

# Effects of speaker gaze on spoken language comprehension: Task matters

Helene Kreysa (hkreysa@cit-ec.uni-bielefeld.de)

Pia Knoeferle (knoeferl@cit-ec.uni-bielefeld.de)

Cognitive Interaction Technology Excellence Cluster  
Bielefeld University, 33615 Bielefeld, Germany

## Abstract

Listeners can use speakers' gaze to anticipate upcoming referents. We examined whether this listener benefit is affected by different comprehension subtasks. A video-taped speaker referred to depicted characters, using either a subject-verb-object or a non-canonical object-verb-subject German sentence. She shifted gaze once from the pre-verbal to the post-verbal referent, a behavior that could allow listeners to anticipate which character would be mentioned next. We recorded participants' eye movements to the characters during comprehension, as well as post-sentence verification times on whether a subsequent schematic depiction correctly highlighted the patient (Experiment 1) or the thematic role relations of the sentence (Experiment 2). Sentence structure affected response times only when verifying thematic roles. The eye movement data also showed reliable differences between tasks, regarding effects of sentence structure and their modulation by speaker gaze. We argue that processing accounts of situated comprehension must consider task effects on the allocation of visual attention.

**Keywords:** spoken sentence comprehension; task effects; speaker gaze; syntactic structuring; eye tracking

## Attention modulation across tasks

When interacting with the immediate visual environment, we can pay attention to all sorts of things: people around us, what somebody says, signs that tell us what to do or where important information is located. Intuitively, these cues affect our visual attention in diverse tasks: while we drive, as we prepare dinner, while we map-read our way to the city sights. Indeed, low-level visual cues have been shown to guide a perceiver's visual attention and improve performance across a range of cognitive tasks. In problem solving, pulsing lines can lead participants to focus on important areas in a diagram, facilitating the solution of insight problems (Grant & Spivey, 2003). In language production, arrows pointing to referents affected the produced sentence structure (Tomlin, 1995), as did brief screen flashes (Gleitman, January, Nappa, & Trueswell, 2007). In a change detection task, participants were faster to detect changes to objects when these were located in the direction of someone else's gaze than when they weren't (Langton, O'Donnell, Riby, & Ballantyne, 2006). In fact, eye gaze stimuli are known to exert a strong pressure to shift attention in the direction of the gaze (e.g., Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002).

In addition to these low-level and largely static cues, dynamic changes in visual context also affect the perceiver's attention and behavior. One such cue is the shifting focus of another person's gaze. Speaker gaze<sup>1</sup> can be informative

<sup>1</sup>We use 'speaker gaze' in a wide sense, as a cue to the direction of attention. In many cases, this explicitly includes head movements.

to listeners, because speakers discussing entities in the visual world robustly gaze at the objects they are about to mention (Griffin & Bock, 2000). Several studies have examined gaze effects by overlaying a moving cursor on a display, thus representing the speaker's gaze to objects without actually depicting the speaker (e.g., Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Carletta et al., 2010; Kreysa, 2009). Listeners can exploit such symbolic gaze cursors in all sorts of tasks. In collaborative visual search, participants detected the target faster when a gaze cursor depicted their interlocutor's focus of attention than when they were provided with no partner information, only voice, or even both cursor and voice (Brennan et al., 2008). Similarly, a dynamic gaze cursor proved helpful in detecting bugs in computer programs (Stein & Brennan, 2004).

Other studies have included a real or video-taped speaker (e.g., Hanna & Brennan, 2007; Nappa & Arnold, 2009; Nappa, Wessel, McEldoon, Gleitman, & Trueswell, 2009). Using a collaborative task, Hanna and Brennan (2007) showed that seeing a speaker attending to the object she was about to mention led listeners to shift attention to the corresponding object in their own workspace even before the speaker mentioned it. Similarly, in a sentence verification task, listeners were able to use the gaze of a robot speaker to anticipate a linguistically ambiguous referent (Staudte & Crocker, 2009). In sum, speaker gaze – whether seen directly or represented by a gaze cursor – allows listeners to anticipate what a speaker will refer to, and can rapidly benefit performance in comprehension, visual search, collaboration, problem solving, and spatial referencing.

