Mapping language onto mental representations of object locations in transfer-of-possession events: A visual-world study using webcam-based eye-tracking

Sarah Hye-yeon Lee (sarahhl@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Philadelphia, PA 19104-6228 USA

Elsi Kaiser (emkaiser@usc.edu)
Department of Linguistics, University of Southern California
3601 Watt Way, Grace Ford Salvatori 301, Los Angeles, CA 90089-1693 USA

Abstract

Source-goal events involve an object moving from the Source to the Goal. In this work, we focus on the representation of the *object*, which has received relatively less attention in the study of Source-goal events. Specifically, this study aims to investigate the mapping between language and mental representations of object locations in transfer-of-possession events (e.g. *throwing*, *giving*). We investigate two different grammatical factors that may influence the representation of object location in transfer-of-possession events: (a) grammatical aspect (e.g. *threw* vs. *was throwing*) and (b) verb semantics (guaranteed transfer, e.g. *give* vs. no guaranteed transfer, e.g. *throw*). We conducted a visual-world eye-tracking study using a novel webcam-based eye-tracking paradigm (*Webgazer*; Papoutsaki et al., 2016) to investigate how grammatical aspect and verb semantics in the linguistic input guide the real-time and final representations of object locations. We show that grammatical cues guide the real-time and final representations of object locations.

Keywords: transfer-of-possession event; event cognition; event representation; object location; visual-world eye-tracking

Introduction

Source-goal events involve an object (*Figure*) moving from its starting point/origin (*Source*) to an endpoint (*Goal*) along a *Path* (e.g. Jackendoff, 1983; Talmy, 1983, 1985). Prior literature in the domain of Source-goal events has largely focused on the fundamental differences between the Source and the Goal – the goal bias (e.g. Do, Papafragou, & Trueswell, 2020; Ihara & Fujita, 2000; Kang, Eerland, Joergensen, Zwaan, & Altmann, 2020; Lakusta & Landau, 2005; Lakusta & Landau, 2012; Papafragou, 2010; Regier, 1996; Regier & Zheng, 2007; Stefanowitsch & Rohde, 2004). However, in order to fully understand how we represent Source-goal events, we also need to understand how we conceptualize and represent the object (figure). In the current work, we investigate the mental representation of the object undergoing change-of-location.

The notion of *change* is central to understanding event representations. There is a growing body of literature focusing on the dimension of event representation related to object state change (e.g. Altmann & Ekves, 2019; Horchak & Garrido, 2021; Ji & Papafragou, 2020a, 2020b; Lee & Kaiser, 2021; Misersky, Silvac, Hagoort, & Flecken, 2021; Sakarias & Flecken, 2019). However, objects can undergo changes in physical *location* as well as in physical states. In linguistics, it has been discussed that there are conceptual similarities between change-of-state and change-of-location (e.g. Gropen et al., 1991; Gruber, 1965; Pustejovsky, 1991). It has also been shown that the cognitive system can not only track multiple representations of the same object as it undergoes change-of-state (e.g. Altmann & Ekves, 2019; Altmann & Kamide, 2007; Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Kang et al., 2020; Solomon, Hindy, Altmann, & Thompson-Schill, 2015) but also as it undergoes change-of-location (e.g. Altmann & Kamide, 2009). In Altmann and Ekves' (2019) Intersecting Object Histories theory, changes in location are understood as a type of change relevant for the representation of object histories.

The current work builds on these insights and aims to investigate the mapping between language and mental representations of object locations. We investigate two grammatical factors that may influence the representation of object location in transfer-of-possession events: (a) grammatical aspect (e.g. *threw* vs. *was throwing*) and (b) verb semantics (guaranteed transfer, e.g. *give* vs. no guaranteed transfer, e.g. *throw*). How do these grammatical factors influence the mapping from language to object location representations? Another aim is to investigate whether the representation of object locations can be dynamically updated during incremental sentence processing. Do comprehenders update their mental representations of object location in real-time as the linguistic input unfolds?

To investigate these questions, we conducted a visual-world eye-tracking study using a novel webcam-based eye-tracking paradigm (*Webgazer*; Papoutsaki et al., 2016). This

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

allowed us to probe how visual attention reflects the real-time mapping of linguistic input onto mental representations of object locations in events where the object undergoes change-of-location from the Source to the Goal. We focus specifically on *transfer-of-possession events*, which are transitive events that involve an object being transferred from a Source/agent to a Goal/patient (e.g. throwing a ball to someone, giving a gift to someone). Using transfer-of-possession events allowed us to investigate effects of grammatical aspect and verb semantics.

