

Action and actor gaze mismatch effects during spoken sentence processing

Dato Abashidze (dato.abashidze@hu-berlin.de)

CITEC, Department of Linguistics, Bielefeld University, Germany
Psycholinguistics, Humboldt University of Berlin, Germany

Pia Knoeferle (pia.knoeferle@hu-berlin.de)

Psycholinguistics, Humboldt University of Berlin, Germany

Abstract

Eye tracking research on situated language comprehension has shown that participants rely more on a recent event than on a plausible future event during spoken sentence comprehension. When people saw a recent action event and then they listened to a German (NP1-Verb-Adv-NP2) past or futuristic present tense sentence, they preferentially looked at the recent event target over another plausible target object (that might be involved in a future action) independent of tense. This preferential inspection persisted even when future events and futuristic present sentences were much more frequent within the experiment, or when a gaze cue biased towards the future action target. The present experiments extend this line of research by introducing incongruence (in Experiment 1 a past tense verb mismatched the recently seen action and in Experiment 2 an actor gaze cue mismatched the past tense sentence condition). Can the verb-action and the gaze-sentence mismatches eliminate the recent-event inspection preference? Would participants recall information in post-experimental memory tests better for matches (the futuristic present tense condition) than mismatches (the past tense condition)? Results revealed inspection of the recent event target as participants processed the verb-action mismatch (Exp 1) and actor gaze incongruence (Exp 2). However, the gaze (but not the verb-action) incongruence eliminated the *overall* recent event preference in the NP2 region. The memory tests also showed some evidence for a reversal of the recent-event preference.

Keywords: Eye-tracking; spoken sentence comprehension; visual world paradigm; recent-event preference; event-sentence incongruence; actor gaze mismatch

Introduction

Every day people see or hear about events in the world and effortlessly integrate language with what they see. Although previous research has examined how people understand language referring to events, little is known about how we interpret reference to a preceding event context in relation to language about future events. Previous research has revealed that both visual and linguistic context can rapidly guide the listeners' visual attention (e.g., Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) and expectations about events. For example, cues to event tense provided by the utterance can help comprehenders in developing expectations about future events (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003). At the same time, listeners tend to prioritize recently inspected event depictions (Knoeferle & Crocker, 2007), or real-world event portrayals (over expectations of future events) when both recent and future events could, temporarily, relate to an utterance (Abashidze, Knoeferle, Carminati, & Essig, 2011; Knoeferle, Carminati, Abashidze, & Essig, 2011). We present two eye-tracking ex-

periments that are situated in the context of these extant findings. The present experiments examine to which extent the priority accorded to recent (vs. future) events in interpreting an utterance holds up when we weaken the congruence between the recent event and the unfolding utterance. To this end, we created mismatches between the recent event context (the action) and (the verb in) an unfolding utterance (Experiment 1) / an actor's gaze behavior (Experiment 2).

The robustness of the recent-event preference

A number of studies have provided evidence for an attentional behaviour that has been dubbed the 'recent event preference' (e.g., Abashidze et al., 2011; Knoeferle & Crocker, 2007): In Experiment 2 by Abashidze et al. (2011) participants saw a person performing an action (e.g., sugaring strawberries) and then they listened to either a past tense sentence (*Der Versuchsleiter zuckerte kürzlich die Erdbeeren*, 'The experimenter recently sugared the strawberries') or a futuristic present tense sentence (*Der Versuchsleiter zuckert demnächst die Pfannkuchen*, 'The experimenter will soon sugar the pancakes'). During the sentence they saw the person in a static position and two objects on the table in front of him (e.g., pancakes and strawberries, i.e., Fig 1-B). After the sentence presentation a second event showed again a sugaring action (the 'future' action) but this time on the other object (e.g., sugaring pancakes). While participants listened to the sentence, their eye gaze to the two potential targets (of the recently seen action, and of a potential future action) were monitored. Results showed that participants preferentially inspected the recent event target (i.e., the strawberries) over the other plausible future event target. This happened even during the futuristic present tense sentence, and they shifted gaze to the pancakes (the plausible future event target) only as it was mentioned. Follow-up studies examined this issue by increasing the number of the future events and of futuristic present tense sentences up to 88% (Abashidze, Carminati, & Knoeferle, 2014), by having the actor gaze at the targets before their mention (Abashidze, Knoeferle, & Carminati, 2015), and by moving linguistic cues that could bias against the recent-event preference to the sentence beginning (Abashidze & Chambers, 2016). Despite these strong visual and linguistic cues in favor of the future event target, these experiments replicated the *overall* recent event preference.