## Task effects: Visual attention & language processing

However, the impressive range of tasks across which visual context cues can influence cognitive processes does not mean that the allocation of visual attention is task-independent. From the very early days of research on eye movements it has been known that images are scanned with different saccade sequences as a function of task: Participants were more likely to fixate on the faces of people in a painting when asked to determine their ages than when estimating their material wealth (Yarbus, 1967; Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010). More recently, task (visual search vs. memorization) has been shown to affect which image areas are inspected (Castelhano, Mack, & Henderson, 2009).

Task effects on gaze behavior have also been reported in language processing, particularly in language production. Thus, a speaker's fixation pattern depends among other things

on whether s/he is inspecting an object or preparing to name it (Meyer, Sleiderink, & Levelt, 1998), producing an active versus a passive description (Griffin & Bock, 2000), telling the time in an analogue versus digital format (Bock, Irwin, Davidson, & Levelt, 2003), or speaking about visible versus remembered objects (Meyer, van der Meulen, & Brooks, 2004). Moreover, eye movements are affected by the processing of linguistic information in language-based tasks (e.g., reading and object recognition), but not in non-linguistic tasks (e.g., visual search) (Rayner & Raney, 1996; Zelinsky & Murphy, 2000). Clearly, the instructions given to participants can affect the interpretation of eye movement data (see Knoeferle, Crocker, Scheepers, & Pickering, 2005, p.109).

Within the domain of spoken language comprehension and “visual world” studies – in which participants’ fixations of objects are monitored as they listen to a related sentence – two typical tasks are acting-out (e.g., Spivey, Tanenhaus, Eberhard, & Sedivy, 2002) and passive listening (e.g., Altmann & Kamide, 1999). Descriptive comparisons of fixation patterns between different studies suggest no obvious task-based discrepancies in the time course with which comprehenders inspect and anticipate objects. But direct and controlled manipulations of task across otherwise similar visual world studies are, to the best of our knowledge, lacking. As a result, the potential effects of more subtle variations of comprehension-related tasks are not explicitly considered in existing accounts of situated comprehension (Altmann & Kamide, 2007; Knoeferle & Crocker, 2006) and associated computational models (Mayberry, Crocker, & Knoeferle, 2009). The linking hypotheses between visual attention and language comprehension that underlie these accounts also don’t take potential top-down task effects into account.

### Task and speaker gaze effects: The present studies

Two eye-tracking experiments connected these separate strands of research: the use of dynamic speaker gaze in on-line sentence processing, and variations in the comprehension task (for further details, see Kreysa & Knoeferle, 2011). In both experiments, people watched videos of a speaker producing German subject-verb-object (SVO) and object-verb-subject (OVS) sentences about characters on a screen. Following each video, participants were instructed to verify specific aspects of the sentence: In Experiment 1, the task was to identify the patient, while Experiment 2 required participants to verify thematic role relations, a task which arguably subsumes the patient identification task.





Consider an example: A speaker looks at a computer display that shows a waiter, a millionaire, and a saxophone player (Fig. 1a). As soon as she begins her sentence with *Der/ Den Kellner* (“the waiter”), case marking identifies the first noun phrase (NP1) as either the subject (*Der*, Table 1, c & d) or the object (*Den*, Table 1, a & b). In German, both subject- and object-initial main clauses are grammatical, but the former are canonical while the latter are not. Understanding an object- (vs. subject-) initial sentence has been shown to slow comprehension (as reflected by eye movements), although

case marking, world knowledge, and factors such as intonation or visual context can modulate this time course (Kamide, Scheepers, & Altmann, 2003; Knoeferle et al., 2005; Weber, Grice, & Crocker, 2006). In the present study, when the sentence continues with *beglückwünscht* (“congratulate”) and the NP2 determiner *den/ der*, neither linguistic information nor world knowledge reveals which of the two other depicted characters (the millionaire or the saxophone player) will be referred to post-verbally. Thus, while the sentence is structurally unambiguous, there is a temporary referential ambiguity at the verb.



