The Goal Bias Emerges Early in Motion Event Inspection and Speech Planning: Evidence from Eye-Movements

Yiran Chen (cheny39@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania,
3401 Walnut Street, Philadelphia, PA 19104 USA

John Trueswell (trueswel@psych.upenn.edu)
Department of Psychology, University of Pennsylvania,
425 S University Ave, Philadelphia, PA 19139 USA

Anna Papafragou (anna4@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania,
3401 Walnut Street, Philadelphia, PA 19104 USA

Abstract

After viewing motion events with a starting-point (Source) and end-point (Goal), people mention the Goal more often and remember it more accurately than the Source. This Goal privilege has been hypothesized to arise from an on-line attentional bias that occurs during event apprehension itself, yet no data exists that: (a) documents this online attentional bias and (b) correlates any online bias with offline memory and linguistic measures. Here we do just that: we recorded participants' eye movements as they viewed or prepared to describe motion events and later tested their memory of Goals or Sources. We find an online attentional bias for Goals over Sources during initial encoding of events. This bias is stronger during free inspection compared to speech planning, an effect likely to reflect the fact that sentence preparation partially promotes encoding and mentioning Sources. Moreover, the extent of the attentional Goal bias is systematically related to both language production and memory, such that the attentional Goal bias is greatest when the Source is not mentioned later during production or not remembered later at test. Thus, we provide the first evidence that an attentional Goal bias appears as soon as one starts to visually encode motion events.

Keywords: Event cognition, Language production, Eye-tracking, Psycholinguistics, Memory

Introduction

Components or relations within an event are not created equal: for any event, we will inevitably only take notice of, comment on and remember some aspects of it but not others. Sometimes, we form systematic biases in terms of which aspects of the event to prioritize. A prime example of such a bias is an asymmetry between the origin (Source) and the endpoint (Goal) of motion events in language production and event memory. When people see an event such as a squirrel going from a mailbox to a trash can (as depicted in Figure 1), they are more likely to mention the Goal ("to the trash can") than the Source ("from the mailbox") in their linguistic description of the event and to more accurately remember Goals than Sources (Chen, Trueswell & Papafragou, 2022; Do, Papafragou & Trueswell, 2020, 2022; Johanson, Semilis, & Papafragou 2019; Lakusta & Landau, 2005, 2012; Lakusta, Muentener, Petrillo, Mullanaphy, & Muniz, 2016; Papafragou, 2010; Regier & Zheng, 2007; Stefanowitsch & Rohlde, 2004). Even prelinguistic infants are more sensitive to Goal changes than Source changes to motion events (Lakusta & Carey, 2015; Lakusta & DiFabrizio, 2017; Lakusta, Wagner, O'Hearn, & Landau, 2007).

Commentators have attributed this privileged status of Goal in language production and event memory to an online attentional bias that occurs during event apprehension itself. For example, Regier and Zheng (2007) presented adults with pairs of joining events that sometimes differed in their end state (e.g., a hand placing a lid onto vs. into a container) and separation events that sometimes differed in their starting state (e.g., a hand taking a lid off vs. out of a container). They found that adults were better able to detect the differences in joining than separation and in turn attributed this difference to an attentional bias favoring end points of motion events (Goals) over starting points (Source). Another influential account proposed that the Goal bias could be driven by online attention to the intentions of a human actor, which is often represented by the Goal landmark. This idea receives further support from the finding that a Goal bias is robust for animate and intentional agents but does not robustly show up when the moving agent is inanimate (e.g., Lakusta & Landau, 2012; Lakusta, Reardon, Oakes & Carey, 2007).

While these proposals attributing the Goal bias to an attentional bias during event encoding are widely cited, the evidence for them is exclusively based on offline measures: either memory for the Goal vs. the Source of the event, or the likelihood of mentioning the Goal vs. the Source in linguistic descriptions. To the best of our knowledge, no study yet has used online measures to examine whether the Goal bias indeed emerges during initial event apprehension. Although it is natural to assume that higher sensitivity to change-detection in memory or higher likelihood of being encoded in language would point to higher attention during encoding, such relations are not guaranteed, especially in the context of complex and dynamic events. For example, while English speakers performed better than Greek speakers at remembering whether a motion event (e.g., a man skating) had an endpoint, the two groups' attentional patterns indicated by real-time eye-movements did not seem to differ (Papafragou, Hulbert & Trueswell, 2008). In another study, when describing motion events such as a boy roller-skating towards a soccer net, adults mentioned Path of motion ("toward the soccer net") much more often than children, the two groups' visual attention to the Path object were almost identical (Bunger, Trueswell & Papafragou, 2012).