This work aims to shed light on how different grammatical properties of an utterance interact in real-time to dynamically update comprehenders' mental representation of changing situations and to ultimately lead to a final understanding of the event. In the following sections, we provide a brief background on the grammatical factors that are under investigation in this study.

Grammatical aspect

Grammatical aspect – in particular the distinction between perfective and imperfective aspect – provides information about whether the described event is represented as completed or ongoing (e.g. Comrie, 1976). When an event is described in perfective aspect, the event is viewed as a completed whole. In contrast, an event described in imperfective aspect is viewed as ongoing and incomplete, with reference being made to the internal temporal phases that make up the event.

Our study aims to investigate how grammatical aspect guides the mental representation of the location of the object being transferred. More specifically, does the completed vs. ongoing distinction lead to different representations of the object's location?

Verb semantics

Transfer-of-possession verbs can be classified into different classes, based on whether the verb's lexical semantics entails (i.e., semantically guarantees) successful transfer or not (e.g. Rappaport Hovav and Levin, 2008). With verbs that entail successful transfer (*give*-type verbs; e.g. *give*, *hand*), it is infelicitous to assert that the transfer was unsuccessful, as demonstrated by the infelicity of the continuation in "Kim gave her brother the ball, # but he never received it." These verbs differ from verbs that do *not* guarantee successful transfer (*throw*-type verbs; e.g. *throw*, *toss*), for which it is felicitous to deny that the transfer was unsuccessful, as in "Kim threw her brother the ball, but he never received it."

In the current study, we investigate whether and how different verb classes with different entailments (*give*-type verbs and *throw*-type verbs) contribute to the mental representations of object locations. More specifically, does the presence vs. absence of the successful transfer entailment lead to different representations of the object's location?

Experiment

To investigate how grammatical aspect and verb semantics in the linguistic input guide the real-time and final representations of object locations, we conducted a visual-world eye-tracking study where participants heard descriptions of transfer-of-possession events and were asked to click on where they think the object is in the scene.

Participants

Participants were recruited on the internet via Prolific and received $5 for participating. All participants were native speakers of American English. 68 participants completed the study. We excluded 12 participants from the analyses due to: visual or hearing impairments (2 participants), poor accuracy on attention check trials (1 participant; 8.33% accuracy; average accuracy of included participants=99.85%), poor calibration (4 participants; initial calibration score < 60, mean pre-trial calibration score < 45), and failure to follow instructions to click on the fixation cross (5 participants). All exclusion criteria were determined prior to data analysis. We included data from 56 participants in the data analysis.

Design and materials

Auditory stimuli All sentences used in the experiment were in past tense. Target sentences (See (1)-(2)) contained a transfer-of-possession verb and animate Source and Goal individuals that differed in gender. The Source and the Goal characters' names differed on each trial. The target stimuli varied on verb type and grammatical aspect. Verb type was manipulated between-items and grammatical aspect was manipulated within-items. For completeness, we used sentences with both dative argument realization patterns (VERB *the ball to GOAL &* VERB *GOAL the ball*).

(a) *Verb type*: Three different of give-verbs (*give*, *hand*, *bring*) and throw-type verbs (*throw*, *kick*, *toss*) were used to develop the target sentences. Each verb was used to create four different items. Each item varied on grammatical aspect and argument realization pattern forms. To account for the two different argument realization patterns used in the study, we only used verbs that sounded natural with both. None of the verbs triggered a strong inference about non-canonical types of balls. These decisions were based on information provided by native speaker informants.

(b) *Grammatical aspect*: Each sentence was presented either in perfective (simple past; e.g. *gave*, *threw*) or imperfective aspect (past progressive; e.g. *was giving*, *was throwing*).

(1) *Give*-type example stimuli
   a. Liam was giving the ball to Paige. [imperfective]
   b. Liam gave the ball to Paige. [perfective]

(2) *Throw*-type example stimuli
   a. Liam was throwing the ball to Paige. [imperfective]
   b. Liam threw the ball to Paige. [perfective]

In addition to 24 target items, the experiment included 34 filler items. Twelve of the fillers also functioned as attention check trials (i.e. had clearly expected click locations). Filler

sentences mentioned a ball or a bird whereas targets always mentioned a ball. In fillers, the ball/bird was in the sentence subject or object position. Sentences were recorded by a female native speaker of American English, using Praat.

