained to date are mixed, with some studies pointing to an effect of language only in the context of speaking (e.g., Gennari et al., 2002; Papafragou et al., 2002), while others find language effects in nonverbal contexts as well (e.g., Flecken et al., 2014b; Flecken et al., 2015a).

A vast majority of the above-cited literature has focused on the perception of spontaneous motion, and has mostly drawn on the typological dichotomy of manner vs path lexicon described by (Talmy, 1985, 2000). The dichotomy groups together "path-languages" such as Spanish, Greek and Turkish, where the path of movement is typically encoded in the verb (e.g., *enter, cross*), and the manner of movement is left implicit or expressed via an adverb, prepositional phrase or gerund (extra-verbal means, i.e., "satellites": *"los ninos entraron a la escuela (corriendo)"*, 'the kids entered the school (running)'). Conversely, in "manner-languages" such as English, German and Russian, the path is specified in a satellite and the manner in the main verb (e.g., *run, drive*: "*the kids ran to the school*"). In an influential study, Papafragou et al. (2008) compared the eye-gaze patterns and memory task performance by speakers of Greek and English, hypothesizing that the most characteristic way of describing events would result in heightened attention to the path component of a motion scene for Greek speakers (here, attention to path was measured by proportion of looks to the path endpoint region; see p. 166), and to the manner component for English speakers (looks to skates, skis etc.). Participants were presented with short animated clips of various motion events: each clip played for 3 seconds and remained on the screen as a still frame for two additional seconds. Encoding was carried out in a verbal and a silent, nonverbal context (note that participants were only made aware of the memory task in the nonverbal condition). Results from the verbal condition showed cross-linguistic differences in eye-gaze patterns during the first second of the video, where the Greek participants were more likely to look at the path endpoint region than the English participants. In the nonverbal condition, a difference only emerged once the video had frozen on the screen, but this time in the opposite direction: Greek speakers were equally likely to fixate on either region, but English speakers were more concerned with the path endpoint region. Results from a task testing memory of the presence or absence of endpoints in each clip showed near-ceiling performance in the verbal condition for both groups, but a poorer score in the nonverbal condition for the Greek, which the authors associated with the group's weaker interest at the path region during this encoding task. The results were interpreted as evidence for the universalist hypothesis, arguing that if the effects of language on perception are task-induced, they are transient and thus of no significance.

Another line of research has investigated differences in event perception that are induced by the presence or absence of aspectual distinctions in the grammar. Aspect refers to the temporal viewpoint under which an event is presented (e.g., ongoing or completed; cf. Comrie, 1976; Klein, 1994). Aspectual languages such as English, Russian, Greek and Arabic express these categories obligatorily via fully grammaticalized verbal morphology, and non-aspectual languages such as German, Dutch, Swedish and Norwegian mark aspect optionally via lexical elements or periphrastic adverbs and particles. It has been hypothesized that speaking an aspectual language and thus expressing aspect by default induces a habit of allocating attention on the durative, action-related properties of events (the currently 'ongoing' phase of the event, "*what is happening right now?*"), while speaking a non-aspectual language correlates with a reverse bias of attending to and including the goal or endpoint of an event in event representation (Slobin, 1996; Bylund, 2009; von Stutterheim and Nuse, 2003). Von Stutterheim et al. (2012) conducted a series of production and eye-tracking experiments with speakers of Arabic, Russian, English, Spanish (+aspect), contrasted with Czech, German and Dutch (-aspect). As predicted, the "-aspect" group mentioned the endpoints more often in their speech, and also fixated this region more and longer than the "+aspect" group (note that the differences in both measures only appeared when the depicted endpoints where not reached during the video, e.g., a woman walking towards a car, but not reaching it). Flecken et al. (2014b) took the study a step further and tested Arabic (+aspect) and German (-aspect) speakers with a non-linguistic, sound-cue distraction task, where the participants were required to watch motion videos and attend to loud beeps that occurred occasionally in a continuous sound stream[1]. The design of the task was motivated by the potential problem with the nonverbal paradigm in Papafragou et al. (2008), where specific instructions for a memory task may have overridden possibly subtle language-specific perceptual biases. Here, even when paying attention to the background sounds of the video instead of preparing to describe it, German speakers showed the same bias for fixating eye-gaze on endpoints of motion. This was taken to suggest that the conceptual implications of language-specific grammatical particularities may be "deeply entrenched and operate as a default in processing visual input" (p. 71).

Next to motion events, a limited number of studies have looked at cross-linguistic descriptions and perception of causative events, which involve an actor performing action on an another actor or object (e.g., peel a banana) (e.g., Klettke and Wolff, 2003; Wolff and Ventura, 2009; Ji et al., 2011). Taking into account the manner-path typology, a recent study by Bunger et al. (2016) studied the perception of caused motion events (e.g., kick a ball into a goal) by English and Greek speakers. Manner verbs are less frequent in Greek, and the combination of one with a resultative phrase is highly restricted (e.g., Giannakidou and Merchant, 1999); thus, the prevalent "compact" event packaging in English (to kick into) is less available in Greek. In this study, participants in both verbal and nonverbal encoding conditions were made aware of an upcoming memory task. Production patterns showed that Greek speakers indeed included information about both event components within one sentence to a lesser extent, but that the groups did

1 As this task was used in the current study, please refer to section 5.2.3 in this paper for more details.

not direct eye-gaze to the two components in a different manner from each other in either of the conditions (cf. Papafragou et al., 2008). These results were taken to demonstrate that the act of speech planning changes event perception, but not according to typologically distinct event encoding patterns – at least not according to the investigated pattern in causative event encoding. Note that the method and results of the memory test were not reported in the paper.

The domain of aspect has been investigated in the context of causative events as well: Flecken et al. (2014a) conducted an eye-tracking study based on the fact that in English, events are marked as specific as opposed to generic or habitual via aspectual morphology on the verb ("*the woman is baking cupcakes*" vs "*the woman bakes cupcakes every Sunday*"). Verbal aspectual markers are not available in German, where the speaker would instead highlight other aspects of the scene, such as properties of the actor, instrument, location or event type to mark that the description refers to a specific action taking place in the here and now (e.g., "*die (ältere) Frau bäckt Kuchen (in der Küche)*", 'the older woman bakes cakes in the kitchen') (Van Beek et al., 2013; Carroll and von Stutterheim, 2011). The results of a verbal event description task showed significantly different gaze allocation patterns before starting to speak, such that German speakers had an overall longer looking time at the visual area of the actor coupled with a higher number of fixations at this area between 1800- 1200 ms before utterance onset. English speakers, on the other hand, showed an earlier predominant increase in fixations to the area of the action. The study did not include a nonverbal paradigm nor a memory task.

To sum up, not only has recent cross-linguistic research obtained divergent findings, but even potentially comparable results seem to have been interpreted in different ways, i.e., a certain effect is considered interesting and attributed to the native language background in the one study, but downgraded in another. A consensus regarding the extent of language-based influence on either motion or causative event perception is far from being reached. It is indeed very hard to compare results from different studies due to large variation in encoding conditions (versions of a "nonverbal" task vary across studies or lack altogether) and in experimental methods (e.g., eye-tracking vs. memory task, static vs. dynamic stimuli). There is a gap in the current literature: it would be important to conduct a study that methodologically collects both eye-gaze and memory data from both verbal and (previously tested) nonverbal encoding conditions while being based on a theoretically well-motivated choice of typological contrast. Moreover, the domain of grammar seems to be the most promising test-case for effects of language on nonverbal event perception, as Flecken et al. (2014b) provides evidence that not only verbal, but also nonverbal event encoding may be influenced by language. This is the only study to date contrasting verbal and nonverbal experiments and targeting a grammatical contrast between languages; hence, the aim of the current study is to further explore cross-linguistic differences in grammar and its implications for online event encoding and memory.

3 CAUSATIVE EVENTS IN ESTONIAN AND DUTCH

An event component that is subject to interesting cross-linguistic variation is the *result* of a causative event. When a person is *"writing a book"*, they can either stop writing before the book is finished, or finish writing it and thus produce a result – a full-blown book. Information about the resultant state can be expressed by different means across languages. In English, a finite transitive sentence "*she wrote a book*" is unmarked for resultativity: in order to convey that the activity was completed and the result (the finished book) has been achieved, one could use a resultative particle ("*she wrote up the book*"), a verb such as *finish* with the gerund form of the verb ("*she finished writing a book*") or the present perfect construction ("*she has written a book*". To stress its incompleteness, it is possible to combine progressive aspect with a temporal adjunct ("*she was writing a book for a while*"), or use a conative prepositional construction ("*she wrote at the book*"). Dutch, the language of the study at hand patterns with English in that a sentence in simple past tense is unmarked for resultativity (see Boogaart, 1999 for a discussion of aspect and tense in both languages). Crucially to the current experiments, the use of linguistic elements to mark the status of an event as having led to a result is optional and carried out via peripheral constructions, suggesting that the related concepts are potentially less accessible and less automatically activated than they would be if encoded by grammatical means (following Lucy, 1992, 2011).

Estonian (a Finno-Ugric language) differs from English and Dutch in that information on the end state of the action and the involved object is expressed grammatically in the opposition between accusative[2] (1) and partitive (2) case-marking of the object noun:

(1) Naine kirjutas raamatu
woman.NOM wrote.PAST.3.SG book.**ACC**
'A woman finished writing a book.'

(2) Naine kirjutas raamatut
woman.NOM wrote.PAST.3.SG book.**PART**
'A woman was writing a book for a while (but did not finish it).'

Accusative case in (1) expresses the perfectivity (i.e. aspectual viewpoint, event is finished) and resultativity (event produced a result) of the action and that the object is affected in its totality. Contrastingly, partitive case in (2) conveys the imperfectivity and irresultativity in the action with respect to the object, and that the object is affected only partially by the action (Metslang, 2013; Tamm, 2004; Rätsep, 1978; Tauli, 1968; see Kiparsky, 1998 for Finnish). It is important to note that fulfilling only the first aspectual criterion, i.e., finishing the action is not enough to warrant accusative case marking: it is crucial for the object to be specific and quantitavely bound, and for the result to be a "tangible" entity. Thus, the object noun in example (3) would not be marked with accusative case, even if the action is completed:

2 It has not yet been agreed upon whether it is most appropriate to refer to the case with the label *accusative*, *total* or *genitive*. Accusative is used here following Lees (2004), who views it as a "blanket term for the non-partitive case" (p. 1).