The impact of incongruence and gaze cues

In language processing research, many studies have employed picture-sentence incongruence and verification as a

method. Experiments using this method have found that participants are sensitive to the incongruence, so that they responded faster to congruent than incongruent picture-sentence pairs (Carpenter & Just, 1975; Underwood, Jebbett, & Roberts, 2004). In an eye-tracking study by Knoeferle and Crocker (2005) participants were presented with depicted scenes and either matching or mismatching np1-verb-adv-np2 sentences. The authors found an incongruence effect in the verb and adverb regions. Participants were faster reading these sentence regions in the congruous than incongruous condition (see also related findings on gender stereotype effects in a picture-sentence verification task by Rodríguez, Burigo, & Knoeferle, 2015). Further studies used the sentence-picture verification procedure while participants were presented with positive and negated sentences (Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999). Results showed that pictures matching the presented sentences (even when the sentences were negated) elicited faster responses than pictures mismatching the sentences. In addition, another cue that has been shown to rapidly influence comprehension (and guide participants' visual attention even when it was incongruous with language) is a speaker's gaze (e.g., Hanna & Brennan, 2007; Kreysa & Knoeferle, 2013; Staudte, Crocker, Heloir, & Kipp, 2014).

The present experiments relied on incongruence in verb-action relations and an actor's gaze (to the future target) as a way to stress-test listeners' preference of inspecting the target of a recent action. The causes underlying the preferential inspection of the recent event are unclear. Perhaps the preferential inspection is guided by the verb. The verb could be linked to representations of the recently inspected action and its location, prompting participants to shift gaze to the location of the action when they encounter the verb. If so, then a match between the recent event and the sentence referring to it could boost the attention towards the recent event. By contrast, a mismatch between the visual and linguistic information could reduce the recent event preference. Alternatively, what we see is a general recency effect (i.e., participants inspect the object that is the target of the recent action, independent of verb meaning). If this were the case, then a mismatch between the recent action and the verb should not interfere with the recent-event preference but a gaze cue (e.g., the actor shifting gaze during the verb to the future target object) might diminish a recency effect and direct the listener's attention to the future target object.

The present experiments

Given that incongruence has been shown to influence participant' eye-movements and their reaction times during picture-sentence verification tasks, two eye-tracking studies examined to which extent incongruence could bias against the recent-event preference. In Experiment 1, the verb of the past (but not futuristic present) tense sentences mismatched the recent action. In Experiment 2, the actor began to inspect the future target object at verb onset in the past tense sentences. We tested to what extent these incongruences will reduce the

preferential inspection of the recent event target during the verb. At this point people could realize that the verb does not match the action they saw in Exp 1 and notice that the actor shifts his gaze towards the future target from the verb onset while the past tense sentence refers to the recent event target in Exp 2. In both experiments the experimental trials were incongruent in the past tense condition only. These experiments used the design from Abashidze et al. (2011, Exp 2), presented above, with one factor, viz. *tense* (past vs. futuristic present). If the recent event preference is sensitive to the incongruence between the recent event and either the past tense sentence (Exp 1) or the actor's gaze (Exp 2), then we should see a decrease in looks to the recent event target starting from the verb region. Can the congruence in the futuristic present tense increase the inspections towards the future event target and override the preferential inspection of the recent event? The incongruence in the past tense might strengthen the congruence in the futuristic present tense condition (e.g., Glenberg et al., 1999).