Display (Visual scenes) Visual scenes were created to accompany the auditory stimuli (See Figure 1). The scenes depicted a Source and a Goal character, illustrated with stick figures, whose positions (right/left) on the screen were counterbalanced throughout the experiment. The arrow next to the Source character indicated the direction of the ball's movement (i.e., signaled who the Source was); this was explained to participants. Crucially, the ball was not visible on the scene. Instead, participants were asked to click on *where they think the ball is* on the scene. Detailed wording provided to participants is given in (3) and discussed more below.

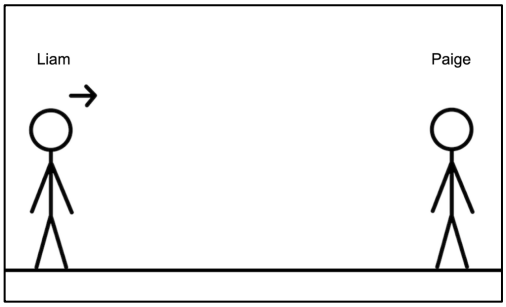

Figure 1: Sample target visual stimuli

Areas of interest To measure the location of the participants' clicks and looks, we look at three areas of interest: the Source area, the Path area, and the Goal area (See Figure 2). The Source and the Goal areas are identical in width, each occupying one fifth of the entire screen's width. The Path area occupies the rest of the screen.

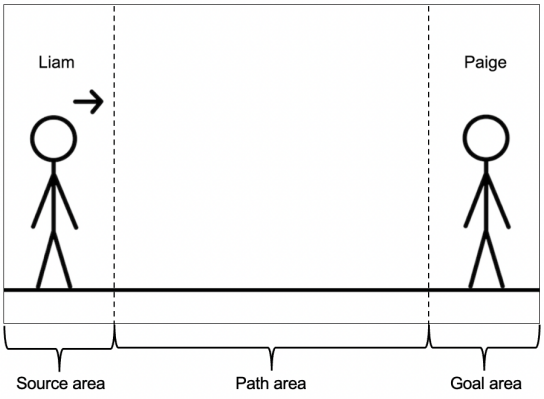

Figure 2: Areas of interest

The Goal area was used to test predictions about whether the object is represented as having reached the Goal or not. In order to test predictions about whether the object is represented as being located on the Path between the Source and the Goal, we look at clicks and looks to the Path area.

Procedure The experiment was hosted online on PennController IBEX (Zehr & Schwarz, 2018), and participants did it remotely via the internet. Eye gaze data was collected using the Webgazer.js library (Papoutsaki et al., 2016), an open-source webcam-based eye-tracking JavaScript library which uses the participant's webcam to compute gaze position.

Each trial started with a central fixation cross. Participants were asked to look at and click on the fixation cross. Participants were given 3.5 seconds to click on the cross before the experiment automatically proceeded to the visual scene. The visual scene (Figure 1) appeared 1000ms after clicking the cross. The audio started 1000ms after the visual scene appeared. Participants were asked to imagine that the world is in a freeze-frame during the moment described by the sentence, and to **click on** where they think the ball is in the scene. The full wordings for these instructions are provided in (3). We did not want participants to think about the location of the ball at the moment the sentence is uttered, as this kind of interpretation may mislead participants to think about the possibility of the presence of other intervening events that may have caused the ball to move after the throwing event. Therefore, we included the phrase "during the moment described by the sentence" to encourage participants to construct event representations relevant to the temporal interval that is discussed by the sentence.

(3) You will hear a sentence, for example "The ball is near Mason". However, the ball is not visible. Now let's imagine that we freeze the world during the moment described by the sentence. Where do you think the ball is in the scene? Your task is to use your mouse to click where you think the ball is.

After the audio finished playing, participants were given 5 seconds to provide their click response before the experiment automatically proceeded to the next trial. Upon clicking, the trial ended and the next trial started after a 250ms pause. The experiment began with two practice trials. The experiment lasted around 20 minutes. Participants' eye movements were recorded during the entire trial (from the onset of the fixation cross until a final click was made), along with their mouse click region and timing.

Predictions

In this section, we discuss predictions regarding (i) the final/post-sentential interpretation, which is informed by the click data and (ii) real-time processing, which is informed by eye gaze data.

