

The role of grammatical cues in tracking object location in transfer-of-possession events: A visual-world eye-tracking study

Sarah Hye-yeon Lee^{1,2} & Elsi Kaiser¹

¹University of Southern California, ²University of Pennsylvania

Source-goal events involve an *object* (Figure) moving from the Source to the Goal (e.g. [3], [5], [6]). Tracking the changes that objects undergo is fundamental to event comprehension ([2]). We investigate how grammatical properties of transfer-of-possession sentences (grammatical aspect, verb semantics) influence comprehenders' mental representations of object *location* changes during real-time processing. We test two main questions:

(a) How do grammatical factors influence sentence-final object location representations? We consider two non-mutually-exclusive hypotheses: The *Grammatical Aspect Hypothesis* predicts that whether the event is described as ongoing (imperfective aspect: *Liam was throwing ...*) or completed (perfective aspect: *Liam threw ...*) influences people's representation of object location.

The *Verb Semantics Hypothesis* predicts that whether the transfer-of-possession verb entails (semantically guarantees) successful transfer influences representation of object location. *Give*-type verbs (e.g. *give, hand, bring*) entail successful transfer, but *throw*-type verbs (e.g. *throw, kick, toss*) do not (e.g. [4]).

(b) Are mental representation of object locations dynamically updated in real-time? Building on prior work (e.g. [1]), we hypothesize that listeners use grammatical cues to update object location representations as the sentence unfolds.

Experiment We used visual-world eye-tracking to investigate effects of grammatical aspect and verb semantics (Table 1) on representations of object locations. Eyegaze data was collected using webcam-based *Webgazer.js* ([7]) and *PCIBex* ([9]). Participants heard transfer-of-possession sentences (e.g. Table 1) about "the ball" and saw scenes with Source and Goal characters (Fig. 1). The ball was never depicted. Instead, participants were asked to imagine that 'we freeze the world' at the moment described by the sentence they heard, and then to mouse-click on where they think the ball is.

Click data (final interpretations) (N=65) There are more SOURCE region clicks in imperfective than perfective aspect ($z=6.33$, $p<0.0001$; *gImer*); more GOAL region clicks in perfective than imperfective aspect ($z=-6.94$, $p<0.0001$), supporting the *Grammatical Aspect Hypothesis*. To test the *Verb Semantics Hypothesis*, we compare clicks on the Center area (and the MIDDLE region) with *give*- vs. *throw*-type verbs, and we indeed find more MIDDLE clicks with non-guaranteed-transfer *throw*-type verbs (Fig. 3, $z=-8.22$, $p<0.00001$).

Eyegaze data (N=56) Looks to the SOURCE/GOAL regions were analyzed from the beginning of the main verb to the end of the sentence, offset by 400ms (to account for a systematic delay in *Webgazer* recordings, e.g. [8]). Proportions of SOURCE looks were higher in imperfective than in perfective aspect (Fig. 4; $t=2.29$, $p=0.026$, *lmer*). Proportion of GOAL looks were higher in perfective than in imperfective aspect (Fig. 4; $t=-15.85$, $p < 0.0001$). Goal-advantage scores (=GOAL minus SOURCE looks) revealed the same pattern (main effect of grammatical aspect; $t=-2.21$, $p=0.032$). These results suggest that grammatical aspect drives the real-time updating of object location representations.

Discussion Our results suggest that both *grammatical aspect* (as shown by gaze data and click data) and *verb semantics* (as shown by the click data) guide the process of constructing event representations. Eye-gaze patterns show that participants use grammatical aspect to dynamically update the object location representations as the sentence unfolds. That is, the process of mapping language onto mental event representations is a dynamic, real-time process. This finding is in line with prior work by [1]. Our study further shows that a temporal-semantic grammatical cue such as grammatical aspect is a relevant cue during this dynamic process. The study also suggests that the novel webcam-based eye-tracking method can provide informative data for psycholinguistic research.