After the eye-tracking session, participants took part in a gated-memory (Exp 1) and a memory (Exp 2) test. Previous studies reported a better recall of the future event (Abashidze et al., 2015, Exp 1) which was not in agreement with the gaze data; however, other findings revealed a better recall of the recent events (Abashidze et al., 2014, Exp 1 and 2), and a better recall of the past tense sentence (Abashidze et al., 2015, Exp 2) underscoring the recent-event preference in the gaze data. If the incongruence affects the recent-event preference and the incongruence effects are long-lasting, then we might see a reduced recall performance for recent compared with future events in the memory test. Alternatively, the incongruence does not affect the recent-event preference and / or its effects are short-lived, in which case we might see better recall of recent than future events.

Experiments 1 and 2: Methods

Participants

Thirty-two native German University students in each experiment (aged 18 to 32) with normal or corrected-to-normal vision gave informed consent and received 6 Euros each for their participation. The study was approved by an ethics vote (Experiment 1: Bielefeld University ethics committee, Experiment 2: DGfS).

Materials and design

The current experiments used the experimental sentences from Abashidze et al. (2014, see Table 1). All sentences ($N=24$) had the structure NP-VERB-ADV-NP and two native German speakers recorded them. The sentences were in two tense conditions and referred either to a recently seen event or a plausible future event. In one condition, the verb was in the present tense and a time adverb (*dennächst*, 'soon') indicated the futuristic present tense condition (Table 1b). In the other condition, the verb was in the simple past, and a time adverb (*kürzlich*, 'recently') indicated the past tense condi-

tion (Table 1a). The critical sentences employed only regular German verbs in which the verb was tense ambiguous up to but excluding the word-final phoneme which disambiguated towards the simple past in the past tense condition. As we can see in Table 1, the experiments used two sentences for each tense condition. With this counterbalancing we ensured that each object was once the target of the recent and once the target of the future event. The critical words in a sentence were matched for spoken syllables and lemma frequency within an item (Baayen, Piepenbrock, & Gulikers, 1995).

Experiment 1 also used the videos from Abashidze et al. (2014) for futuric present tense sentences. These videos ($M_{\text{duration}}=5015$) showed an actor sitting at a table in front of two objects (e.g., strawberries and pancakes, one on the left and one on the right; both of the objects can be sweetened). We additionally recorded new videos (for the recent events). For example, instead of sugaring strawberries, a first video presented the actor tasting the strawberries (see Fig 1-A), and the verb in the past tense sentence (translated literally: ‘The experimenter sugared recently the strawberries’) never matched the recent action. By contrast, the futuric present sentence always matched the future event (see Fig 1-C).

In Experiment 2, for each experimental trial, participants saw a short video before and after hearing a sentence about a person performing an action. For the incongruence, we used the gaze videos from Experiment 2 by Abashidze et al. (2015). For instance, when participants listened to a past tense sentence they saw a video of the actor shifting his gaze towards the future event target (i.e., pancakes) from the onset of the verb, where it remained until the end of the sentence (the gaze cue mismatched the referential past tense sentence). By contrast, in the futuric present tense condition no gaze cue was present. Participants saw a snapshot from the last frame of the first video showing the actor in a static position looking straight ahead (i.e., see Fig 1-B). In both experiments the incongruence biased against the recent event preference.

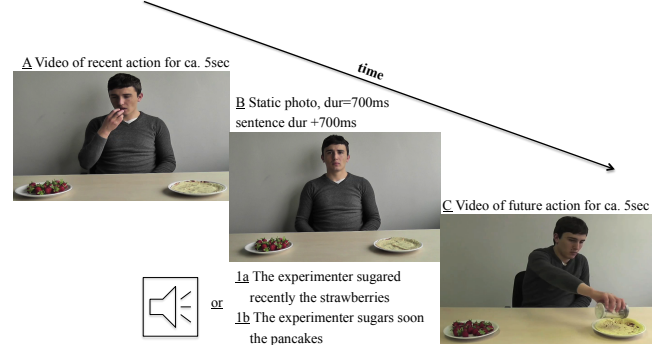
In addition to the experimental items we created 36 filler sentences. To balance the incongruence across/within experiments, 12 fillers in Experiment 1 featured an incongruence between the futuric present verb and future event. Similarly, 12 fillers in Experiment 2 showed the actor looking at the recent event target during a futuric present sentence. In both experiments, recent and future events appeared equally often.