Predictions about post-sentential interpretations One goal of this study is to investigate how different grammatical factors contribute to the final interpretations of transfer-of-possession events, in particular the representation of object

location. We consider two non-mutually-exclusive hypotheses regarding the grammatical factors we tested (grammatical aspect, verb semantics).

1. Grammatical Aspect Hypothesis: The Grammatical Aspect Hypothesis states that the mental representations of object locations are guided by grammatical aspect information. This hypothesis predicts that upon hearing imperfective sentences (e.g. *…was throwing*…), comprehenders will construe the event as ongoing, leading to an event representation where the ball may have not reached the final goal yet (e.g. The source individual may still have it, or it may be mid-air.) Perfective sentences (e.g. *…threw*…), however, are predicted to be more likely to elicit a completed event representation where the ball is at its final location.

If the Grammatical Aspect Hypothesis is on the right track, we expect the proportion of Goal area clicks to be greater in the perfective than in the imperfective aspect condition.

2. Verb Semantics Hypothesis: According to the Verb Semantics Hypothesis, verbs' entailment patterns constrain the mental representation of events that comprehenders construct, such that *give*-type verbs, which entail successful transfer, give rise to event representations where successful transfer occurs – i.e. the ball successfully ends up at the Goal. Conversely, *throw*-type verbs, which do not entail successful transfer, may give rise to event representations in which the ball does not successfully end up at the Goal and ends up on the Path between the Source and the Goal.

The Verb Semantics Hypothesis predicts that there will be more clicks to the Path area (and conversely, fewer clicks to the Goal area) in sentences with *throw*-type verbs than in sentences with *give*-type verbs. We interpret clicks on the Path area as indicating that the participant constructed an event representation where the transfer was not successful (i.e. did not reach the goal.)

Predictions about real-time processing In addition to assessing end-of-sentence interpretation, this experiment also investigates whether the mental representation of object locations are dynamically updated during real-time sentence processing. Do comprehenders update mental representations of object location in real-time as the linguistic input unfolds? Specifically, we are interested in whether the object location representations are updated during the *verb phrase*. We assume that eye movements will provide a measure of where participants think the object (the ball) is in the scene. Observing participants' eye gaze as the sentence unfolds can shed light on their dynamically changing mental representations (e.g. Tanenhaus et al., 1995).

The two hypotheses outlined above are also relevant here. If mental representations of object locations are dynamically updated in real-time during the unfolding of linguistic input, we expect to see that the patterns outlined for each hypothesis above will be reflected in eye movements *while participants are still listening to the verb phrase*. If grammatical aspect cues are considered in real-time to update the mental representation of object location, it can be predicted that the proportion of Goal looks will be greater in the perfective aspect condition than in the imperfective aspect condition, *during the unfolding of the verb phrase*. The same logic goes for the Verb Semantics Hypothesis. If verb semantics cues guide the real-time updating of the mental representation of object location, it can be predicted that the proportion of Path looks will be greater in the *throw*-type verb condition than in the *give*-type verb condition.

Data processing and analysis

Click data In order to investigate the hypotheses regarding the effects of grammatical cues on participants' final interpretations, as indicated by click locations, we conducted two different statistical analyses on the click data. First, we conducted analyses on the proportions of Goal area clicks in different conditions. Second, we conducted analyses on the proportions of GOAL region clicks in different conditions. This analysis allows us to test the Grammatical Aspect Hypothesis. Second, we conducted analyses on the proportions of Path area clicks in different conditions. This analysis informs us of the Verb Semantics Hypothesis.

For statistical analyses, we used Generalized Linear Mixed Effects models (*glmer*) with grammatical aspect, verb type, grammatical aspect x verb type interaction, and argument realization pattern as fixed effects. We used the maximal random effect structure justified by model comparison.

Eye gaze data To investigate whether the mental representation of object location is dynamically updated during the unfolding of the sentence, we looked at the time window from the onset of the verb phrase (e.g. *was throwing*, *threw*) to the end of the sentence. The analysis time window was offset by 300ms, instead of the usual 200ms, given that it has been reported that there is a systematic delay in Webgazer recordings (e.g. approximately 300ms additional delay in Slim & Hartsuiker, 2021). In order to test the Grammatical Aspect Hypothesis, we conducted analyses on the proportion of looks to the Goal area. In order to test the Verb Semantics Hypothesis, we conducted analyses on the proportion of looks to the Path area.