Therefore, with offline measures alone, we are not able to pinpoint at what stage a conceptual asymmetry like the Goal bias emerges – attention during event inspection, conceptualization after inspection, or consolidation after encoding. This is particularly important since there is an alternative account for the Goal bias that does not reflect an online attentional bias but rather recency effect in memory (Regier & Zheng, 2007). According to this possibility, the Source and Goal could have received equivalent attention as people visually examine the event but the information about Source (i.e., the starting point of the event) could have just decayed faster at test since it was attended to earlier. Therefore, online measures of attention allocation during event apprehension are called for in order to determine at what stage the Goal bias emerges, thus better adjudicating between competing accounts of the nature of the bias.

The current study does just that: we examine attention patterns during the initial apprehension of motion events through eye-tracking and ask (a) whether there is indeed an attentional preference to Goal and (b) if so, whether the extent of the attentional Goal preference is directly related to the robust offline behavioral signatures of the Source-Goal asymmetry (i.e., different likelihood of being encoded in a linguistic description and different accuracy in memory tests). Specifically, we recorded participants' eye movements as they freely viewed or prepared to describe motion events similar to that in Figure 1 and later tested their memory on Goal or Source changes. This work thus bridges research in event apprehension, memory and language.

Experiment

Methods

Participants A hundred and twenty-eight native speakers of American English recruited from University of Pennsylvania subject pool participated for course credit.

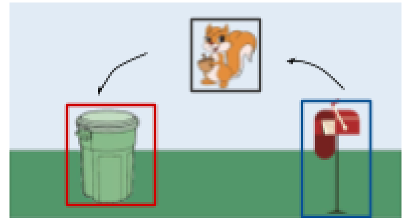

Figure 1. The midpoint of a sample critical event with Areas of Interest (Goal in Red, Source in Blue, Figure in Black). Arrows represent the motion trajectory in the original videos. Participants did not see the boxes or arrows, and saw animated videos instead.

Materials We created 28 video clips depicting motion events (each 5s). In critical events (n = 16), an animate Figure moved from an inanimate Source landmark to an inanimate Goal landmark. Filler events (n = 12) did not involve a Source/Goal path (e.g., a ghost moves around the moon). The Figure, Source and Goal were all represented by clipart images (See Figure 1a for an example of a critical clip). The motion was achieved through Powerpoint Animation and was saved as video files.

The direction of the motion in the critical clips were left-right counterbalanced such that half of our clips showed a Figure moving from left to right and the other half showed a Figure moving from right to left. We constructed two experimental lists to counterbalance Source and Goal landmarks such that objects which were the Sources in one list were the Goals in the other. We generated two additional lists with reversed presentation order to control for potential order effect.

To probe speakers' conceptual encoding of Sources and Goals in memory, we also constructed foil videos that involved either a Source Change or a Goal Change relative to the critical videos. Source and Goal changes were always within-category (e.g., the mailbox was changed to another mailbox, see comparison of Figure 1 and Figure 2a or 2b).

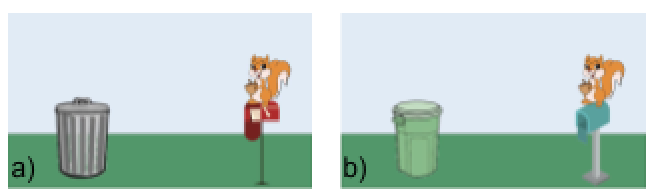

Figure 2. First-frame of a memory test trial of the critical event shown in Figure 1 involving a) a Goal change and b) a Source change.

For each critical clip, three Areas of Interest (AOIs) – Figure, Source and Goal – were manually defined a priori in Tobii lab pro studio (See Figure 1 for an example). The AOIs of the same clipart image (e.g., the squirrel) were of the same size across different counterbalancing videos and foil videos.

Procedure Participants were randomly assigned to either the Linguistic or the Non-linguistic condition and to one of the four presentation lists. First, we familiarized all participants with all the clipart images that would later appear in either the target video clips or the foil video clips in the memory task. This was done to minimize possible image familiarity at test. These pictures were presented one at a time at the center of the screen and proceeded automatically every two seconds.