(3) Mees segas suppi
man.NOM stirred.PAST.3.SG soup.**PART**
'A man stirred soup'

For an activity such as *writing a book* in (1-2) that does have the potential of achieving a result, the two case endings convey opposite information. Therefore, if information about the end phase of the activity is available to the speaker, it is pragmatically infelicitous to apply accusative case to an unfinished or irresultative event and partitive case to a finished or resultative event. That is, when the speaker knows that the event has reached its end state (action finished) and has produced a tangible result object, i.e., a finished book, the use of accusative case on the object noun would be required. Importantly, as case-marking is a grammatical feature, information about the end state for this type of events is a part of every present or past tense sentence in Estonian (as long as it contains an object); there is no "no-case"-variant of the object available. This means that when an activity has the potential of culminating in a result, conveying information about the end state of the action via the appropriate case marker is in fact *obligatory* in Estonian – and crucial for successful communication.

4 AIMS OF THE PRESENT STUDY

The current study employed the contrast between the encoding of resultativity in Estonian (obligatory in given contexts) and Dutch (optional marking only) as a window for investigating the nature of the relationship between language and perception. We used causative events (agent performing action on object, e.g., girl peeling a banana), for which Estonian participants have to convey information about its resultativity (whether or not the action finished and produced a tangible result object) via grammatical case marking on the object. Grammatical case-marking of noun phrases is not a feature of the Dutch language.

Speakers of the two languages viewed short (3-second-long) video-clips of caused action while their eye-gaze was being recorded, after which they performed a surprise memory recall task that tested their memory of the end state of the action depicted in each video (action finished or unfinished). The participants were randomly assigned to two encoding conditions. In one condition, participants were required to describe each video in one sentence after it had finished playing. In the other, nonverbal condition, they had to inspect the videos silently while performing a distracter task that instructed them to pay attention to the continuous background sound of the videos, and remember the content of those videos in which an additional sound cue was played (following Flecken et al., 2014b). In both experimental sessions, two additional neuropsychological tasks (Digit Span and Corsi Blocks tests) assessed general visuo-spatial and verbal memory abilities of the participants in order to counter potential concerns for overall memory differences between the two groups.

We hypothesize that grammatically encoding the absence and presence of a result of a causative event (e.g., girl peels a banana) in given contexts in Estonian should be reflected in how Estonian speakers distribute visual attention to particular event components while viewing such events, as indicated by their eye-gaze fixations in two spatially distinguishable pre-defined areas of interest, *Agent* (girl) and *Action+Object* (banana) (described in detail in section 6.3). More specifically, Estonian speakers are expected to show a heightened degree of attention to the region of the Action+Object towards the *end* of the video clips (i.e., final frames of the video), where the potential result of the action is revealed. The properties of this subcomponent of the event are critical for the selection of the appropriate case-marker in the Estonian language. Dutch speakers are predicted not to show an attention bias towards the Action+Object region during this time window, as information regarding the resultant state is not relevant to (and is typically not encoded in) their descriptions of the events.

Second, the frequent linguistic marking of the result of an action was hypothesized to affect Estonian speakers' performance during a subsequent memory recall task: we predicted superior performance for the Estonian group at recalling how each video ended, i.e., whether the action had terminated and the activity had produced a result or not.

By registering and analysing eye movement and memory recall under two different types of task conditions during the encoding phase, we address an important issue in the debate on the relation between language and perception – namely, whether potential language-specific viewing and memory patterns surface while thinking for speaking (verbal condition only), or whether such language effects operate by default and play a role in nonverbal scene encoding as well. Specifically, we investigate whether Estonian speakers, compared to Dutch speakers, display heightened attention to the action and object and enhanced memory of the events' resultant states in both contexts, or whether their scene encoding and memory patterns diverge under different encoding instructions. If Estonian speakers show a "resultativity bias" regardless of experimental condition, it would constitute evidence of a strong effect of language on event perception and recall. If Estonian speakers show this bias when verbal encoding is needed, but not in a nonverbal task, it would indicate that language mainly affects event perception when the task requires thinking for speaking. The latter result would be reflected in an interaction of language and encoding condition for our dependent measures, i.e., allocation of gaze to the Action+Object region over the agent region, during the final phases of the video clips, and memory response accuracy and reaction time in relation to the event's end state.

Following up on the exploration of a potential interaction of language and condition, the present study set out to obtain a fine-grained understanding of the extent of language effects within each condition for each of the two languages by including two subtypes of causative events: resultative actions (Type A) and non-resultative actions (Type B). Type A included actions in which the object is totally affected and that hold the potential of culminating in a tangible result within the duration of the short videos, such as *cut a circle, peel a banana, fold a paper plane* etc. The descriptions of this type of actions were expected to yield a consistent case-marking alternation from Estonian speakers: unfinished actions would trigger partitive case (e.g., "lennuk-it", 'plane.PART'), and finished actions (having

produced a tangible result) accusative case (e.g., "lennuk-i", 'plane.ACC'). Type B included actions in which the object is partially affected and does not undergo dramatic change, and no clear result it present at the end of the action, e.g., *mix cards, grate cheese, read a book* etc. Here, the use of partitive case is appropriate not only for unfinished actions, but for finished actions as well (e.g., "raamatu-t", 'book.PART' in both cases). It was an open question whether a potential resultativity bias for the Estonian group applies only to Type A items, where the case-marking paradigm is always evoked, or whether Type B items are processed in a similar way. In other words, by manipulating event type, we can obtain a more detailed picture of how the encoding of resultativity affects cognition: whether the potential influence of language is restricted to only those cases and contexts in which the actual grammatical distinction is accessed online, or whether an enhanced focus on the end state of events become a default and automatized way of parsing causative events in Estonian speakers. The latter would indicate a more global effect of language on causative event cognition.

In sum, the study aims to contribute to our understanding of the influence of language on event perception and cognition in several aspects. First, the choice of the particular language pair allowed uncovering potential modulations of event perception that are induced by the presence or absence of grammatical features in a language. While previous studies have mainly correlated differences in verbal or nonverbal processing with lexical structures (e.g., manner vs path verbs), it is likely that grammaticalized distinctions bear a stronger effect on cognition (Lucy, 1992). The alternation of case-marking in Estonian is obligatory in specific contexts, which could lead to a more automatized activation of the connected cognitive processes. Second, using a combination of offline and online measures (memory performance and registration of eye movements) within the same group of speakers enables us to obtain a more detailed picture of potentially subtle language-based differences than compared to employing one research method alone. Third, by using an identical set of stimuli and experimental set-up to study verbal and nonverbal encoding in two languages, we are able to draw more definite conclusions about the extent of the influence of the particular grammar-based linguistic feature. By additionally including a manipulation of event type and systematically controlling for the overt linguistic encoding of the relevant feature, we can provide a fine-grained analysis of the scope of potential language-based processing patterns. Finally, by including Estonian, a typologically distinct and much less-studied language, the study broadened the scope of cross-linguistic psycholinguistic studies, which have so far only included a regrettably small sample of languages and a limited range of linguistic phenomena as test cases for language-on-cognition effects.

5 METHODS

5.1 Verbal encoding condition

5.1.1 *Participants*

The Estonian group included 33 participants, recruited from the the University of Tartu community, Estonia. Out of the initial group, ten participants had to be excluded due to technical error or failure to fulfil task requirements[3]. The final sample consisted of 23 native speakers of Estonian with a mean age of 22.57 (SD 3.47, n=17 female and n=6 male). The Dutch group included 26 native speakers of Dutch, recruited from Radboud University Nijmegen, the Netherlands. Here, data from two participants were excluded due to technical error[4], leaving the final group of 24 participants with a mean age of 22.46 (SD 3.32, n=17 female and n=6 male). The participants were right-handed and had normal or corrected-to-normal vision, none reported neurological or psychological disorders. All participants gave written consent to take part in the experiment and received payment for their participation.

5.1.2 *Materials*

The items of the encoding task were video-clips of 3000 ms in duration, recorded for the purpose of this study at the Max Planck Institute for Psycholinguistics, which depicted four different actors (3 female, 1 male) performing common every-day actions on various objects. The actions were selected and acted out with the objective to spatially separate the main elements of the event (Agent and Action) from each other to the maximal extent, which ensured that, given the current size of the videos on the screen and thus the visual angle for our participants, the viewers had to place overt fixations in each region to retrieve information on it (this excluded e.g., *eat an apple*, where one can extract information parafoveally on both elements at the same time). Spatial separation of the elements permitted the distinction of two Areas of Interest in the analysis of eye-gaze data, one corresponding to Agent, the other to Action. The actions were filmed against a white background with no distracting items.

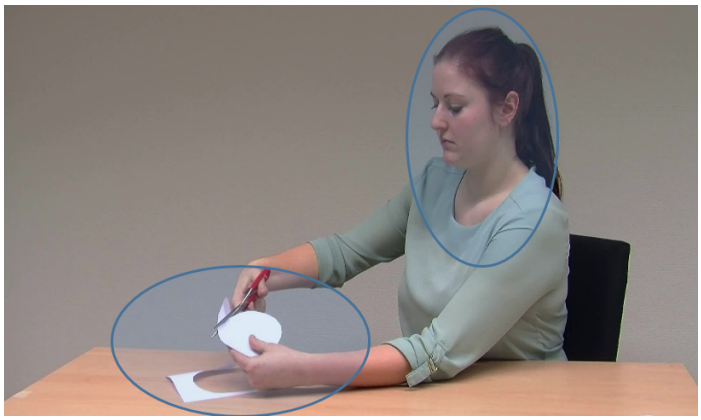
The events belonged to three different categories: resultative events (Type A, n = 18), non-resultative events (Type B, n = 18) and filler events (n = 18). Type A included transitive actions on a singular, specific object, which produce a tangible result, such as *draw a flower* or *cut a circle*. Descriptions of this type of actions are most likely to undergo the case alternation in Estonian (see section 3). Type B included actions that do not produce a clear result, such as *stir soup* or *beat cream*, and are thus not subject to case alternation in the current context. The fillers were intransitive events (n = 4), ditransitive events (n = 6) or transitive events (n = 8) in which it was impossible to obtain spatial distinction

3 Data from five participants were discarded due to a low sampling rate (60Hz) or tracking ratio (<70%). Five more participants were excluded based on their descriptions of the events: instead of a finite sentence, they used nominalizations of the action such as "*the peeling of the banana*", where no direct object and thus no information about the end-state of the action was present.