Statistical analyses were conducted using the *lme4* package (version 1.1.26) (Bates et al., 2015) and *lmerTest* (version 3.1.3) (Kuznetsova et al., 2017) in the R software environment (R Development Core Team, 2019). As fixed effects, we used grammatical aspect, verb type, grammatical aspect x verb type interaction, and argument realization pattern. We used the maximal random effect structure justified by model comparison.

Results

Post-sentential interpretations Figure 3 shows the proportion of clicks on each of the areas (Source, Path, and Goal), by verb type and grammatical aspect.

Two different patterns can be seen in Figure 3. First, there are more Goal area clicks in perfective conditions (*gave*, *threw*) than in imperfective conditions (*was giving*, *was*

throwing). Second, we also find more Path area clicks in *throw*-conditions than in *give*-conditions. We additionally see an overall Source preference, which we attribute to various reasons, including the visual salience of the source character due to the presence of the arrow. Because the overall preference for the Source is not relevant for our hypotheses or the main claims being made in this chapter, we do not discuss it further.

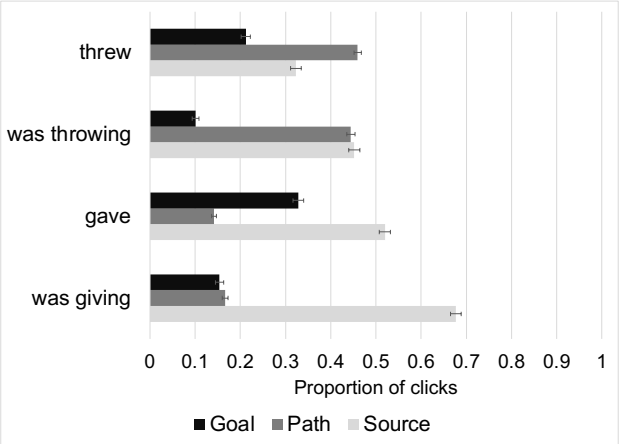

Figure 3: Proportion of clicks on each area of interest (Error bars show +/- 1 SE)

In terms of the proportion of **Goal area clicks**, there is a main effect of grammatical aspect ($p < 0.05$): Perfective sentences elicited more Goal area clicks than imperfective sentences. These results support the Grammatical Aspect Hypothesis.

In terms of the proportions of **Path area clicks**, there was a main effect of verb type ($p < 0.05$): Sentences with *throw*-type verbs elicited more Path area clicks than sentences with *give*-type verbs, supporting the Verb Semantics Hypothesis.

Real-time processing In order to test the Grammatical Aspect Hypothesis, we analyzed the proportion of looks to the Goal area. Figure 4 shows the proportions of looks to the Goal area in imperfective aspect trials (red) and in perfective aspect trials (blue), relative to the onset of the verb phrase (e.g. *was throwing*, *threw*). The area within the dashed lines represent the time window that we are interested in (from the onset of the verb phrase to the end of the sentence). For the purposes of plotting Figures 4 and 5, the dashed line that represents the offset of the sentence was averaged across trials. However, in the analyses, we used the exact offset of each sentence. In our analyses, the time window was offset by 300ms.

Figure 4 shows that after the onset of the verb phrase, the proportion of Goal area looks is greater in perfective aspect trials than in imperfective aspect trials.

We conducted statistical analyses on the proportion of looks to the Goal area and found a main effect of grammatical aspect ($p < 0.05$): The proportions of looks to the Goal area were greater in perfective aspect sentences than in imperfective aspect sentences. These results support the Grammatical Aspect Hypothesis.

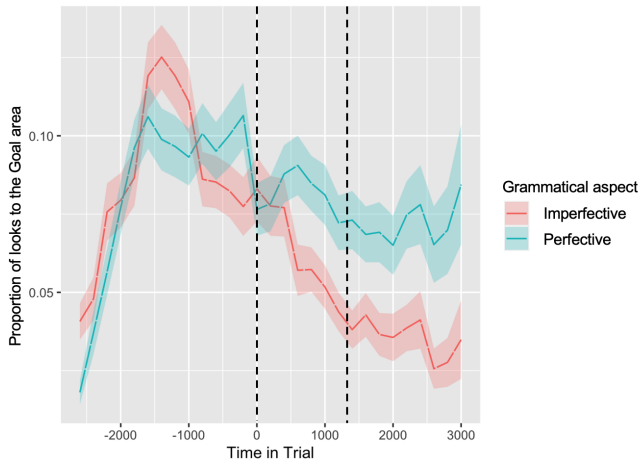

Figure 4: Proportions of Goal area looks by grammatical aspect; 0 on the x-axis indicates the onset of the verb phrase; Data is collapsed by participant for plotting