Thus, both experiments manipulated 1 factor: sentence tense (past vs futuric present); in half of the trials the sentence was in the past tense and in the other half in the futuric present tense (see Table 1 for counterbalancing). The resulting four lists used a Latin square design. Each experimental list contained every critical item in only one condition and all fillers. Each subject saw an individually pseudo-randomized version of one of the four experimental lists.

Procedure

An Eyelink 1000 eye-tracker recorded participants’ eye movements. After a successful 9-point calibration, the experiment began. Participants were asked to inspect the scene

Figure 1: Sequence of events of a typical experimental trial for Experiment 1



and to listen carefully to the sentences. As in the previous studies by Abashidze et al. (2011), on a given trial, a participant first saw a video of a person (the actor) performing one action before the sentence (e.g., tasting strawberries for Exp 1 and sugaring strawberries for Exp 2); then participants saw a static photo (see Fig 1-B). 700 ms after the onset of the static photo, a sentence was presented via the loud speakers either in (a) the past tense or (b) the futuric present tense (see Table 1). In Experiment 1 (experimental items), the past tense verb did not match the recent event; but the verb of the futuric present matched the future event (shown after the sentence had ended). In Experiment 2, during the past tense sentence the actor directed his gaze towards the future event target from verb onset (the gaze cue mismatched the past tense sentence and its NP2 referent). However, during the futuric

Table 1: Example experimental sentences. The indices (‘) indicate counterbalancing versions

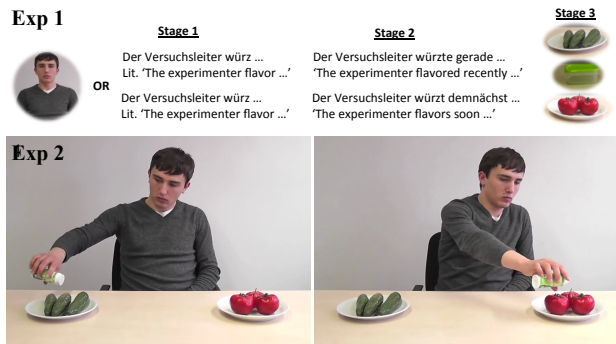
Tense condition & counterbalancing	Sentences
1a past tense	<i>Der Versuchsleiter zuckerte kürzlich die Erdbeeren</i> ‘The experimenter recently sugared the strawberries’
1a’ past tense	<i>Der Versuchsleiter zuckerte kürzlich die Pfannkuchen</i> ‘The experimenter recently sugared the pancakes’
1b futuric present	<i>Der Versuchsleiter zuckert demnächst die Pfannkuchen</i> ‘The experimenter will soon sugar the pancakes’
1b’ futuric present	<i>Der Versuchsleiter zuckert demnächst die Erdbeeren</i> ‘The experimenter will soon sugar the strawberries’

present sentence participants saw the actor in a static position throughout the sentence (as in Experiment 1, see Fig 1-B). 700 ms after the sentence had ended, participants saw a video of the actor performing the second action event (e.g., sugaring pancakes, Fig 1-C, both experiments). Post-experiment, participants completed a gated memory test in Experiment 1 and a memory test in Experiment 2 (Fig. 2). At the end, they were debriefed. Each experiment lasted approximately 50-55 minutes.

Memory tests

Experiment 1 tested participants' later memory of the linguistic information and Experiment 2 examined participants' later memory of the visual information. Experiment 1 shows an example sentence as presented in the gated memory test in a 3-stage procedure (Fig 2, Exp 1). At the first stage, participants saw only the first noun phrase and the verb stem and had to verbally complete the verb tense. The second stage added the temporal adverb, and they had to recall the second noun phrase. If they were unable to do so, they received a further prompt at the third stage and had to select the correct referent out of three objects. Two of these were from that sentence trial and the third was a distractor from another filler item.