4 Data were recorded at a low sampling rate (60Hz).

between the agent and the action. An example of a stimulus item from Type A is depicted in Figure 1, and a full list of test and filler items is provided in Appendix A.

Figure 1: Example stimuli: *cut a circle*
Eye-tracking Areas of Interest are marked by ellipses (Agent: area of head and shoulders of the actor; Action: area of action and the object)



Each Type A and Type B action had a finished and unfinished counterpart, all edited to 3000 ms in length. The unfinished versions always depicted the mid-state of the action. The finished versions of Type A items additionally showed the completion of the action with the achievement of a result, and the finished versions of Type B items showed the completion of the action without reaching a result. Four stimulus lists with 54 video clips in each were constructed, such that two lists included the finished version of the action and the other two included the unfinished version of the same action. This meant that the number of finished and unfinished actions was the same in each list - 18 finished and 18 unfinished (plus 18 fillers - 9 finished and 9 unfinished), and that each participant saw each action once, either in the finished or unfinished version. The four actors and the position of the object (to the left or right of the actor) were pseudo-randomized within the lists. The end of an action was indicated by either a clear, visible culmination (e.g. paper was cut in half) or the actor putting down the object and removing hands, if the first option was not available (e.g. stirring soup in a bowl). The duration of the end phase was kept as constant as possible across the different events.

The items of the memory test were screenshots of the last frames of both the unfinished and finished versions of each video (see discussion of method in section 5.1.3 and examples of the setup in Figure 2).

5.1.3 *Procedure*

The participants were first asked to sign a consent form and fill in a language background questionnaire, where information about their languages and education level was collected. They were then seated to approximately 60 cm from the remote, contact-free SMI RED250m eye-tracker (SensoMotoric Instruments), attached to the lower part of the laptop screen. The display resolution of the laptop was set to 1920x1080, and the eye-tracker recorded the movements of both eyes at 250Hz (i.e., every 4 ms). The

participants then performed the four components of the experiment in the following order: (1) Encoding task; (2) Corsi Blocks task; (3) Memory task; (4) Digit Span task.

The eye-tracker was set to record only during the first, encoding task. The software package Presentation NBS (Neurobehavioral Systems, Albany, CA) was used to control the eye-tracker, present the stimuli, register button presses and record speech (for later transcription); during each trial. The software sent time-stamps to the eye-tracking system for the events of interest (video onset and offset, speech onset). Due to the nature of the one-computer set-up, each component of the experiment had to be set up manually by the experimenter, and the initial calibration and mid-experiment re-calibrations were performed by the participant in a semi-automatic fashion, i.e., an occasional button-press was required for the (re-)calibration and validation to proceed. Each experimental session lasted approximately 35 minutes.

PART 1 – VERBAL ENCODING TASK The participants first read instructions of the task on the computer screen, translated into their native language by a native speaker. The instructions required the participant to watch the video until the end (signalled by a beep) and describe the event in each video clip in one sentence, answering the question "What happened in the video?" ("*Wat gebeurde in de video?*", "*Mis videos juhtus?*"), and press the spacebar to proceed to the next video. The responses were recorded by an external microphone. Note that each video disappeared from the screen after playing, meaning that the participants were looking at an empty white screen rather than a frozen final frame of the video while speaking (cf. Papafragou et al., 2008). The participants then carried out two practise trials during which they could ask clarification questions or receive further instructions from the experimenter. In the responses to the participants' questions, the critical question regarding the encoding task was always formulated in the same way as in written instructions. Moreover, the participants did not receive feedback regarding their chosen sentence constructions: they were not told to use transitive sentences or employ the past or present tense, but encouraged to describe the videos in the way that seemed most suitable to them.

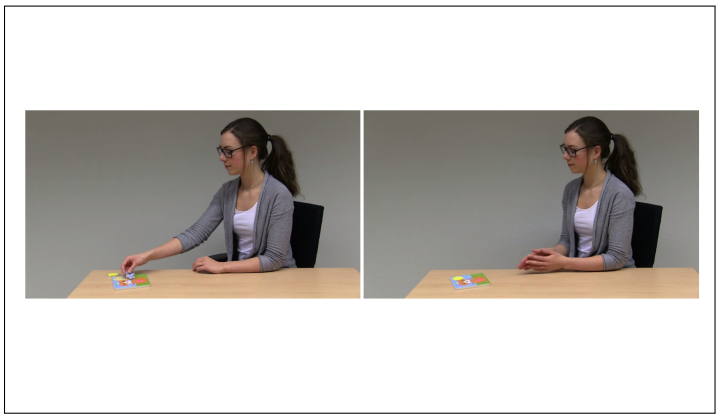
The two practise trials were followed by a semi-automatic 5-point calibration procedure, which was repeated after every 12 trials (i.e., about 2 minutes) to ensure maximum tracking quality throughout the encoding phase. Together with the videos, verbalizations and (re-)calibrations, the task lasted between 9 minutes and 11 minutes in total, depending on each individual's speaking rate and length of utterances.

PART 2 – CORSI BLOCKS TAPPING TASK The encoding task was followed by a mouse-based version of the Corsi Blocks Tapping task (Cognitive Experiments III v3, www.neurobs.com), a common method for measuring visuo-spatial memory (Kessels et al., 2000). Participants were presented with 9 blue rectangles on a grey background, which turned red one by one. The task was to memorize the order in which the rectangles switched colour, and recreate the order after each trial by clicking on the rectangles with the computer mouse. The testing began at a trial length of 3 rectangles and increased in a 1:2 staircase method (following Woods et al., 2011), where a single

correct response increased the length of the subsequent list by one rectangle and two incorrect responses reduced the length by one rectangle. The task ended after the participant had completed 14 trials. The duration of the task varied slightly due to differences in the list length that each participant reached, but the test was timed by the experimenter to confirm that it always lasted between 3,5-4 minutes. The participant did not receive feedback on their accuracy after the task.

PART 3 – MEMORY TASK During the next component, we used a two alternative forced-choice (2AFC) task to test the participants' memory of the end state of the previously encoded videos. In this task, the participants saw two screenshots side by side: one of the screenshots showed the actual ending of the video (e.g., woman completed drawing a flower) while the other screenshot was made of its counterpart with respect to completion (e.g., woman was in the middle of drawing a flower). The actor and the position of the object (left or right of the actor) was the same as during the encoding phase; the order of presentation was similar but not identical to the one before. Participants were required to press the appropriate button on the keyboard (left or right) as fast as possible. An example screenshot from the memory task is shown in Figure 2.

Figure 2: Screenshot of the memory task: item *do a puzzle* (Type A)



PART 4 – DIGIT SPAN TASK The final component of the experiment was the forward version of the Digit Span task (Cognitive Experiments III v3, www.neurobs.com). In each trial, a series of digits was presented one by one in the centre of a white screen. Participants had to memorize the digits, and type them in by using the keyboard in the sequence after each trial. As for the Corsi Blocks task, the list length of the trials started from 3 and increased in a 1:2 staircase method until 14 trials had been presented.

5.2 Nonverbal encoding

5.2.1 *Participants*

The Estonian group included a different sample of 25 native speakers of Estonian. Data of two participants were excluded from the analysis[5], so the final group consisted of 23 participants with a mean age of 24.39 (SD 3.96, n=16 female and n=7 male), recruited from the community of the University of Tartu, Estonia. The Dutch group included 24 native speakers of Dutch, recruited from Radboud University Nijmegen, the Netherlands. After excluding 2 participants[6], the group of 22 speakers had a mean age of 21.68 (SD 2.40, n=17 female and n=5 male). The participants were right-handed and had normal or corrected-to-normal vision, none reported neurological or psychological disorders. All participants gave written consent to take part in the experiment and received payment for participation.

5.2.2 *Materials*

Both the items and the presentation lists of the encoding task and the memory task were identical to the ones in the verbal condition, described in section 5.1.2.

5.2.3 *Procedure*

Except for the encoding task, the procedure of the non-verbal condition was exactly the same as in the verbal condition (see section 5.1.3). Instead of describing the videos after encoding, the participants received a sound-cue task (closely following but slightly adapted from Flecken et al., 2014b).

The participants were asked to watch the videos in silence while listening to the sound of ocean waves in the background. They were told that a short beep would be played during randomly selected videos, and that their task was to remember during which of the videos the beep was played. The videos were presented automatically one after another with 3500 ms breaks in between (2000 ms white screen; 1500 ms fixation cross). After having seen a block of 6 videos, the participants saw a screenshot of one of the videos, which was accompanied by the question *"Did you hear a beep during this video?"* in the participant's native language. Having pressed a button that corresponded to YES (green) or NO (red), the experiment continued either with a new set of 6 videos, or a re-calibration, if 2 blocks of 6 videos had been played.

The beeps were played at the middle of each video. Crucially, they only occurred during filler trials and not during Type A or Type B items, which avoided potential heightened attention to the critical event scenes and unwanted familiarization with their end states

5 Data from one Estonian participant were discarded due to a wrong sampling rate (120Hz). Another participant was excluded because the language background questionnaire revealed they were an early bilingual speaking both Russian and Estonian since birth.

6 Data from 2 Dutch speakers were excluded due to a technical problem during the memory task.

by having seen a screenshot of it. Furthermore, the number of beeps within a block alternated between 1-3, which made the task more difficult.