Figure 1: Screen displays: (a) Example of still from the videos used in Experiments 1 and 2; (b) Template for patient verification (Exp. 1): Does the circled character correspond to the patient of the sentence?; (c) Template for verifying role relations (Exp. 2): Does the arrow reflect the thematic roles of the sentence?

Table 1: Overview of the experimental conditions (Congruency is excluded here). The English translation of the SVO sentence is “the waiter congratulates the millionaire”, while the OVS sentence implies that the waiter is *being* congratulated by the millionaire.

Condition	Picture	Sentence	
OVS & NoGaze	a 	<i>Den beglückwünscht Millionär.</i>	<i>Kellner der</i>
OVS & Gaze	b 	<i>Den beglückwünscht Millionär.</i>	<i>Kellner der</i>
SVO & NoGaze	c 	<i>Der beglückwünscht Millionär.</i>	<i>Kellner den</i>
SVO & Gaze	d 	<i>Der beglückwünscht Millionär.</i>	<i>Kellner den</i>

However, if the speaker now shifts gaze from the referent of the NP1 to the post-verbal referent, the direction of her gaze could allow the listener to anticipate the latter even before hearing the NP2. If speaker gaze in a setup such as Figure 1a is used to anticipate post-verbal referents, listeners should begin to fixate this “target” referent shortly after the

speaker begins to gaze at it, and more often than when the display doesn't show the speaker. Such speaker-gaze based anticipation could either be independent of, or modulated by, syntactic structuring and thematic interpretation. If it is independent of sentence structuring, then post-verbal referent anticipation should occur to the same extent and with the same time course for both SVO and OVS sentences. Alternatively, if speaker gaze effects on referent anticipation interact with syntactic structuring, then we should see differences in the time course and/ or extent to which a listener inspects the target referent for OVS relative to SVO sentences. If such effects are long-lasting, they could also affect the post-sentence verification response latencies.

Crucially, all or none of these speaker gaze effects on a listener's visual attention could vary as a function of the two different comprehension tasks: patient verification (Exp. 1) and role relations verification (Exp. 2). Observing similar speaker gaze effects across these two tasks would suggest that the use of speaker gaze is independent of subtle task differences. Alternatively, people's eye gaze and verification times may be affected by the different aspects of comprehension that each verification task focuses on. If so, we should see differences in speaker gaze effects and sentence structure effects as a function of task.

## Experiments 1 and 2

### Methods

**Participants** Thirty-two Bielefeld University students took part in Experiment 1 (15 male; 3 replacements), and a further 32 participated in Experiment 2 (5 male). All were native German speakers with normal or corrected-to-normal vision. All gave informed consent.

**Materials and Design** We created 72 characters in the virtual world SecondLife, and 48 critical sentences (NP1-VERB-NP2-PP). We grouped the characters into 24 triplets, and took a snapshot of them. Each snapshot was paired with two German sentences (SVO and OVS) to create 24 items. Each sentence described a transitive action between the central character (e.g., the waiter) and one of the two outer characters (e.g., the millionaire; Table 1). None of the nouns in the sentence were semantically associated, nor was there a semantic connection with the verb. Actions were not depicted. A naming pretest ensured all characters were recognizable.

We recorded two videos for each item, showing the speaker producing the sentences about the characters. She was seated to the right of a 20" Apple iMac 8.1 screen, which displayed the SecondLife triplet. A Canon PowerShot G10 camera was positioned in such a way that both screen and speaker were visible in the recording. Videos began with the speaker looking at the camera and smiling briefly. She then inspected all three characters in a fixed order, so that participants could establish what a gaze to each of them looked like. Finally, her gaze returned to the central character, who was always the referent of the NP1. She then began producing the sentence, which had been read out to her previously. The speaker

always looked at the character she was mentioning. Thus, shortly after uttering the verb, her gaze shifted from the NP1 referent to the NP2 referent. A second pretest ensured that this gaze shift was easy to see (98% correct; detection latency  $M = 498$  ms,  $SD = 386$ ).