Figure 2: An example of a sequence of stages in the gated memory test, Exp 1 and display for the memory test, Exp 2



For the memory test in Experiment 2, we created two snapshots of the first and second video of each experimental item, i.e., showing the experimenter performing one of the two actions (Fig 2, Exp 2). The two snapshots associated with each item were combined into one display and shown to participants. Two versions were created in which the respective location of the two pictures was counterbalanced and participants responded with a button press. Above the picture, one of two questions appeared:

- (a) Welche Aktion wurde VOR dem Satz durchgeführt?
"Which action was performed before the sentence?"
- (b) Welche Aktion wurde NACH dem Satz durchgeführt?
"Which action was performed after the sentence?"

Experiments 1 and 2: Analyses and results

Eye tracking

For the eye-tracking data we divided each experimental sentence into three time regions, (the verb, the adverb and the NP2). Each word region lasted from its onset to the onset of the following word region and NP2 ended at sentence offset. The measure of interest was inspection of the recent and future target objects. Because looks to one of the objects implied fewer looks to the other target objects, we computed mean log gaze probability ratios for the recent relative to the future target ($\ln(P(\text{recent target})/P(\text{future target}))$). A score of zero indicates that both targets are inspected equally often; a positive value means more looks to the recent event target; a negative value means more looks go to the future event target (see Knoeferle et al., 2011).

For the inferential analyses, we performed separate ANOVAs on the mean log ratio averaged for each condition (past vs. futuric present) and word region by participants and by items respectively. The independent variable was tense, with two levels, past and future tense. We tested the significance of the intercept *overall* (a positive intercept represents a preference of inspection of the recent event target).

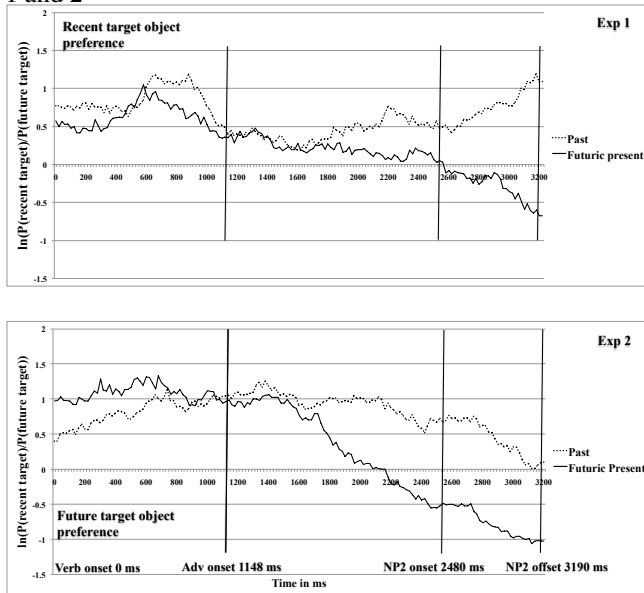
In Figure 3, the dotted lines indicate the past tense condition and the solid lines indicate the futuric present tense condition. As we can see, the incongruence influenced target inspection during sentence comprehension. In Experiment 1, participants decreased their attention towards the recent event target at the end of the verb region (i.e., following the mismatch); however both lines (in the mismatching past and matching futuric present tense) remain above zero, meaning that people continued to preferentially inspect the recent (vs. future) event target. In Experiment 2, the preferential looks towards the recent event target lasted until the middle of the adverb region in both tense conditions. Interestingly, despite the incongruent gaze in the past tense condition, participants' attention towards the recent event target persisted until sentence end but decreased as the target was mentioned.

A noticeable difference between Experiment 2 and Experiment 1 is thus that the gaze incongruence (but not the verb-action incongruence) seems to have prompted participants to decrease their attention to the recent target object during NP2. At the end of the NP2 (the name of the recent event target), participants inspected the recent and future event targets equally often in the past tense condition (unlike in Experiment 1).