This task was designed to keep participants engaged and focus their attention on the video clips without biasing them towards any of the event elements. Previous nonverbal paradigms have involved giving participants the instructions to inspect the scenes carefully for an upcoming memory task, which might have biased their attention to all aspects of the scenes, including details which would typically not be focused (e.g., Papafragou et al., 2008). Moreover, with such instructions the use of inner speech and verbalization strategies cannot be excluded; a reliance on language to memorize the scenes seems likely. Following Flecken et al. (2014b), we claim that the present task reduces the use of inner verbalization strategies, at least sentence-level verbalizations.

6 DATA CODING, PREPROCESSING AND ANALYSIS

6.1 Control data

The method of obtaining a participant score for Corsi Blocks and Digit Span control tasks was adopted from Woods et al., 2011, who advocate the use of a novel mean span (MS) metric as the most reliable and precise measure for quantifying the results of neuropsychological tasks [7]. The MS baseline was set at 2.5 (0.5 digits less than the initial list length) and the score was calculated by adding the baseline to the rate of accuracy at each list length (see Woods et al., 2011, p. 5 for an example).

An accuracy score for the secondary sound-cue task in the nonverbal condition was calculated by dividing the correct number of responses by the total number of sound-cue prompts (n=9). The results provided an indication of whether the participants understood and paid attention to the task.

6.2 Language production data

The transcribed data were coded for object case-marking (in Estonian), aspectual switches (e.g., "*was cleaning*" or "*has cleaned the mirror*"), tense of utterances (past vs. present) and agent specificity (e.g., "*a man with glasses*"). Coding was carried out by two independent coders. Discrepancies between the coders (regarding aspect, tense, case marking and agent specificity) only existed in a very low number of trials (about 5% of all trials) and were resolved after discussion.

7 The authors only discussed the application of the MS for the Digit Span task, but as the Corsi Blocks task was administered in an identical manner in this experiment, it is reasonable to assume that the method can be used for this task as well.

6.3 Eye-gaze data

Eye movement patterns during the encoding phase of each condition were analysed for the critical items (Type A: n=18, Type B: n=18), testing potential main and interaction effects of Condition (Verbal/Nonverbal encoding) and Language (Estonian/Dutch). In follow-up analyses, the factor Event type (Type A/Type B) was included as well.

Two identically sized and spatially distant elliptical areas of interest (AoI) were defined for each stimulus after all data had been collected: one AoI included the head and shoulders of the actor (Agent), and the other AoI included the region of the action (Action), encompassing the agent's hands as well as the object and the instrument fully (see Figure 1)[8]. The size of the two AoIs was kept constant across all videos. Fixations in these two AoIs were computed for the entire time that the videos were playing and until participants started speaking, with SMI BeGazeTM software (SensoMotoric Instruments, SMI)[9]. During the recording, Presentation NBS software sent timestamps to the eye-tracker, providing the time of stimulus onset, stimulus offset and speech onset for each trial for later preprocessing.

We plotted fixations during the time that the videos were playing and shortly thereafter (5000 ms in total). Fixation data were preprocessed in R (version 3.2.3), using a script which detected for each participant whether a fixation fell into a particular AoI in each of 100 successive 50 ms bins. For plotting, we aggregated fixations across participants for each AoI and time bin; data was presented as the proportion of fixations at a particular AoI during a given time bin (i.e., number of fixations out of all registered fixations during the time bin) (following Flecken et al., 2014a). In all cases, finished and unfinished action trials were collapsed.

Our analyses of eye movements focused on a subpart of the overall time course, namely the final phase of the unfolding of the event (see section 7.3 for details), during which it became clear whether the event finishes and the object reaches a resultant state or not. We computed the total number of fixations in the Action and Agent AoIs for each participant in this time window, and subtracted all Agent fixations from the total number of Action fixations in order to obtain a so-called "action-bias" for the language groups in each condition (cf. Papafragou et al., 2008). We then computed the logit-transformed odds ratio of the action-bias (log(action-bias / total number of AOI fixations – action-bias)), following the quasi-logistic regression analysis proposed by Barr (2008). Logistic mixed effect regression models were used to predict the probability of Action-over-Agent fixations on the basis of the predictors Language, Condition and Event type (the latter factor is only included in follow-up analyses for each Language separately).

First, an overall Language (Dutch vs. Estonian) by Condition (Verbal vs. Nonverbal) analysis was run on the dependent variable. Follow-up tests focused separately on each

8 The Agent region was matched with the face of the actor, as facial features are most relevant for the identification of the human agent and have been shown to be most interesting to viewers. The Action region admittedly included the visual areas of both the action and the object, but as these areas necessarily overlap, it was impossible to disentangle attention to the two elements from each other in the analysis.

9 Although both the number and duration of fixations at stimulus components have been shown to relate to language processing and attention allocation (Griffin, 2004), only the first measure was analysed in this paper due to space limitations.

Condition (Verbal vs. Nonverbal) and each Language (Estonian vs. Dutch), and in the latter case include Event type: Type A (resultative) vs. Type B (non-resultative).

6.4 Memory data

The memory analyses focused on accuracy of responses (correct/incorrect choice of variant of the event's endstate) as well as reaction times of the responses. We plotted the proportion of accurate responses for all items for each Condition and Language. Data was analysed with logistic mixed effect regression models, taking trial-by-trial binomial responses (accurate: yes "1", no "0") as the dependent variable. The analysis was exactly the same as for eye-tracking data: we first ran an overall analysis of Language (Dutch vs. Estonian) by Condition (Verbal vs. Nonverbal), and if motivated, followed up on each Condition (Verbal vs. Nonverbal) and each Language (Estonian vs. Dutch), including Event type: Type A (resultative) vs. Type B (non-resultative).

7 RESULTS

7.1 Control data

7.1.1 *Corsi Blocks*

The mean scores for the Corsi Blocks task are presented in Figure 3, and were analysed with a two-way analysis of variance (ANOVAs) with two levels of CONDITION (verbal, nonverbal) and two levels of LANGUAGE (Dutch, Estonian). The analysis showed no main effect of Language ($F(1,86) = .010$, $p = .755$, n.s.), no main effect of Condition ($F(1,86) = .506$, $p = .479$, n.s.) and no interaction between Language and Condition ($F(1,86) = 2.037$, $p = .157$, n.s.). These results indicated that the mean scores from the Corsi Blocks task of any of the four participant groups did not significantly differ from each other.

7.1.2 *Digit Span*

The mean scores for the Digit Span task are presented in Figure 4. As for Corsi Blocks, the means were analysed with ANOVAs with two levels of CONDITION (verbal, nonverbal) and two levels of LANGUAGE (Dutch, Estonian). The analysis showed no main effect of Language ($F(1,86) = .345$, $p = .559$, n.s.), no main effect of Condition ($F(1,86) = .012$, $p = .912$, n.s.) and no interaction between Language and Condition ($F(1,86) = 1.140$, $p = .239$, n.s.). It appeared that the mean scores from the Digit Span task of any of the four participant groups did not significantly differ from each other either.

Figure 3: Corsi blocks results (error bars indicate +/- SE)

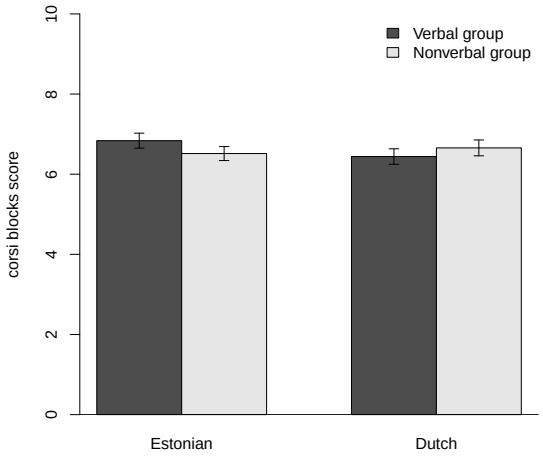
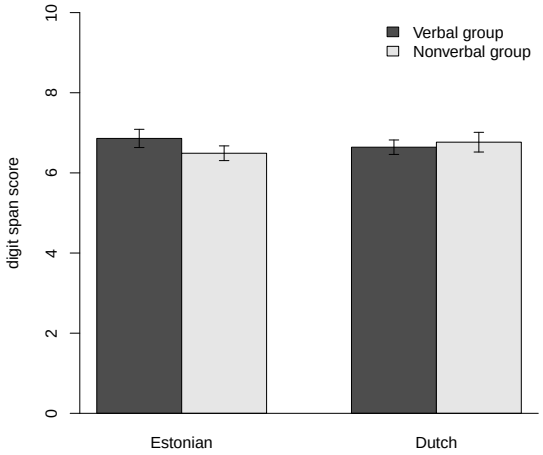


Figure 4: Digit Span results (error bars indicate +/- SE)



7.1.3 *Nonverbal encoding: Sound-cue detection task*

Regardless of the language background, all participants demonstrated high accuracy scores at the sound-cue detection task (average scores about 95% for both language groups). This can be taken as evidence that both groups understood the task equally well and paid attention to the videos throughout the encoding phase of the experiment.

7.2 Verbal encoding: Language production data

Table 1 shows the absolute and relative frequencies of case-marking in the Estonian data. The statistics justify our division of stimulus items into two types, and demonstrate that the group was to a large extent sensitive to the manipulation of the completedness and resultativity of the events: the Estonian group marked finished Type A items with accusative case and the other 3 groups of items (Type A unfinished; Type B finished and unfinished) predominantly with partitive case.

Table 1: Case-marking in Estonian data (total items: n=828)

	Type A		
	unfinished	finished	overall (% of total items in Type A)
PART	176	79	61.6%
ACC	30	124	37.2%
n/a [a]	1	4	1.2%
	Type B		
	unfinished	finished	overall (% of total items in Type B)
PART	188	165	85.3%
ACC	6	20	6.3%
n/a	14	21	8.4%

a Instances of intransitive sentences (without any object), indirect objects (marked with a different case) or technical problems with the recording.