The design included three within-subject factors (Table 1): *Gaze* (speaker vs. not), *Structure* (SVO vs. OVS), and *Congruency* between the sentence content and a post-sentence response template (see Procedure). The display versions and sentence manipulations were allocated such that each sentence role (agent or patient) was equally distributed across screen positions over the course of the experiment. In addition, the NP2 referent appeared on the same side of the screen equally often, so that the speaker shifted her gaze to the right just as frequently as to the left. The 24 experimental items were supplemented by 48 fillers with different sentence structures and images. The speaker was visible on 50% of trials.



Figure 2: Eyetracking setup: Participants watched the video on the screen, then pressed a button in response to the template.

**Procedure** We monitored eye movements using an Eye-link 1000 desktop head-stabilized tracker (SR Research), and recorded post-sentence verification latencies (see Fig. 2 for the experimental setup). On Gaze trials, participants saw the speaker talking about the SecondLife characters on the screen. On NoGaze trials, the same video was shown, but the speaker was occluded behind a grey bar. Thus, only the static screen with the three characters was visible (see Table 1, a & c). Immediately following the end of each video, participants saw a template like Figures 1b or 1c. Their task was to press a button depending on whether the template accurately depicted the sentence content ('yes' vs. 'no'). For Experiment 1, the template represented the character who had been mentioned in the patient role in the sentence (see Fig. 1b). For a video such as Figure 1a and the sentence *Den Kellner beglückwünscht der Millionär* ("The waiter is congratulated by the millionaire", OVS), the correct response to the template in Figure 1b is 'yes': The position of the waiter (i.e., the middle character) is circled. In Experiment 2, participants verified whether the arrow on the template correctly depicted who-does-what-to-whom in the sentence (Fig. 1c). Thus, for the same sentence, Figure 1a followed by Figure 1c would also require a 'yes' response, because the arrow points from the position of the millionaire on the right (the agent of the sentence) to the waiter (the patient) in the middle.

**Eye movement analysis** For the eye movement analyses, we selected two critical time windows during the video. The first (“SHIFT”) comprised all fixations that began between the speaker’s gaze shift and the onset of the NP2. The second time window (“NP2”) comprised all fixations starting during the NP2. The x-y coordinates of participants’ fixations were assigned to four areas of interest: NP1 referent, target (= NP2 referent), competitor (= the non-mentioned character), and the area around the speaker. The main dependent variable was the number of fixations to the target, i.e., the referent of the NP2. Log-linear models were used for the inferential analysis, combining characteristics of a standard cross-tabulation chi-square test with those of ANOVA. They included the factors *Gaze* (Gaze vs. NoGaze), *Structure* (SVO vs. OVS), and either participants ( $N = 32$ ) or items ( $N = 24$ ). Finally, a model including *Experiment* as a factor allowed us to assess the generalizability of effects across tasks.

## Results Experiment 1 (Verifying the patient)

**Response time results** Response times were measured from the onset of the verification template until participants’ button press (96% accuracy). A  $2 \times 2 \times 2$  (Structure\*Gaze\*Congruency) repeated-measures Anova on log-transformed response times revealed faster responses to matching than mismatching templates ( $ps < .001$ ). Neither Structure nor Gaze had any effect.

**Eye movement results** Figure 3 shows proportions of fixations in all interest areas, for the Gaze vs. NoGaze conditions during the SHIFT time window. Generally, participants still tended to fixate on the NP1 referent, who had just been mentioned. However, in the Gaze condition, fixations to the as-yet-unmentioned NP2 referent increased shortly after the speaker shifted her gaze. Note that the speaker herself was rarely fixated at all.

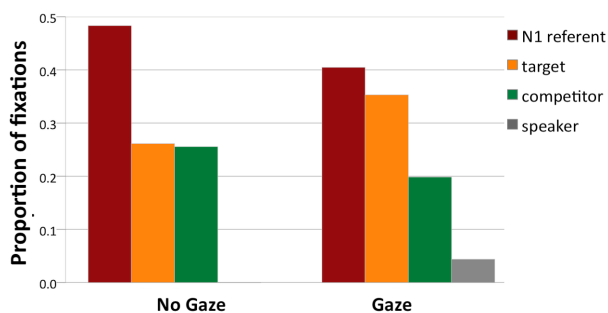


Figure 3: Distribution of fixations beginning in the SHIFT time window across areas of interest, by speaker visibility (Exp. 1).