Then, participants proceeded to the encoding phase, where the Linguistic and Non-linguistic condition differed. In the Linguistic condition, participants were asked to describe each video freely after the video played once and disappeared. In the Non-linguistic condition, participants were just told to watch the videos carefully.

After they encoded (linguistically or non-linguistically) all 28 videos, all participants received a memory test: on each critical trial, participants either saw foils with within-category Source changes or Goal changes[1] and were asked to say whether the video was the "same" or "different" as before after the video played once. All participants were told in advance that there would be a memory test.

Eye movements were recorded with a Tobii TX300 Pro table-top eye-tracker when participants freely inspected the videos (Nonlinguistic condition) or inspected the video in preparation for language description (Linguistic condition) and when all participants inspected the videos during memory test. Calibration was performed at the beginning of the experiment and a centered fixation cross was used prior to the start of each video clip. Verbal responses were recorded and later transcribed for analysis.

Analysis

Language Production Based on prior work (Lakusta & Landau, 2005, 2012; Papafragou, 2010; Do et al., 2020, 2022; Chen et al., 2022), we expect participants to be more likely to encode the Goal landmark than the Source landmark in their linguistic description of the critical events. To test that, we built a logistic mixed-effects model where we predicted whether the landmark object was mentioned with Role (Goal vs Source, sum coded) as a fixed effect and included by-subject and by-item random intercepts as well as by-subject and by-item random slopes of Role.

[1] In order to be able to compare one participant's looking pattern depending on Goal (Source) correctness, Change Type was manipulated between-subjects such that one participant would be asked to detect Goal changes on all trials or Source changes on all trials.

Memory for Sources and Goals Based on prior work (Lakusta & Landau, 2005, 2012; Papafragou, 2010; Do et al., 2020, 2022; Chen et al., 2022), we also expect that participants will be more likely to detect changes to the Goal landmarks than those to the Source landmarks. To test that, we built a logistic mixed-effect models predicting whether the response on a critical trial was correct with a fixed effect of Change Type (Goal vs Source, sum coded), Condition (Non-linguistic vs Linguistic, sum coded) and by-subject and by-item random intercepts and by-item random slope of Change Type.

Eye Movements First, to test whether the Goal bias emerges during event apprehension, we ask whether, in general, there is a longer fixation to Goal than Source as participants inspect our critical events. In addition, to determine whether the eye-movement patterns in the Linguistic task are linked to the specific process of language production or whether they reflect conceptual representation of motion events, we also compare eye-movement during Linguistic and Nonlinguistic encoding. With these two goals, we built a linear mixed-effect model predicting total fixation duration (ms) with AOI (Goal vs Source, sum coded), Condition and their interaction as fixed-effect and by-subject and by-item random intercepts, by-item random slopes for AOI and Condition, and by-participant random slope for AOI.

We then asked whether the extent of Goal bias interacts with language production and memory performance. To quantify the extent of the Goal look preference, we created a Goal Look Preference Index (GLPI) by subtracting the logit of Source look proportion from that of Goal in each time bin (50ms). To ask whether GLPI interacts with language production, we ran a mixed-effect linear model on participants in the Linguistic condition, predicting their mean GLPI during encoding with whether they mentioned Source (Mentioned vs Not mentioned, sum coded) as fixed effect, with by-subject and by-item random intercepts and by-subject and by-item random slopes for Source mention. Similarly, to ask whether GLPI interacts with memory, we ran a mixed-effect linear model, predicting mean GLPI during encoding with Condition, Change type, Memory Accuracy (Correct vs Incorrect, sum coded) and a three-way interaction as fixed effects and by-subject and by-item random intercepts and by-item random slopes for Memory Accuracy and Change type as random effects.

Results

Language Production We replicated the Goal bias in language observed in prior work: the participants in the Linguistic condition were more likely to mention the Goal than the Source in their descriptions of the events (M_{Goal} = 97.5%, M_{Source} = 87.3 %, β = - 1.128, SE = 0.399, p = 0.005).