However, Table 1 also shows that Type A items were occasionally marked with accusative case, even when the event was shown as unfinished. Likewise, there were instances where finished Type A actions were described using partitive case, and finished Type B items with accusative case. Potential reasons for these findings will be taken up in the discussion section. Importantly, the overall patterns are in line with our expectations, namely, case alternation for Type A items (62% partitive, 37% accusative), and mostly partitive case (85%) for Type B items.

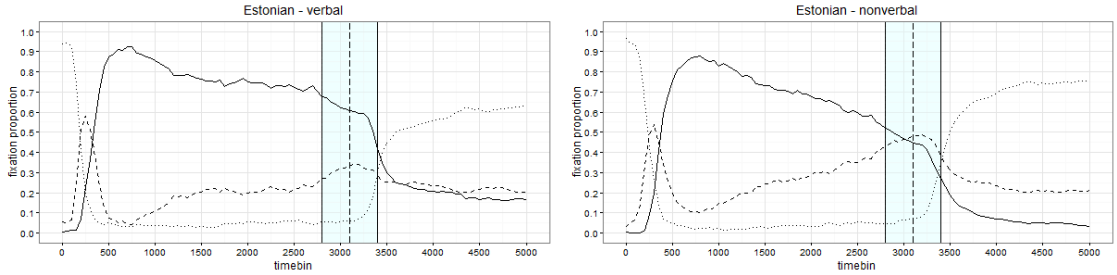
The use of different constructions in Dutch turned out to be negligible (24/864 'aan het' constructions (2.8%); no present perfect constructions), so no further descriptive statistics on this language will be provided.

7.3 Eye-gaze data

Figure 5 and Figure 6 present the proportion of fixations out of all registered fixations in the Agent and Action AoIs, and Outside of both (background) (in both verbal and nonverbal encoding conditions) by Estonian and Dutch groups respectively. The proportions are plotted as a sequence of 50 ms timebins from the onset of the video up to 5000 ms. Due to a 0-100 ms time delay in video playback by the presentation program, the videos finished playing at ~3100 ms (dashed line on the graphs).

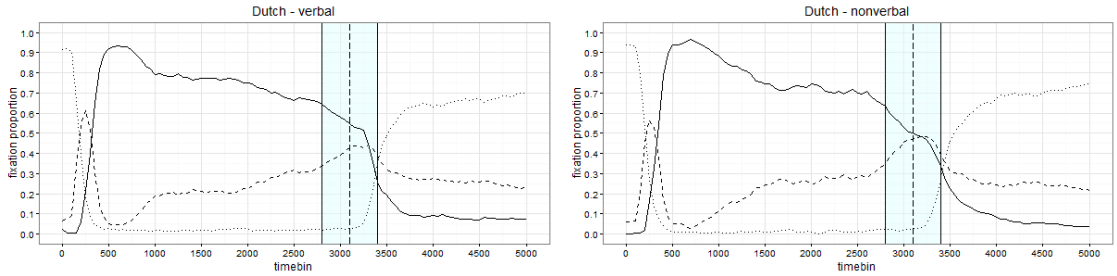
The figures clearly demonstrate that regardless of condition and language, participants first turned their attention to the Agent region, before moving their eyegaze to the Action region and maintaining their gaze in this region for most of the video duration. For one, this matches the order of mentioning the event components in each language in the subsequent event descriptions. Moreover, the continuous dominant allocation of attention to the action region reflects that the action region displays dynamic content and thus continues to attract gaze.

Figure 5: Estonian eye-gaze data (event types combined)



Line types: solid = Action, dashed = Agent, dotted = Outside

Figure 6: Dutch eye-gaze data (event types combined)



Line types: solid = Action, dashed = Agent, dotted = Outside

We were interested in eye-gaze patterns during the specific part of the video where information about the resultant state of the event was conveyed. We therefore computed the log odds of Action-over-Agent fixations within a selected time window: 2800-3400 ms (blue-shaded area in Figure 5 and Figure 6). The cut-off mark for the beginning of the analysis (2800 ms) was the point in time when the action reached culmination in the finished version of each clip. The video ended at 3100 ms, i.e., 300 ms after the resultant state of the event became visible, and visual inspection of the plots showed a maximum of 300 ms delay in removing eyes from that region; this is also the point in time in the plots where the lines for each of the AoIs come together. These reasons motivated the selection of 3400 ms as the cut-off mark for the end of the analysis. This data-driven approach differs from comparing eye-gaze patterns at every second (e.g., Papafragou et al., 2008; Bunger et al., 2016, but arguably provides a more sensitive picture of the processes underlying the subpart of event that is of most interest to the current investigation.

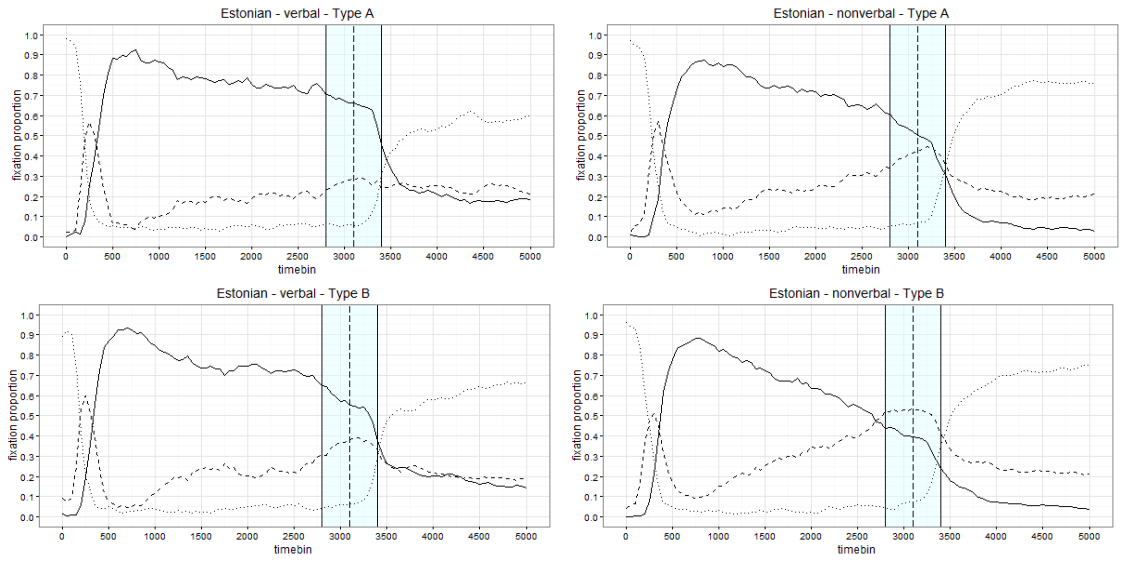
We ran linear mixed effect regression models (separate interaction and main effects models) on the empirical logits of action-over-agent ("action-bias") fixations (see description in section 6.3), taking LANGUAGE (Estonian vs. Dutch) and CONDITION (verbal vs. nonverbal) as fixed factors, and a random effects structure with random intercepts for PARTICIPANTS and ITEMS, as well as a by-item random slope for CONDITION. Dutch language and verbal condition were set as baseline categories. The model showed no main effect of Language (β = .063, SE = .078, t = .807, p = .422, n.s.), but a main effect of Condition (β = –.238, SE = .078, t = –3.031, p < .001), and a significant Language by Condition interaction (β = –.366, SE = .152, t = –2.396, p < .05). Overall, the nonverbal condition elicited a lower action-bias compared to the verbal condition. The significant

interaction was followed up by Condition analyses, targeting effects of Language within each condition. Furthermore, we split up the data by Language to compare conditions and to explore potential main and interaction effects of the factor EVENT TYPE (here, Type B items were assigned as the baseline category).

DATA SPLIT BY CONDITION The models included LANGUAGE as a fixed factor, and PARTICIPANT and ITEM as random intercepts. Analysis of the results from the verbal condition showed a significant Language effect (β = .242, SE = .117, t = 2.080, p < .05), showing a larger action-bias in the Estonian group compared to the Dutch group. When running the same model on data from the nonverbal condition, the results showed no effect of Language (β = –.124, SE = .098, t = –1.264, p =.213, n.s.).

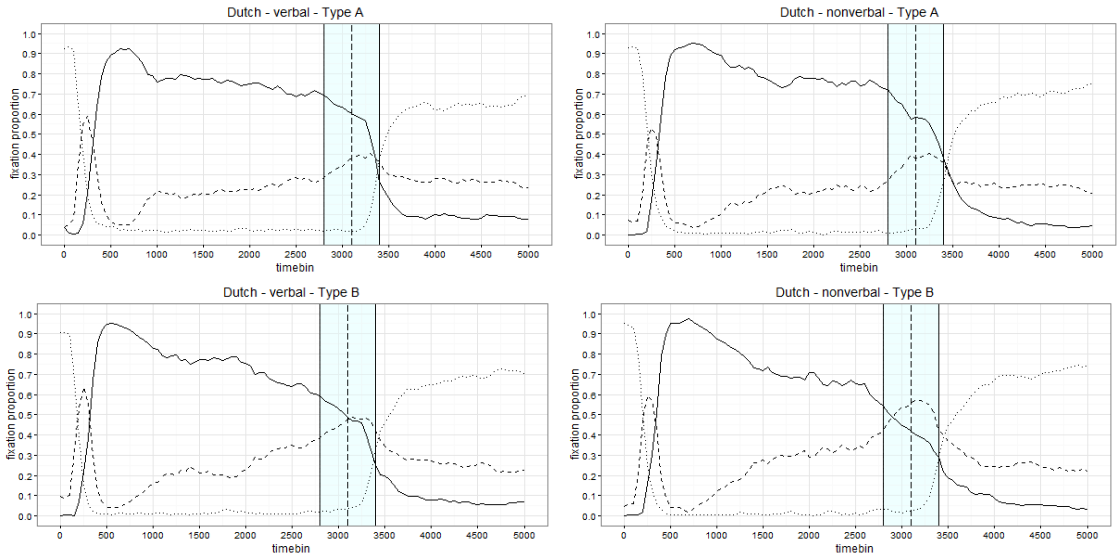
DATA SPLIT BY LANGUAGE Figure 7 and Figure 8 show the eye-gaze data split up by two conditions and two event types for Estonian and Dutch speakers respectively. The models of analysis included CONDITION and TYPE as fixed factors, random intercepts for PARTICIPANT and by-participants random slopes for EVENT TYPE. Results for the Estonian group showed main effects of Condition (β = –.410, SE = .108, t = –3.789, p < .001) and Event type (β = .291, SE = .062, t = 4.665, p < .001), but no interaction between Condition and Event type (β = .086, SE = .125, t = –.687, p = .496), which underlines a lower action-bias in the nonverbal compared to the verbal condition in Estonian participants, as well as an overall higher action-bias for Type A compared to Type B items. Models on the Dutch data showed no Condition effect (β = .0236, SE = .105, t = .226, p = .823, n.s.), but a main effect of Event type (β = .365, SE = .049, t = 7.519, p < .001). There was also a marginally significant interaction between Condition and Event type in the Dutch data (β = .187, SE = .094, t = 1.988, p = .053). This analysis shows that the Dutch participants also displayed a larger Action bias for Type A items than for Type B items, and mainly so in the nonverbal condition.