In order to test the Verb semantics Hypothesis, we analyzed the proportion of looks to the Path area. Figure 5 shows the proportion of looks to the Path area in *give*-type verb trials (red) and in *throw*-type verb trials (blue), relative to the onset of the verb phrase (e.g. *was throwing, threw*). Again, the area within the dashed lines represent the time window of interest.

Figure 5 shows that after the onset of the verb phrase, the proportion of Path area looks is greater in *throw*-type sentences than in *give*-type sentences.

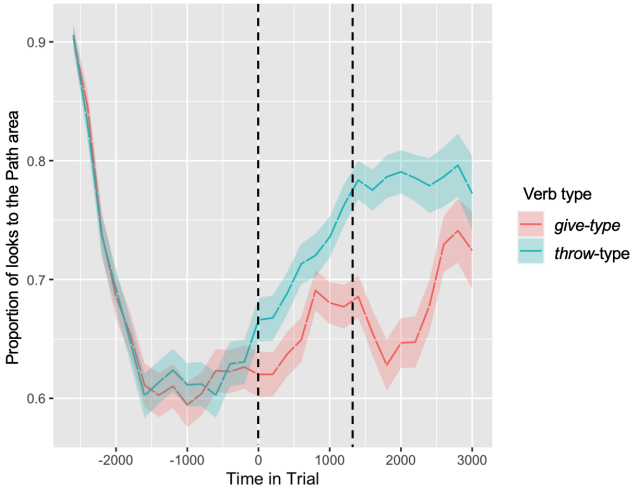

Figure 5: Proportions of Path area looks by verb type; 0 on the x-axis indicates the onset of the verb phrase; Data is collapsed by participant for plotting

We conducted statistical analyses on the proportion of looks to the Path area and found a main effect of verb type ($p < 0.05$): The proportions of looks to the Path area were greater in *throw*-type sentences than in *give*-type sentences. These results support the Grammatical Aspect Hypothesis.

Taken together, the eye gaze data suggest that the mental representations of object locations are reflected in the real-time eye movements during the unfolding of the sentence.

Discussion

This study sheds light on the cognitive processes involved in the representation of the object in Source-Goal events. How does language guide the mental representations of object location during real-time language processing of transfer-of-possession events? We investigated two grammatical factors that can influence the construction of object location representations: grammatical aspect and verb semantics. The study was designed to test how grammatical cues impact real-time eye movements to the visual scene and the final interpretation of the event.

The first aim of the study was to assess how different grammatical cues influence the final object location representations that comprehenders post-sententially reach at. Our results support the Grammatical Aspect Hypothesis, according to which the ongoing vs. completed event representations lead to different representations of object locations. The click data suggests that in perfective aspect sentences, comprehenders are more likely than in imperfective aspect sentences to construct an event representation in which the object is located at the Goal, whereas in imperfective sentences, they are more likely than in perfective sentences to construct an event representation where the object is not at the Goal.

The results also support the Verb Semantics Hypothesis, suggesting that verbs' entailment patterns contribute to the event representations that comprehenders construct. When the sentence describes a *throwing* event, comprehenders are more likely to construct an event representation in which the object is located in the middle ground than when the sentence describes a *giving* event. These results suggest that the semantic differences between the two verb classes (the presence vs. absence of the successful transfer entailment) lead to different representations of object locations. In sum, comprehenders consider verb semantics and grammatical aspect as cues to guide the final representation of the event.

The second aim of the study was to investigate whether the mental representation of object locations can be dynamically updated during incremental sentence processing. Our eye-tracking data indicate that object locations are updated in real-time while the sentence unfolds. The data suggest that grammatical aspect is indeed a cue that comprehenders consider in order to dynamically update the object location representations during the unfolding of the sentence. While listening to the verb phrase in perfective aspect sentences, comprehenders are more likely to look at the Goal than when they are listening to imperfective aspect sentences (Grammatical Aspect Hypothesis). While listening to the verb phrase in sentences with *throw*-type verbs, comprehenders are more likely to look at the Path than when listening to sentences with *give*-type verbs (Verb Semantics Hypothesis).