ANOVAs revealed a tense effect in the NP2 region in Experiment 1 and in the Adverb and NP2 regions in Experiment 2, reflecting that tense modulated the listeners' looks to the recent (vs. future) event target. The grand mean (i.e., the mean of both conditions / the intercept) was positive in all regions in Experiment 1 and in the Verb and Adverb regions in Experiment 2, which indicates an *overall* recent-event preference (significant intercept in all the ANOVAs by region). Thus, Experiment 1 replicated the *overall* preference to look at the recent event target in all three word regions indepen-

Discussion

Figure 3: Mean log gaze probability ratios ($\ln(P(\text{recent target})/P(\text{future target}))$) by condition from verb onset for Exps 1 and 2



dent of tense condition, whereas Experiment 2 did not reveal a significant intercept in the last word region (NP2), suggesting the gaze incongruence eliminated the *overall* recent event preference in the NP2 region. Furthermore, a post-hoc two-way ANOVA on the mean log gaze ratios of the NP2 region revealed a tense effect and a marginal experiment effect but no interaction. This experiment effect suggests that the gaze incongruence not only eliminated the overall recent event preference but also marginally decreased participants' attention towards the recent event target in Experiment 2 compared with Experiment 1.

Memory tests

Gated memory test We calculated the percentage of correct responses by conditions. Participants correctly answered questions from all three stages on average with 61%. They correctly recalled 54% at stage one, which is more accurate than at stage two with 40%. Importantly, the highest accuracy emerged at stage three (89%). Subjects recalled the futuristic present sentence (match) better than the past tense sentence (mismatch). Logistic linear mixed effect (LME) analyses showed a marginal tense effect at stage 1 ($p < .08$) and a fully significant tense effect at stage 2 ($p < .02$), indicating higher accuracy for the futuristic present than past tense sentence condition. **Memory test** In Experiment 2 we calculated the percentage of correct answers in indicating the event targeted in the image. Participants correctly answered questions on average with 61%. They were slightly more accurate in recognizing the future (matching) events (60%) than the recent (mismatching) events (59%). The LME analyses did not reveal any significant difference in recalling the recent versus future events.

Across two experiments, we examined the recent-event preference (e.g., Abashidze et al., 2014; Knoeferle et al., 2011) and stress-tested it with two types of incongruence (verb-action and the actor's gaze to the future target in the past tense condition).

We had predicted that if the recent event preference were guided by the verb, then participants' overall preference to inspect the recent event target should disappear when they realize that the recent action mismatched the verb in meaning (this did not happen, Exp 1). Analyses of the data from both eye-tracking experiments did not show an early preferential inspection (during the verb and adverb) of the future event target but rather corroborated participants' preference to gaze at the recent event target. It is possible that the effect of the verb-action incongruence was weak since the past tense sentence, while mismatching at the verb, did mention the correct event target. Perhaps for this reason, inspection of the recent event target persisted during the Adverb and increased during NP2 in the past tense sentences of Experiment 1.

Furthermore, if a recency effect underlies the recent-event inspection preference and more generally object inspection, then an incongruent actor's gaze to the future target (as the most recent cue) should have guided the listeners' attention to that target during the verb and adverb for the past tense sentences, which did not happen; rather participants preferentially inspected the target of the past event i.e., in line with the sentence tense and the recent-event preference.

While previous studies revealed an immediate gaze effect in a congruent environment (at around 300-500 ms after its onset, e.g., Kreysa & Knoeferle, 2013), participants in Experiment 2 fully decreased their inspection of the not-gazed-at recent event target only at sentence end (i.e., after mention of the target), thus ignoring that the actor gazed at the future target from verb onset. This suggests a strong reliance on the recently-seen event and a relatively slow effect of the actor's gaze when it had to compete with the preceding action referenced by the verb. Although the gaze mismatch eliminated the overall recent event preference in the NP2 region of Experiment 2, a between-experiment comparison of the same word region did not reveal a fully significant experiment effect between Experiment 1 and 2. The two types of incongruence hence did not differ reliably in the extent to which they disrupted the recent event preference at sentence end. For the Adverb region, however, between-experiment analyses clarified that participants were more likely to inspect the recent event target in Experiment 2 than 1, corroborating that gaze did not immediately modulate this inspection preference. Tense effects in the Adverb and NP2 regions in Experiment 2 (compared with only the NP2 region in Experiment 1) revealed that the gaze incongruence (and the actor's attention to the future event target) did boost the integration of tense after the verb in Experiment 2 compared with 1.