Figure 7: Estonian eye-gaze data, split into Type A and Type B items



Line types: solid = Action, dashed = Agent, dotted = Outside

Figure 8: Dutch eye-gaze data, split into Type A and Type B items



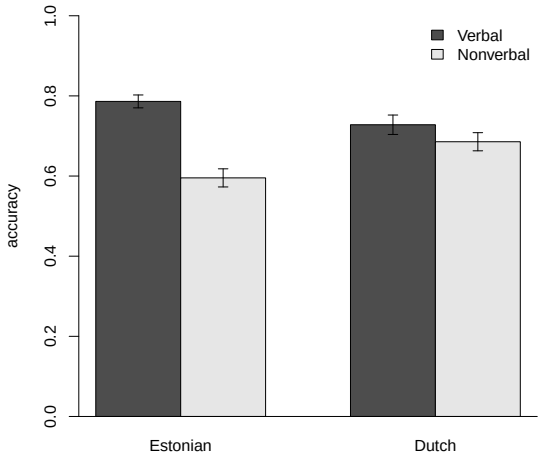
Line types: solid = Action, dashed = Agent, dotted = Outside

7.4 Memory data

7.4.1 *Accuracy results*

Figure 9 depicts accuracy scores for both language groups in verbal and nonverbal conditions.

Figure 9: Accuracy scores - language by condition
(error bars indicate +/- SE)



We used logistic mixed effects regression models (separate interaction and main effect models) to predict accuracy at the task (correct = 1, incorrect = 0). The models included two fixed factors: LANGUAGE (Estonian vs. Dutch) and CONDITION (verbal vs. nonverbal), random intercepts for PARTICIPANTS and ITEMS, as well as a by-item random slope for CONDITION. Dutch language and Verbal condition were set as baseline categories. The model showed no main effect of Language (β = –.043, SE = .115, z = –.371, p = .711, n.s.),

but a significant main effect of Condition ($\beta = -.558$, SE = .115, $z = -4.850$, $p < .001$) and a significant interaction between Language and Condition ($\beta = -.690$, SE = .218, $z = -3.167$, $p < .01$). These findings show that the overall scores for accuracy were higher in the verbal condition than in the nonverbal condition, but that the difference between scores in the two conditions was bigger in Estonian than in Dutch.

The significant interaction between Language and Condition motivated splitting up the data by Condition (verbal and nonverbal) to target direct language comparisons. Furthermore, we split up the data by Language to contrast conditions and to explore potential main and interaction effects of the factor Event type (here, Type B items were assigned as the baseline category).

DATA SPLIT BY CONDITION The model to test for an effect of Language in each of the two conditions included LANGUAGE as a fixed factor, and PARTICIPANT and ITEM as random intercepts. Analysis of the results from the verbal experiment showed a marginally significant effect of Language ($\beta = .317$, SE = .166, $z = 1.908$, $p = .056$) such that the Estonian group was better than the Dutch group in the verbal encoding condition. When running the same model on data from the nonverbal condition, the results showed a significant effect of Language ($\beta = -.369$, SE = .144, $z = -2.568$, $p < .05$): the Dutch group appeared to perform better than the Estonian group in the nonverbal encoding condition.

DATA SPLIT BY LANGUAGE The models included CONDITION and EVENT TYPE as fixed factors, random intercepts for PARTICIPANT and by-item random slopes for CONDITION. As for the eye-gaze data above, we were further interested in potential differences within each of the language groups in how well they remembered the two types of events (Type A: resultative vs. Type B: non-resultative). Barplots of accuracy means are shown in Figure 10 for the Estonian group, and in Figure 11 for the Dutch group.

Figure 10: Accuracy - Estonian group by condition and type (error bars indicate +/- SE)

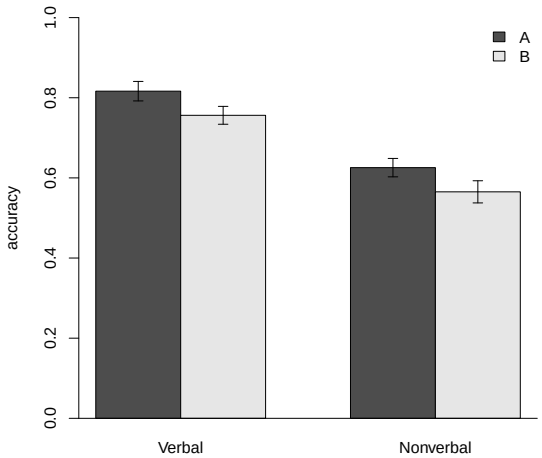
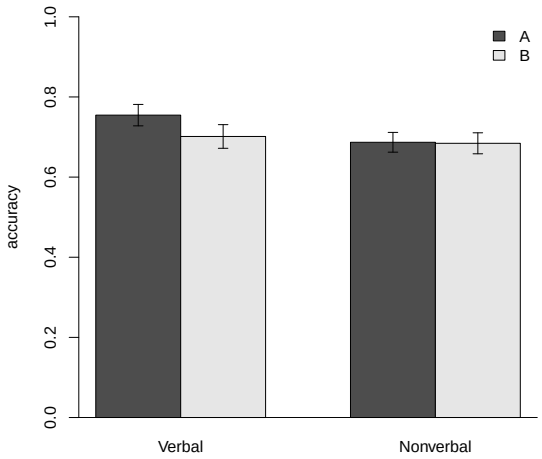


Figure 11: Accuracy - Dutch group by condition and type (error bars indicate +/- SE)



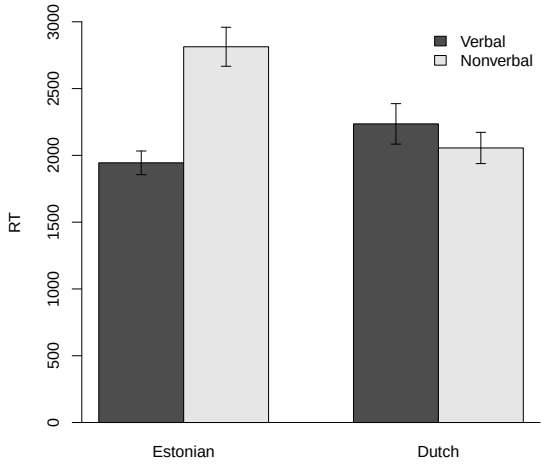
The analysis of Estonian data showed a significant effect of Condition (β = –.956, SE = .156, z = –6.119, p < .001), a significant effect of Event type (β = .318, SE =.155, z = 2.056, p < .05), but no interaction between Condition and Event type (β = –.009, SE = .253, z = –.355, p = .722, n.s.). These results indicate that the Estonian group was overall better at remembering the events' endstates in the verbal than the nonverbal condition. Moreover, the group had higher accuracy scores for Type A events than Type B events regardless of condition. When applying the same models to the Dutch data, we found no effect of Condition (β = –.230, SE = .167, z = –1.379, p = .168, n.s.), no effect of Event type (β = .103, SE = .147, z = .704, p = .482, n.s.), and no interaction between Condition and Event type (β = –.266, SE = .226, z = –1.175, p = .240, n.s.). Thus, diverging from the Estonian results, the Dutch group did not differ in memory accuracy after verbal or nonverbal encoding; they also did not appear to remember Type A items differently from Type B items in either of the experiments.

7.4.2 *Reaction time (RT) results*

The mean RT scores for the two language groups in verbal and nonverbal conditions are shown in Figure 12. In order to obtain normal distribution of the data, RT values were log transformed, after which outlier values above or below 2.5*SD were excluded from the analysis (following Baayen, 2008).

To analyse log-transformed RT data, we used linear mixed effects models (separate models testing interaction and main effects), which included LANGUAGE and CONDITION as fixed factors, random intercepts for PARTICIPANT, and by-item random slopes for CONDITION. The analysis showed no main effect of Language (β = –.008, SE = .060, t = –1.339, p = .184, n.s.), a significant main effect of Condition (β = –.137, SE = .056, t = –2.289, p < .05), and a significant interaction between Language and Condition (β = –.397, SE = .112, t = –3.527, p < .001). Corresponding to the pattern in accuracy results, we see that both groups were faster in the verbal task than for the nonverbal task, but that the difference was not similar in both language groups.

Figure 12: Reaction times - language by condition
(error bars indicate +/- SE)



DATA SPLIT BY CONDITION For the verbal condition, the model with LANGUAGE as a fixed factor, and random intercepts for PARTICIPANT and ITEM showed no effect of Language (β = .114, SE = .076, t = 1.514, p = 0.137, n.s.), indicating that the groups did not differ in reaction times in the verbal encoding condition. For the nonverbal condition, a significant effect of Language appeared (β = −.282, SE = .083, t = −3.38, p < .01), such that the Dutch group was faster than the Estonian group at responding in the nonverbal condition.