Figure 4 presents the time course of participants’ fixations to the target character only, from the onset of the speaker’s gaze shift, as a function of structure and gaze. Like Figure 3, it shows an earlier rise of looks to the target character in the Gaze than in the NoGaze conditions, for both sentence structures. This begins about 500 ms after speaker gaze shift, and well before the onset of the NP2. Only much later, roughly at the offset of the NP2, do participants in the NoGaze condi-

tions fixate the target character to the same extent.

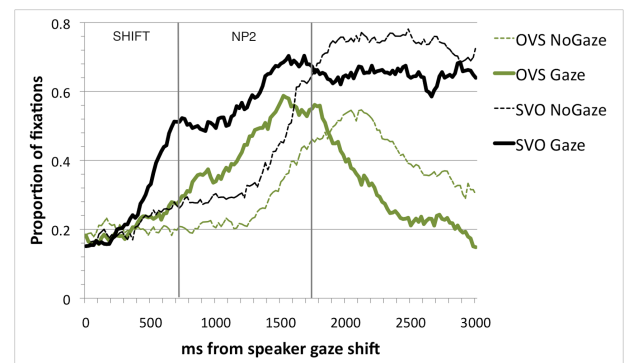


Figure 4: Time course of participants’ fixations to the target character (the NP2 referent) in ms from speaker gaze shift, depending on structure and gaze (Exp. 1). The mean on- and offset of the NP2 are marked as vertical lines.

Analyses for the SHIFT time window confirmed that participants were more likely to inspect the target character when they could see the speaker (35%) than when they could not (26%;  $ps < .05$ , Fig. 3). An effect of Sentence Structure further revealed that people fixated the target character more often in the SVO (36%) than OVS conditions (25%;  $ps < .05$ ). The main effect of Speaker Gaze was also present in the NP2 time window ( $ps < .001$ ). When the speaker was present, participants fixated the target character more (55%) than in her absence (43%). The Structure effect from the previous time window carried through too: Participants looked at the target character more while hearing an SVO sentence (54%) than during OVS (45%;  $ps < .001$ ). There was no reliable interaction of Gaze and Structure in either time window.

## Results Experiment 2 (Verifying role relations)

**Response time results** Participants’ responses to the who-does-what-to-whom template were 96% accurate. Just as in Experiment 1, matching templates elicited faster responses than mismatches ( $ps < .001$ ), and Speaker Gaze had no reliable effect on response times. Unlike for the patient verification task, however, role relations verification led to a significant main effect of Sentence Structure ( $ps < .05$ ), such that SVO sentences elicited faster responses than OVS (71 ms).

**Eye movement results** Figure 5 shows proportions of fixations in all interest areas for the Gaze vs. NoGaze conditions during the SHIFT time window. Figure 6 presents the time course of participants’ fixations to the target character. As in Experiment 1, these began to increase almost as soon as the speaker shifted her gaze, well before this character was mentioned (and earlier than when no speaker gaze was available). At the end of the sentence, the gaze pattern also differed from Experiment 1 (Fig. 4): There, participants in the SVO condition predominantly fixated the sentence-final patient, whereas this was not the case for Experiment 2 (Fig. 4).

During the SHIFT time window, log-linear analyses confirmed an effect of Gaze on fixations to the target character: Just like in the patient verification task, participants were



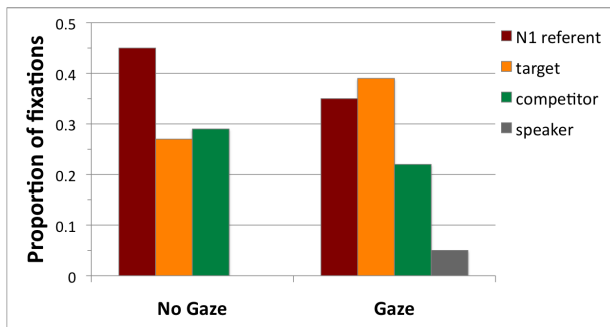


Figure 5: Distribution of fixations in the SHIFT time window across interest areas, depending on speaker visibility (Exp. 2).