That is, the process of language getting mapped onto mental event representations is a dynamic, real-time process. This finding is in line with prior work by Altmann and Kamide (2009), where they showed that in comprehending sentences like *The woman will put the glass on the table*, the event representations of object locations are dynamically updated. Our study further shows that a temporal-semantic grammatical cue such as grammatical aspect is a relevant cue during this dynamic process.

This study uses a novel webcam-based eye-tracking paradigm and shows that it can provide a useful way to collect eye-tracking data. More research is needed to better understand the exact nature of the time latencies and the factors that attribute to them, but this study provides some initial confirmation that the method provides data that can be informative for psycholinguistic research.

Acknowledgments

This study is supported by the National Science Foundation under the Doctoral Dissertation Research Improvement Grant #BCS-2041261. We thank Mila Mathias and Jesse Storbeck for help designing and implementating the experiment. We also thank anonymous CogSci reviewers for invaluable suggestions and feedback.

References

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817-840.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502-518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55-71.

Comrie, B. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.

Do, M. L., Papafragou, A., & Trueswell, J. (2020). Cognitive and pragmatic factors in language production: Evidence from source-goal motion events. *Cognition*, *205*, 104447.

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, *41*(1), 153-195.

Gruber, J. S. (1965). *Studies in lexical relation*. Doctoral dissertation, Massachusetts Institute of Technology.

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on

event processing: do objects compete with themselves? *The Journal of Neuroscience*, *32*(17), 5795-5803.

Horchak, O. V., & Garrido, M. V. (2021). Dropping bowling balls on tomatoes: Representations of object state-changes during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(5), 838-857.

Ihara, H., & Fujita, I. (2000). A Cognitive Approach to Errors in Case Marking in Japanese Agrammatism: The Priority of the Goal -ni over the Source -kara. *Constructions in Cognitive Linguistics*, *123*.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Ji, Y., & Papafragou, A. (2020a). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition*, *197*, 104197.

Ji, Y., & Papafragou, A. (2020b). Midpoints, endpoints and the cognitive structure of events. *Language, Cognition and Neuroscience*, *35*(10), 1465-1479.

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2020). The influence of state change on object representations in language comprehension. *Memory & Cognition*, 390-399.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, *82*, 1-26.

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

Lakusta, L., & Landau, B. (2012). Language and Memory for Motion Events: Origins of the Asymmetry Between Source and Goal Paths. Cognitive Science, *36*(3), 517-544.

Lee, S. H., & Kaiser, E. (2021). Does hitting the window break it? Investigating effects of discourse-level and verb-level information in guiding object state representations. Language, Cognition and Neuroscience, 36(8), pp. 921-940.

Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition*, *214*, 104744.

Papafragou, A. (2010). Source-Goal Asymmetries in Motion Representation: Implications for Language Production and Comprehension. Cognitive Science, 34(6), 1064-1092.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 3839-3845).

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*(1), 47-81.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rappaport Hovav, M., & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of Linguistics*, *44*(1), 129-167.

Regier, T. (1996). *The Human Semantic Potential: Spatial Language and Constrained Connectionism*. MIT Press.

Regier, T., & Zheng, M. (2007). Attention to Endpoints: A Cross-Linguistic Constraint on Spatial Meaning. *Cognitive Science*, *31*(4), 705-719.

Sakarias, M., & Flecken, M. (2019). Keeping the Result in Sight and Mind: General Cognitive Principles and Language-Specific Influences in the Perception and Memory of Resultative Events. *Cognitive Science*, *43*(1), e12708.

Slim, M. S., & Hartsuiker, R. (2021). Visual world eyetracking using WebGazer.js. PsyArXiv.

Solomon, S. H., Hindy, N. C., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). Competition between mutually exclusive object states in event comprehension. *Journal of Cognitive Neuroscience*, *27*(12), 2324-2338.

Stefanowitsch, A., & Rohde, A. (2004). The goal bias in the encoding of motion events. In K.-U. Panther, & G. Radden (Eds.), *Studies in linguistic motivation*. Berlin: Mouton de Gruyter.

Talmy, L. (1983). How language structures space. In H. Pick, & L. Acredolo (Eds.), *Spatial orientations: Theory, research, and application*. New York: Plenum Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description*. New York: Cambridge University Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).