DATA SPLIT BY LANGUAGE The models (interaction and main effects) included CONDITION and EVENT TYPE as fixed factors, random intercepts for PARTICIPANT and ITEM, a by-item random slope for CONDITION and a by-participant random slope for EVENT TYPE. Mean RTs are shown in Figure 13 for the Estonian group and in Figure 14 for the Dutch group.

Figure 13: Reaction times - Estonian group by condition and type
(error bars indicate +/- SE)

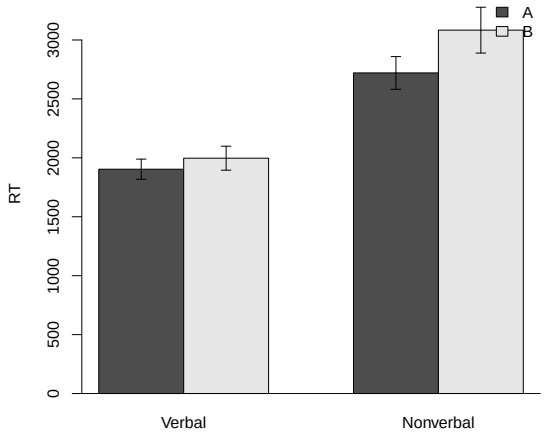
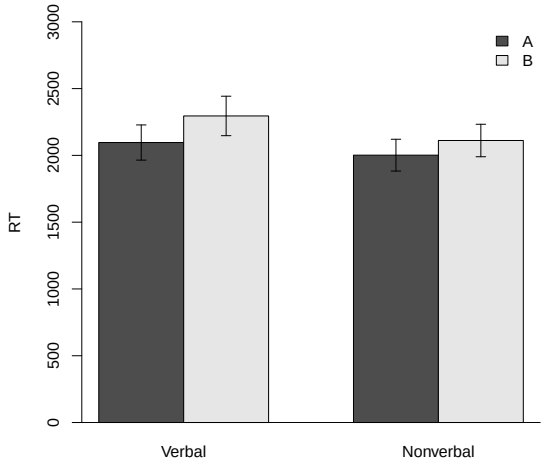


Figure 14: Reaction times - Dutch group by condition and type (error bars indicate +/- SE)



The analysis of Estonian data showed a significant main effect of Condition (β = –.300, SE = .068, t = –4.391, p < .001), a trend for a main effect of Event type (β = .071, SE = .038, t = 1.889, p = .067), and no interaction between Condition and Event type (β = .062, SE = .040, t = 1.541, p = .132, n.s.). The Estonian group appeared to be faster at in the verbal condition than the nonverbal condition, and there was a tendency in the data for faster responses for Type A items compared to Type B items. The analysis of Dutch data showed no main effect of Condition (β = .062, SE = .087, t = .715, p = .478, n.s.), a trend for a main effect of Event type (β = .0613, SE = .032, t = 1.900, p = .065), and no interaction between Condition and Event type (β = –.020, SE = .038, t = –.543, p = .590, n.s.). The analysis shows that the Dutch group responded equally fast in the two experimental conditions. Rather surprisingly however, the analysis further shows an advantage for the Dutch group at responding to Type A items in both conditions.

8 SUMMARY AND DISCUSSION

The present study set out to explore the effects of specific language background and task demands on how people encode and memorize events. The test-case for investigating the potential influence of language on event cognition was the differential system of case-marking in Estonian, where the grammatical case of an object of a transitive sentence depends on the properties of the action and the object: the object noun is marked with accusative case if the activity reaches completion and produces a tangible result, meaning that the object is affected in its totality (e.g., woman finished drawing a flower), and with partitive case if the activity is not completed or it terminated without a result, so only a part of the object becomes affected (e.g., a woman stopped drawing a flower before it was finished). We hypothesized that the grammaticalization (by case markings) of the resultant state in Estonian affects the perception (encoding) and memorization of causative events. Dutch speakers were included as a control group, as a simple present or

past tense event description in this language does not typically include this specific type of information.

In order to explore the extent of potential effects on cognitive processes induced by the grammatical marking of resultativity in Estonian, the cross-linguistic experiments of the current study employed online and offline methods (eye-tracking and a memory test, respectively), two different experimental task conditions (verbal and nonverbal) and two types of causative events (resultative and non-resultative). Previous empirical findings had suggested a range of potential directions for the interaction between language, perception and experimental condition, such as no influence of language at all, an influence of language in a verbal encoding context, or a context-independent effect of language-specific encoding patterns (e.g., Gennari et al., 2002; Papafragou et al., 2008; Trueswell and Papafragou, 2010; Flecken et al., 2014b; Bunger et al., 2016).

In the current study, we expected an attentional bias towards the action and the objects depicted in the videos in Estonian participants (compared to Dutch), in particular during the final phases of the events' unfolding – our manipulation of completedness (finished/unfinished) would only reveal itself in this time window. We also expected Estonians to show better performance at recalling the end states of the events in the verbal encoding condition. We further anticipated a particularly strong bias in terms of overt attention and better memory for resultative events, as in these cases the marking of the result is obligatory. It was an open question to what extent non-resultative events were processed in the same way: even though differential case marking would not be reflected in the linguistic product, the specifics of the action and the properties of the object matter for the conceptualization and formulation of events, as also in these cases a specific case marker would have to be selected; the system would not be "switched off", so to speak.

It was yet another open question to what extent similar patterns were to be obtained in the nonverbal experiment: here, participants might still rely on linguistically-biased processing patterns, which would indicate an effect of language on global event cognition. On the other hand, an effect of 'thinking for speaking' (i.e., language-specific influence on encoding and memory is present only when given explicit verbalization instructions) was also likely (see Bunger et al., 2016).

First, language production data from the verbal encoding condition confirmed that Estonian speakers were sensitive to the manipulation of both completion state (finished/unfinished) as well as event type in the video stimuli, as they consistently marked the presence or absence of a result at the end of an event by the use of the appropriate case marker. As expected, accusative case was most often applied to the finished versions of Type A items, where the result of the action was achieved within the duration of the video. This contrasted with a bias towards using partitive case to describe the unfinished versions of Type A items, and both finished and unfinished versions of Type B items, where no result was reached. Admittedly, the patters were clear but not absolute: this shows that the videos left room for free interpretation, and that our manipulation was not fully constraining. Note that even when a resultative action is displayed as finished, partitive case is still not ungrammatical, as for a large part of the event that had to be described, it was in fact unfinished. Some participants on

certain trials may have finished their conceptualization and formulation of the events before the videos had completely ended – there was nothing in the instructions which forced them to wait until the final phases before initiating speech preparation processes. Importantly however, the overall patterns confirmed out expectations for Estonian participants; furthermore, Dutch speakers did not show any consistent linguistic patterns in marking the incomplete or complete status of the event.

Analysing the patterns of eye-gaze fixations during the encoding phase of the video stimuli allowed us to explore whether the presence or absence of obligatory encoding of resultativity in the given context bears an effect on how speakers of the two languages gather information about the dynamic events, either when preparing to describe them or while performing a nonverbal sound-cue detection task. We were particularly interested in the eye movements of the participants during a time window at the end of each video, where crucial information about the resultant state of the event was conveyed, and on the blank screen immediately after it had finished playing (i.e., end of video +/- 300 ms).

In the current study, language-related preferences of attention allocation to the Agent and Action event components did surface in the eye-gaze data: the difference between the total number of fixations to the two areas of interest (Action minus Agent fixations) during the relevant time window was significantly greater for the Estonian group than for the Dutch group in the verbal encoding condition, indicating increased attention to the Action region over the Agent region ("action-bias") at the end phase of the event for this group of speakers. However, the same finding did not surface in the data from the nonverbal encoding condition, where at least in the selected time window, both language groups divided their attention between the Agent and the Action in a similar manner. When exploring condition effects within languages more closely, we found that Estonian speakers showed a greater action-bias in the verbal condition than in the nonverbal condition, but no such difference surfaced in the Dutch data. This finding provides evidence for the thinking for speaking hypothesis (Slobin, 1996), which postulates that the more salient and readily encodable constructions in a language influence the way in which speakers allocate their attention to the relevant event components, but that this influence is restricted to the context of preparing to speak. With respect to the type of event (Type A: resultative vs. Type B: non-resultative), participants from both language and condition groups appeared to show a larger action-bias for Type A items than Type B items. This suggests that the features of the resultant objects were universally more salient to the viewers.

The data from the memory task corresponded with the eye-gaze patterns in that the Estonian group displayed an advantage for remembering the end states of causative events after having overtly described them: when contrasting the languages against each other within each condition, the Estonian group proved to have higher accuracy scores than the Dutch group in the verbal condition but not in the nonverbal condition. Note that the reaction time (RT) values of the two groups were not significantly different. The analysis further revealed a surprising advantage for Dutch speakers in the nonverbal condition, which was significant in terms of both accuracy and RT values. We will return to plausible explanations for this finding below.

When exploring condition and event type effects for each language individually, we found that Estonian speakers had significantly higher accuracy scores and shorter RTs in the verbal condition compared to the nonverbal condition, and that the group appeared to perform better in both aspects with Type A items than with Type B items, regardless of condition. In other words, Estonian speakers seemed to remember better the end state of those activities to which the case-marking alternation applied and had to be activated during the task (finished/unfinished versions of Type A items in the verbal task) and to those same items in the nonverbal task. The Dutch group did not demonstrate significantly different performance in the verbal and the nonverbal condition, nor did the group display a differentiation between Type A and Type B items in their accuracy scores (note the analysis did show a trend for faster reaction times for Type A items in both encoding conditions). These findings indicate that verbalizing the event resulted in a significant boost in performance for the Estonian group, but not for the Dutch group; moreover, an additional advantage of Type A items for Estonian speakers was clear in both verbal and nonverbal conditions. While an overall advantage for the Estonian group in the verbal condition once again confirms the thinking for speaking hypothesis, the latter finding hints at a more global influence of language on cognition, suggesting that Estonian speakers were committing resultative (Type A) items into memory in a different manner from non-resultative (Type B) events that are not sensitive to case-alternation, even outside the context of language.