Memory for Sources and Goals As expected, we also replicated the Goal bias in memory production: participants were more likely to detect Goal changes than Source changes (β = 0. 201, SE = 0.073, p = 0.006). Unsurprisingly, since they likely engaged with the events more deeply, participants in the Linguistic condition performed better on the memory test overall than participants in the Non-linguistic condition (β = 0. 248, SE = 0.073, p < 0.001; Linguistic: M_{Goal} = 62.5%, M_{Source} = 53.1%; Non-Linguistic: M_{Goal} = 49.2%, M_{Source} = 42.9%). No significant interaction was detected between Change type and Condition.

Eye Movements Participants fixated longer on the Goal than the Source when initially inspecting the critical events regardless of conditions (β = 199.41, SE = 55.04, p < 0.001, M_{Goal} = 1390 ms, M_{Source} = 960 ms, Figure 3). There was also a main effect of Condition: participants fixated on both AOIs longer in general in the Linguistic condition (β = 53.71, SE = 23.02, p = 0.021). There was also a significant interaction between Condition and AOI (β = -113.42, SE = 14.30, p < 0.001), suggesting a stronger Goal look preference in the Non-linguistic (Figure 3). The description task prompted more examination of the Source, which was otherwise even more likely to be overlooked in free inspection.

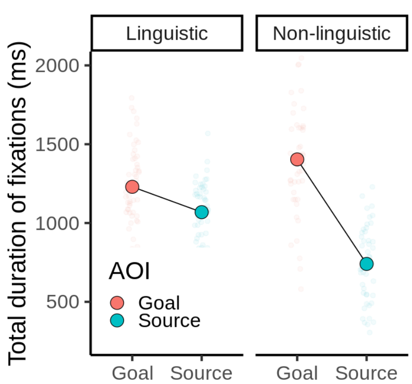

Figure 3. Mean duration of fixation on Source and Goal during initial encoding in the Linguistic and Non-linguistic condition. Translucent dots in the background represent participant means.

Beyond the overall looking preference, we turn next to how the extent of Goal look preference interacts with language production and memory. Focusing first on the Linguistic condition, we found that participants had weaker Goal look preference when preparing descriptions in which they mentioned the Source compared to when they did not (β = 0. 291, SE = 0.094, p = 0.005, Figure 4).

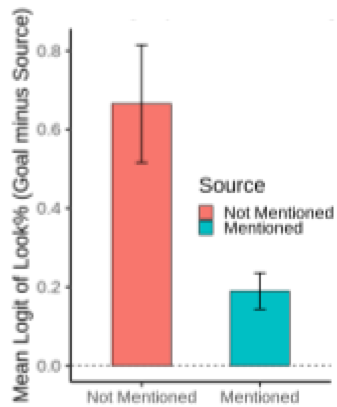

Figure 4. Goal look preference during speech planning, split by Source mention in language production. Goal look preference is represented by the difference between Goal and Source look (Elogit transformed proportion of look in every 50ms window).

By the same token, online Goal look preference during event inspection is systematically related to later memory performance. Our model revealed a main effect of Condition (β = 0. 310, SE = 0.049, p < 0.001), which is consistent with the overall more pronounced Goal preference in the Non-linguistic condition reported above. Importantly, the model also revealed a significant interaction between Change Type and Memory Accuracy (β = 0.055, SE = 0.027, p = 0.047). Combined with patterns shown in Figure 5, the interaction can be interpreted as follows: while success in detecting a Goal change was linked to a more pronounced Goal look preference, success in detecting a Source change was linked to a reduced Goal look preference.

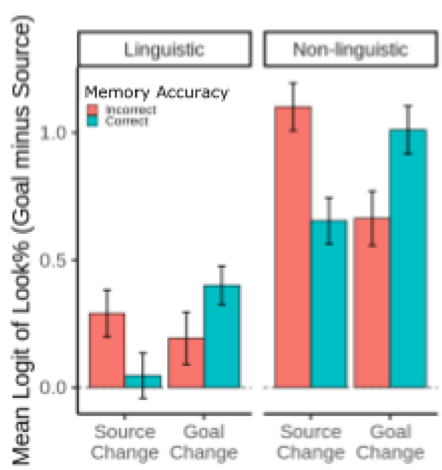

Figure 5. Overall Goal look preference (averaged across the 5s event) in the Linguistic and Non-linguistic condition by Change type (Source vs. Goal) and Memory Accuracy (Incorrect vs. Correct) in the memory test.

In addition to the planned analysis, we examined the time-course of Goal look preference as the events unfolded during initial perception. We found that the differential looking pattern by Memory Accuracy emerged globally rather than being driven by differences located during specific moments in time (Figure 6).