more likely to fixate the target when they could (vs. couldn't) see the speaker (39% vs. 27%;  $ps < .001$ ). Sentence Structure also had a significant effect in the SHIFT window, although unlike in Experiment 1, participants fixated the target character more often when hearing an OVS relative to an SVO sentence (35% vs. 30%;  $ps < .05$ ). Finally, also unlike Experiment 1, the interaction of Gaze and Structure was significant ( $ps < .05$ ): The facilitative effect of Gaze was considerably larger for subject- than object-initial sentences, as can be seen in Figure 6. In the NP2 time window, the only reliable effect was one of Gaze ( $ps < .001$ ), with participants fixating the target character more often with (63%) than without (47%) the speaker.

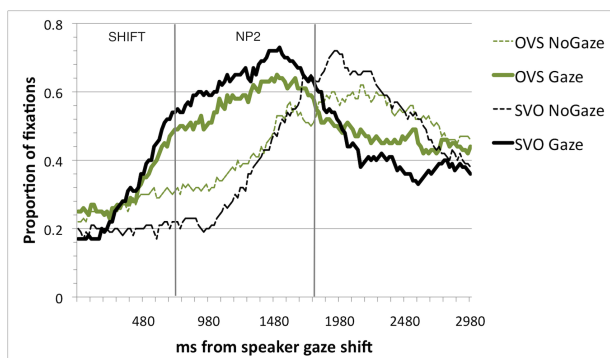


Figure 6: Time course of participants' fixations to the target character, depending on structure and gaze (Exp. 2). The mean on- and offset of the NP2 are marked as vertical lines.

In cross-experiment analyses, the factor Experiment had no reliable effect on response latencies. Crucially however, it affected fixation patterns: During the SHIFT time window, participants were more likely to fixate the N2 referent in Experiment 2 than in Experiment 1 ( $ps < .05$ ). Structure interacted with Experiment, with increased fixations to the NP2 referent when hearing an SVO sentence in Experiment 1, but when hearing an OVS sentence in Experiment 2 (relative to the respective other structure,  $ps < .001$ ).

## General Discussion

We assessed whether speaker gaze effects on both response latencies and visual attention during comprehension for a ver-

ification task varied as a function of subtle task differences. To this end, we recorded participants' gaze as they listened to NP1-VERB-NP2 sentences mentioning two out of three characters on a computer screen. On half the trials, they saw a speaker shifting gaze at the verb from the NP1 referent to the NP2 referent. Subsequently, participants verified whether a circled character corresponded to the patient of the sentence (Exp. 1), or whether an arrow between two characters correctly depicted their thematic role relations (Exp. 2; note that this task requires having identified the patient correctly).

As expected, response latencies in both experiments were shorter when the template matched (vs. mismatched) the video in the to-be-verified aspects. However, sentence structure affected response times only when people judged thematic role relations, but not when they verified the identity of the sentential patient. This suggests that the thematic role task may have required more in-depth syntactic processing. The moment-by-moment allocation of visual attention supports this conclusion: While sentence structure reliably affected anticipatory eye movements to the post-verbal referent in both experiments, its effect differed between the two associated tasks. Patient verification led to more target fixations during SVO than OVS sentences, while this pattern flipped for thematic role verification. This may be due to task differences in the informativity of the gaze shift: For patient verification, only gaze shifts in SVO sentences are task-relevant (the patient is already uniquely identified in OVS sentences). In contrast, for thematic role verification, the gaze shift is informative in both sentence structures, since this task relies on identifying two characters. In addition, it seems possible that during normal sentence processing (i.e., in Exp. 2), there may be a tendency to fixate the agent while hearing the verb. In contrast, if the task is explicitly to identify the patient (Exp. 1), an efficient strategy would be to locate this character as early as possible and pay less attention to the remainder of the sentence.