Importantly, the two control tasks administering general visuo-spatial and verbal working memory ability did not show any differences between the tested participant groups. This shows that language and condition effects could not be attributed to individual or group-related differences in the cognitive skills tested (attention and memory).

Before proceeding to conclusions about the influence of language on perception, some methodological limitations of the present study will have to be discussed. For one, although the inclusion of the control tasks was an important addition to the study, the between-subject design does not enable asserting with confidence that the findings can be attributed to the native language of the participants, and not to the variation in individual differences between the groups. Visuo-spatial and verbal working memory are only a few factors that could have influenced task behaviour; most importantly, the inclusion only a single language pair does not enable ruling out potential cultural differences between the groups. This applies to the nonverbal task in particular, where the results from the memory task showed an advantage for the Dutch speakers over the Estonian speakers: the two participant groups may have interpreted the task requirements in a different way, or paid attention to the sound-cue task to a varying extent. On a different note, it is somewhat difficult to position the findings in the context of previous studies, as the verbalizations of the events were collected offline (following Papafragou et al., 2008), meaning that the participants were explicitly instructed to wait before starting to speak. Such a time lag between event apprehension and speech onset does not allow targeting potential language-based differences in the process of conceptualization in particular (i.e., early component of language production, cf. Levelt,

1989), which has been the topic of most of the previous language-and-cognition studies (e.g., von Stutterheim et al., 2012; Flecken et al., 2014a; Bunger et al., 2016).

Nevertheless, the study contributes to our current knowledge about language-induced effects on event perception in several aspects. First, it improved upon the methodology of previous studies by using video-clips rather than pictures (Bunger et al., 2016) or clip-art scenes (Papafragou et al., 2008), which created a more realistic and ecologically valid experimental setting, and by including a larger number of items in each condition (54 videos, compared to 12 in Bunger et al., 2016 and Papafragou et al., 2008). It was also the first study to date to include a verbal and nonverbal condition as well as control measures of general working memory abilities. More importantly however, it broadened the scope of the empirical basis of event encoding theories by exploring the domains of resultativity and causative events, both of which have not been subject to enough research in the past. We inspected a grammar-based rather than lexical contrast between languages, which has previously been hypothesized to be more susceptible to "deeper" influence by language (Lucy, 1992). Our results mostly revealed *thinking for speaking* effects (Slobin, 1996), proving that the task of speaking alters the allocation of cognitive resources and varies depending on the language background: in the context of verbal description of the videos, differential object case-marking in Estonian induced heightened attention to the relevant visual region of the videos by Estonian speakers. The results of the memory task further hinted at a context-independent, potentially habitual and automatic attention to resultativity for the Estonian participants, showing superior memory of action-related properties of resultative events over non-resultative events in both verbal and nonverbal encoding conditions for this group of speakers. In order to further specify the processes in hand, additional research is needed.

APPENDICES

APPENDIX A

Table 2: Complete list of stimuli items

Type A items	Type B items	fillers
break chocolate	beat cream	bandage a hand
build a lego tower	clean glasses	blow one's nose
crunch paper	clean knife	calculate
cut an apple	clean mirror	dance
cut a circle	cut nails	give a flower
draw a flower	grate cheese	juggle
fold a paper plane	put cream on hands	look in mirror
cut paper in half	knit a scarf	put a book on one's head
open a can	measure a box	put a book on the table
open a jam jar	mix cards	put on a hat
open a letter	play a drum	shake hands
peel banana	polish glass	sleep with head on the table
peel a mandarin	pour water from flask	stretch muscles
peel a potato	read a book	take someone's glasses
pour coke	salt soup	talk on the phone
put a puzzle together	staple paper	throw ball
roll wool	stir soup	throw paper in the bin
tear paper	wipe table	yawn

REFERENCES

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

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of memory and language*, 59(4):457–474.

Boogaart, R. (1999). *Aspect and temporal ordering: A contrastive analysis of Dutch and English*. Den Haag: Holland Academic Graphics.

Boroditsky, L. (2011). How language shapes thought. *Scientific American*, 304(2):62–65.

Bunger, A., Skordos, D., Trueswell, J. C., and Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: Implications for language production and attention. *Language, Cognition and Neuroscience*, pages 1–23.

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

Carroll, M. and von Stutterheim, C. (2011). Event representation, event-time relations and clause structure: A cross-linguistic study of English and German. In Pederson, E. and Bohnemeyer, J., editors, *Event representation in Language and Cognition*, pages 68–83. Cambridge: Cambridge University Press.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*, volume 2. Cambridge university press.

Fausey, C. M. and Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic bulletin & review*, 18(1):150–157.

Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, 15(4):466–485.

Finkbeiner, M., Nicol, J., Greth, D., and Nakamura, K. (2002). The role of language in memory for actions. *Journal of Psycholinguistic Research*, 31(5):447–457.

Flecken, M., Athanasopoulos, P., Kuipers, J. R., and Thierry, G. (2015a). On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition*, 141:41–51.

Flecken, M., Carroll, M., Weimar, K., and Von Stutterheim, C. (2015b). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French–German L2 users. *The Modern Language Journal*, 99(S1):100–122.

Flecken, M., Gerwien, J., Carroll, M., and von Stutterheim, C. (2014a). Analyzing gaze allocation during language planning: A cross-linguistic study on dynamic events. *Language and Cognition*, 7(1):138–166.

Flecken, M., von Stutterheim, C., and Carroll, M. (2014b). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition*, 6(01):45–78.

Gennari, S. P., Sloman, S. A., Malt, B. C., and Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83(1):49–79.

Giannakidou, A. and Merchant, J. (1999). Why Giannis can't scrub his plate clean: On the absence of resultative secondary predication in Greek. In *Proceedings of the 3rd international conference on Greek linguistics*, pages 93–103.

Gleitman, L., January, D., Nappa, R., and Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of memory and language*, 57(4):544–569.

Gleitman, L. and Papafragou, A. (2005). Language and thought. In Holyoak, K. and R, M., editors, *Cambridge Handbook of Thinking and Reasoning*, pages 633–661. New York: Cambridge University Press.

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

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In Bloom, P., Peterson, M., and Garrett, M., editors, *Language and space*, pages 1–30. Cambridge, MA: MIT Press.

Ji, Y., Hendriks, H., and Hickmann, M. (2011). How children express caused motion events in Chinese and English: Universal and language-specific influences. *Lingua*, 121(12):1796–1819.

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., and De Haan, E. H. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied neuropsychology*, 7(4):252–258.

Kiparsky, P. (1998). Partitive case and aspect. In Butt, M. and Geuder, W., editors, *The projection of arguments: Lexical and compositional factors*, volume 265, pages 265–308.

Klein, W. (1994). *Time in language*. London: Routlege.

Klemfuss, N., Prinzmetal, B., and Ivry, R. B. (2012). How does language change perception: a cautionary note. *Frontiers in psychology*, 3:78.

Klettke, B. and Wolff, P. (2003). Differences in how English and German speakers talk and reason about CAUSE. In Alterman, R. and Kirsh, D., editors, *Proceedings of the 25th annual conference of the Cognitive Science Society*, pages 675–680. Mahwah, NJ: Lawrence Erlbaum Associates.

Lakusta, L. and Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive science*, 36(3):517–544.

Lees, A. (2004). Partitive-accusative alternations in Balto-Finnic languages. In Muskovsky, C., editor, *Proceedings of the 2003 Conference of the Australian Linguistic Society*.

Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. *Language and space*, pages 109–169.

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. A. (2011). *Language and cognition: the view of anthropology*. New York: Psychology Press.

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

Metslang, H. (2013). *Grammatical relations in Estonian: subject, object and beyond*. PhD thesis, University of Tartu.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1):28.

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

Papafragou, A., Massey, C., and Gleitman, L. (2002). Shake, rattle,'n'roll: The representation of motion in language and cognition. *Cognition*, 84(2):189–219.

Papafragou, A. and Selimis, S. (2010). Event categorisation and language: a cross-linguistic study of motion. *Language and cognitive processes*, 25(2):224–260.

Pavlenko, A. and Volynsky, M. (2015). Motion encoding in Russian and English: Moving beyond Talmy's typology. *The Modern Language Journal*, 99(S1):32–48.

Pinker, S. (1995). *The language instinct: The new science of language and mind*, volume 7529. London: Penguin Books UK.

Radvansky, G. A. and Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6):608–620.

Rätsep, H. (1978). *Eesti keele lihtlausete tüübid [Types of simple clauses in Estonian]*. Tallinn: Valgus.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In Levinson, S., editor, *Rethinking linguistic relativity*, pages 70–96. Cambridge: Cambridge University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In Gentner, D. and Goldin-Meadow, S., editors, *Language in mind: Advances in the study of language and thought*, volume 157–192. Cambridge, MA: MIT Press.

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

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In Shopen, T., editor, *Language typology and syntactic description*, volume 3, pages 57–149. Cambridge: Cambridge University Press.

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

Tamm, A. (2004). Estonian transitive verb classes, object case, and progressive. *Nordlyd*, 31(4).

Tauli, V. (1968). Totaalobjekt eesti kirjakeeles [The total object in written Estonian]. *Suomalais-ugrilaisen Seuran Toimituksia*, pages 216–224.

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

Trueswell, J. C. and Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63(1):64–82.

Ünal, E. and Papafragou, A. (2016). Interactions between language and mental representations. *Language Learning*, 66(3):554–580.

Van Beek, G., Flecken, M., and Starren, M. (2013). Aspectual perspective taking in event construal in L1 and L2 Dutch. *International Review of Applied Linguistics in Language Teaching*, 51(2):199–227.

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., and 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.

von Stutterheim, C. and Nuse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics*, 41(5; ISSU 387):851–882.

Wolff, P. and Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3):253–265.

Wolff, P. and Ventura, T. (2009). When Russians learn English: How the semantics of causation may change. *Bilingualism: Language and Cognition*, 12(02):153–176.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., and Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of clinical and experimental neuropsychology*, 33(1):101–111.

Zacks, J. M. and Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2):80–84.

Zacks, J. M. and Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin*, 127(1):3.