The post-experiment memory tests results in Experiment 1 did not agree with the overall recent event preference in the

gaze data (recall was reliably better for the futuristic present tense sentences, in conflict with the recent event inspection preference (see also Abashidze et al., 2015)). In Experiment 2, no reliable difference in recall emerged for the recent versus future events, suggesting short-lived effects of the gaze mismatches. The better recall of the futuristic present condition in Experiment 1 could be explained if we assume that the congruent recent events and linguistic information evoked more in-depth processing and increased attention to the stimuli that then also benefitted the later recall of event information.

In conclusion, the incongruences in the past tense sentences did not reduce the overall recent event preference immediately (during the verb and adverb in both experiments); but at least actor gaze incongruence did eliminate the overall preference eventually, in the NP2 region of Experiment 2 and it boosted the tense effects. What these results suggest is that the recent-event inspection preference is robust, and that it is not entirely dependent upon verb reference or cue recency. The recall accuracy in Experiment 1, by contrast, suggests that the verb-action mismatches affected short-term memory of the events. Gaze mismatches, by contrast, seem to have had immediate effects in the sense that they boosted tense effects but they neither reduced the overall inspection preference more than verb-action mismatches at NP2, nor did they modulate recall of the stimuli.

Acknowledgments

This research was funded by the Cognitive Interaction Technology Excellence Center 277 (German Research Foundation, DFG). We thank Johanna Bokelmann, David Nothdurft and Luise Henneberg for help in conducting Experiment 2.

References

- Abashidze, D., Carminati, M. N., & Knoeferle, P. (2014). How robust is the recent event preference? In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 92–97). Cognitive Science Society.
- Abashidze, D., & Chambers, C. G. (2016). The role of early linguistic cues in the recent event preference. In *Proceedings of the Conference on Architectures and Mechanisms for Language Processing*.
- Abashidze, D., Knoeferle, P., & Carminati, M. N. (2015). Eye-tracking situated language comprehension: Immediate actor gaze versus recent action events. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 31–36). Cognitive Science Society.
- Abashidze, D., Knoeferle, P., Carminati, M. N., & Essig, K. (2011). The role of recent real-world versus future events in the comprehension of referentially ambiguous sentences: Evidence from eye tracking. In *Proceedings of the EuroCogSci*. New Bulgarian University Press.
- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The celex lexical database philadelphia: University of Pennsylvania. *Linguistic Data Consortium*.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.
- Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, *1*(1), 19–33.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.
- Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55.
- Knoeferle, P., Carminati, M. N., Abashidze, D., & Essig, K. (2011). Preferential inspection of recent real-world events over future events: Evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology*, *2*.
- Knoeferle, P., & Crocker, M. (2005). Incremental effects of mismatch during picture-sentence integration: Evidence from eye-tracking. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1166–1171).
- Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye-movements. *Journal of Memory and Language*, *75*, 519–543.
- Kreysa, H., & Knoeferle, P. (2013). Reference-related speaker gaze as a cue in online sentence processing. In D. Kluge A. & Söffker (Ed.), *Mensch Teams Systeme und Automaten*. Duisburg: DuEPublico.de.
- Rodríguez, A., Burigo, M., & Knoeferle, P. (2015). Visual gender cues elicit agent expectations: different mismatches in situated language comprehension. In *Proceedings of the Euro Asian Pacific CogSci* (pp. 234–239).
- Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, *133*(1), 317–328.
- 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*, 632–634.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology Section A*, *57*(1), 165–182.