Importantly, while the availability of speaker gaze led to substantially earlier anticipation of the NP2 character across the board, this benefit was also modulated by sentence structure – the greatest facilitation occurred for canonical SVO sentences in the role verification task. Task differences became even more obvious later in SVO sentences, when it was advantageous for listeners who had to verify the patient to maintain fixation on the NP2 referent. It seems then that task can critically affect syntactically-driven eye movements in online spoken language comprehension. In sum, to accurately account for effects of visual context (e.g., speaker gaze) and syntactic structure on the deployment of visual attention, processing accounts of situated language comprehension must include a model of task constraints.

## Acknowledgments

This research was funded by the Cognitive Interaction Technology Excellence Center (German research foundation, DFG). We thank Eva Mende, Linda Krull, Anne Kaestner,

Lydia Diegmann, and Eva Nunnemann for their assistance with preparing the stimulus materials and/ or collecting data.

## References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- 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. *JML*, *57*, 502–518.
- Bock, K., Irwin, D., Davidson, D., & Levelt, W. (2003). Minding the clock. *JML*, *48*, 653–685.
- Brennan, S., Chen, X., Dickinson, C., Neider, M., & Zelinsky, G. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*, 1465–1477.
- Carletta, J., Hill, R., Nicol, C., Taylor, T., de Ruiter, J., & Bard, E. (2010). Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods*, *42*, 254–265.
- Castelhano, M., Mack, M., & Henderson, J. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*, 115.
- Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the “give” and “take” between event apprehension and utterance formulation. *JML*, *57*, 544–569.
- Grant, E., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*, 462–466.
- Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.
- Hanna, J., & Brennan, S. (2007). Speakers’ eye gaze disambiguates referring expressions early during face-to-face conversation. *JML*, *57*, 596–615.
- Kamide, Y., Scheepers, C., & Altmann, G. (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., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*, 481–529.
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*, 95–127.
- Kreysa, H. (2009). Coordinating speech-related eye movements between comprehension and production. Unpublished doctoral dissertation, Edinburgh University, UK.
- Kreysa, H., & Knoeferle, P. (2011). Peripheral speaker gaze facilitates spoken language comprehension: Syntactic structuring and thematic role assignment in German. In *Proceedings of the European Conference on Cognitive Science in Sofia*.
- Langton, S., O’Donnell, C., Riby, D., & Ballantyne, C. (2006). Gaze cues influence the allocation of attention in natural scene viewing. *Quarterly Journal of Experimental Psychology*, *59*, 2056–2064.
- Mayberry, M., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, *33*, 449–496.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye-movements during noun phrase production. *Cognition*, *66*, B25–B33.
- Meyer, A. S., van der Meulen, F., & Brooks, A. (2004). Eye movements during speech planning: Talking about present and remembered objects. *Visual Cognition*, *11*, 553–576.
- Nappa, R., & Arnold, J. (2009). Paying attention to intention: Effects of intention (but not egocentric attention) on pronoun resolution. In *Proceedings of the CUNY Conference*.
- Nappa, R., Wessel, A., McElloodon, K., Gleitman, L., & Trueswell, J. (2009). Use of speaker’s gaze and syntax in verb learning. *Language Learning and Development*, *5*, 1–32.
- Rayner, K., & Raney, G. E. (1996). Eye-movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*, 245–248.
- Ricciardelli, P., Bricolo, E., Aglioti, S., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual’s gaze. *NeuroReport*, *13*, 2259–2264.
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye-movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.
- Staudte, M., & Crocker, M. (2009). The effect of robot gaze on processing robot utterances. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 431–436). Cognitive Science Society, Inc.
- Stein, R., & Brennan, S. (2004). Another person’s eye gaze as a cue in solving programming problems. In *Proceedings of the 6th International Conference on Multimodal Interface* (p. 9–15). Penn State University.
- Tatler, B., Wade, N., Kwan, H., Findlay, J., & Velichkovsky, B. (2010). Yabus, eye movements, and vision. *i-Perception*, *1*, 7–27.
- Tomlin, R. S. (1995). Focal attention, voice, and word order. In P. A. Dowing & M. Noonan (Eds.), *Word Order in Discourse* (p. 517–552). Amsterdam: John Benjamins.
- Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*, B63–B72.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, *11*, 125–131.