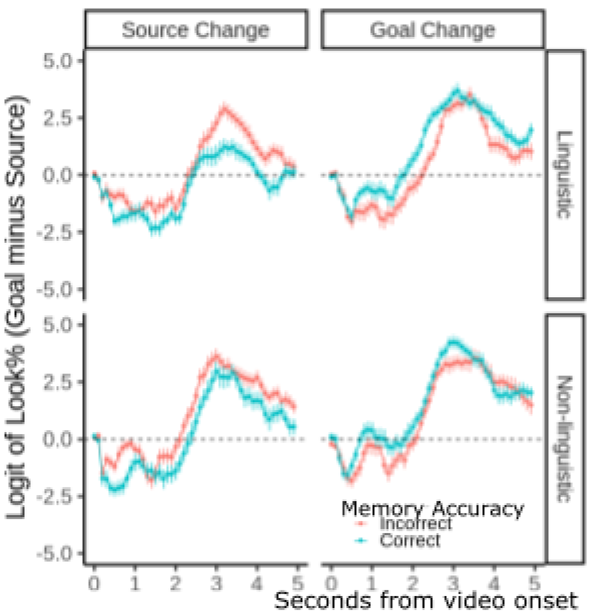

Figure 6. Goal look preference over the course of the entire event in the Linguistic and Non-linguistic condition, split by the Change type and accuracy in the memory test.

General Discussion

Our eye-tracking experiment contributes the first evidence that the Goal bias reflected in language production and memory is present at initial online event apprehension. For the same landmark object, participants fixated on it longer if it assumed the role of Goal in a motion event compared to when it assumed the role of Source during both free inspection and inspection for speech planning. This bias was more pronounced in free inspection and was reduced if participants were tasked with speech preparation. This shows that participants visually approach the same events in different ways in different contexts (i.e., free inspection vs. planning a description). This in turn highlights the fact that the Goal bias has a cognitive basis independent from linguistic processes. Further, the extent of such a Goal bias in visual attention was found to be systematically related to the result of language production as well as memory performance: attention to the Goal over the Source is greatest when the Source is not mentioned during production and/or not remembered later at test. Below, we further discuss the implications of these findings for the nature of the Goal bias and the broader relation between visual attention, speech production and event memory.

A critical aspect of our findings is that an attentional bias that emerges during inspection is responsible for the more accurate encoding of the Goal and its later memory advantage compared to the Source. What accounts for this early attentional bias towards the Goal? It is possible that bottom-up sources of information drive this effect, as it has been shown that viewers often launch predictive eye-movement to project the trajectory of a motion path (Michael & Melvill Jones, 1966; Stark, Vossius, & Young, 1962), which in our materials was the Goal landmark. However, given that there are large effects of top-down control on eye movements (e.g., Land & Tatler 2009) and that the gist of an event and its roles can be extracted within milliseconds of observing it (Hafri, Papafragou & Trueswell, 2013; Hafri, Trueswell & Strickland 2018), it is also entirely possible that early eye movements reflect rapid apprehension of the event structure - including the assessment of the intention of the animate character.

Our finding that Source mention was directly linked to reduced visual attention to Goal during speech planning resonates with and extends previous psycholinguistic findings showing a tight relation between visual attention allocation during speech planning and language production. Although previous research has robustly established that speakers in general direct their attention to event components that they plan to talk about and in the order in which they plan to talk about them (e.g., Griffin & Bock, 2000; Bock et al., 2004; Gleitman et al., 2007), these investigations were often restricted to static representation displays (often only with an agent and a patient). Though other studies have looked at more complex events and the relation between attention and mention of these event components (Papafragou et al., 2008; Bunger et al., 2012, 2016), attentional patterns were not typically analyzed depending on which event component was encoded in the utterance produced (but see Bunger et al., 2021). Therefore, the current study contributes some of the first evidence that participants' relative attention to different components of an event during utterance planning is directly related to whether they later mention these components.