Table 1: Design (For completeness, both dative argument realizations were tested (*V the ball to GOAL & V GOAL the ball*). Because they did not pattern differently, we collapse them.

Grammatical aspect	<i>give</i> -type verb	<i>throw</i> -type verb
imperfective	Liam was giving Paige the ball.	Carly was throwing Oliver the ball.
perfective	Liam gave Paige the ball.	Carly threw Oliver the ball.

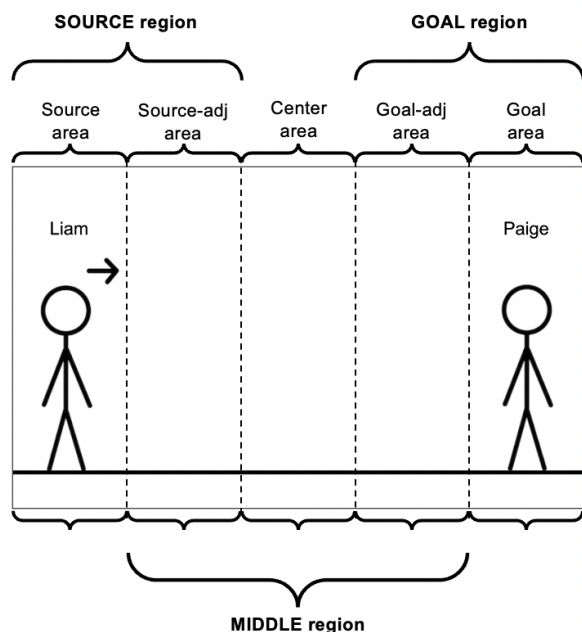


Figure 1. Visual scene

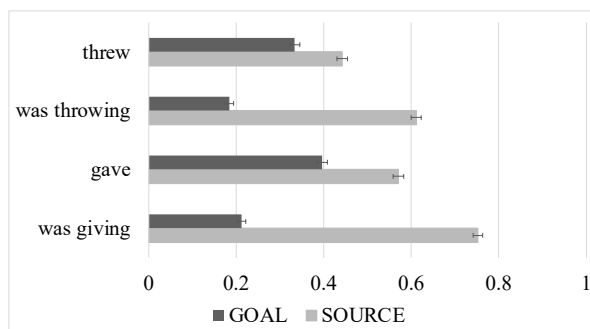


Figure 2. SOURCE vs. GOAL region clicks

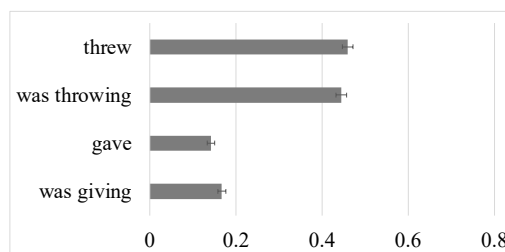


Figure 3. MIDDLE region clicks



Figure 4. Proportions of GOAL region looks by grammatical aspect (left), Proportions of SOURCE region looks by grammatical aspect (right); Center area looks not plotted; 0 on the x-axis indicates the onset of the verb; Data is collapsed by participant for plotting

References

- [1] Altmann & Kamide. 2009. *Cognition*.
- [2] Altmann & Ekves. 2019. *Psychological Review*.
- [3] Jackendoff. 1983. *Semantics and cognition*.
- [4] Rappaport Hovav & Levin. 2008. *Journal of Linguistics*.
- [5] Talmy. 1983. How language structures space. In *Spatial orientations: Theory, research, and application*.
- [6] Talmy. 1985. Lexicalization patterns: Semantic structure in lexical forms. In *Language typology and syntactic description*.
- [7] Papoutsaki et al., 2016. *Proceedings of the International Joint Conference on Artificial Intelligence*.
- [8] Slim & Hartsuiker. 2021. Visual world eyetracking using WebGazer.js.
- [9] Zehr & Schwarz. 2018. PennController for Internet Based Experiments (IBEX).