In future work, we hope to engage in a finer-grained analysis of how preparing different event descriptions of the same dynamic event changes the way that the event was visually inspected. Currently, our linguistic analysis remained at the level of whether each event component was mentioned or not. However, many rich details about *how* an event component was manifested in the utterance were not differentiated. For example, while the majority of participants produced a linguistic description reflecting a typical Source-Goal construal (e.g., "The squirrel moved from the mailbox to the trash can"), some construed the Source landmark as a modifier of the Figure ("The fairy in her house flew to the tree") and others set the stage with the

Source landmark but arguably did not include it as part of the motion event ("There was a windmill on the left. The dolphin was swimming towards the beach that's on the right".) Future analyses will examine whether and how looking patterns differ by the type of encoding and order of mention to clarify how early decisions about linguistic encoding of event roles are made during the process of speech planning.

Lastly, the current work encourages the use of online measures to understand the representation of dynamic events. As discussed in the introduction, commentators have discussed whether the Goal bias in language production and memory is driven by a memory recency effect (Regier & Zheng, 2007) or the rapid apprehension of intentional structure (e.g., Lakusta & Landau, 2012). Our investigation with online measures seems to rule out the recency effect as the sole explanation for the Goal bias and suggests that observers build real-time event representations that include an assessment of intentionality. Further investigation using online measures is needed to directly examine this possibility. For instance, future work can examine participants' visual attention during inspection of similar motion events but with an inanimate Figure to see whether the preference for Goal over Source in online visual attention is reduced or even entirely eliminated. Future studies can also ask whether manipulations that implicitly direct attention to Source in turn modulate event description and event memory.

Acknowledgments

This research was supported by the graduate research grant issued by the Language and Communication Sciences program under the MindCORE initiative at the University of Pennsylvania. We thank members of the Language and Cognition Lab and the Language Learning Lab at the University of Pennsylvania for helpful feedback.

References

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In: *The interface of language, vision, and action: Eye movements and the visual world*, 249-278.

Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension and utterance formulation in children: Evidence from linguistic omissions. *Cognition*, 122(2), 135-149.

Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: implications for language production and attention. *Language, Cognition and Neuroscience*, 31(8), 1015-1037.

Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2021). How children attend to events before speaking: crosslinguistic evidence from the motion domain. *Glossa: a journal of general linguistics*, 6(1).

Chen, Y., Trueswell, J., & Papafragou, A. (2022). The Source-Goal Asymmetry in Motion Events: Sources Are Robustly Encoded in Memory but Overlooked at Test. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

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

Do, M. L., Papafragou, A., & Trueswell, J. (2022). Encoding motion events during language production: effects of audience design and conceptual salience. *Cognitive Science*, 46(1), e13077.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4), 544-569.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. Psychological science, 11(4), 274-279.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, 142(3), 880.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, 175, 36-52.

Johanson, M., Selimis, S., & Papafragou, A. (2019). The Source-Goal asymmetry in spatial language: Language general vs. language-specific aspects. *Language, Cognition and Neuroscience.*

Lakusta, L., & Carey, S. (2015). Twelve-month-old infants' encoding of goal and source paths in 'agentive' and 'nonagentive' motion events. *Language Learning and Development,* 11, 152–175.

Lakusta, L., & DiFrabrizio, S. (2017). And, the Winner Is…A Visual Preference for Endpoints over Starting Points in Infants' Motion Event Representations. *Infancy*, 23, 323–343.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, 96, 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, 517–544.

Lakusta, L., Muentener, P., Petrillo, L., Mullanaphy, N., & Muniz, L. (2016). Does Making Something Move Matter? Representations of Goals and Sources in Motion Events With Causal Sources. *Cognitive Science*, 41, 1–13.

Lakusta, L., Spinelli, D., & Garcia, K. (2017). The relationship between pre-verbal event representations and semantic structures: The case of goal and source paths. *Cognition*, 164, 174-187.

Lakusta, L., Wagner, L., O'Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language Learning and Development*, 3, 179–197.

Land, M., & Tatler, B. (2009). Looking and acting: vision and eye movements in natural behaviour. Oxford University Press.

Michael, J. A., & Jones, G. M. (1966). Dependence of visual tracking capability upon stimulus predictability. Vision research, 6(11-12), 707-IN20.

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

Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, 34, 1064–1092.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism.* Cambridge, MA: MIT Press.

Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science*, 31, 705–719.

Rohde, H., Kehler, A. and Elman, J. (2006). Event Structure and Discourse Coherence Biases in Pronoun Interpretation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 28, 697-702.

Stark, L., Vossius, G., & Young, L. R. (1962). Predictive control of eye tracking movements. *IRE Transactions on Human Factors in Electronics*, (2), 52-57.

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

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