Title
The Facilitatory Effect of Referent Gaze on Cognitive Load in Language Processing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Sekicki, Mirjana
Staudte, Maria

Publication Date
2017

Peer reviewed

The Facilitatory Effect of Referent Gaze on Cognitive Load in Language Processing

Mirjana Sekicki (mirjana@coli.uni-saarland.de)

Maria Staudte (masta@coli.uni-saarland.de)
Department of Language Science and Technology; Saarland Univeristy, Germany

Abstract

This paper considers prediction in language processing by examining the role of the visual context, and specifically, the role of speaker referent gaze on cognitive load. We inspect the anticipatory visual attention during sentence processing together with the cognitive load induced at the points of the gaze cue, and the linguistic referent. Employing a novel measurement of cognitive load - the Index of Cognitive Activity (Marshall, 2000) allowed us to simultaneously consider both anticipatory eye-movements and cognitive load. Our results show that the gaze cue is being followed, and considered as a relevant piece of information, which subsequently reduces the cognitive load on the linguistic referent. In addition, we found that considering the gaze cue is in itself not costly, unless it cues an object mismatching with the previous linguistic context.

Keywords: Gaze; Cognitive Load; Index of Cognitive Activity; Prediction; Eye-tracking

Introduction

A series of investigations in the visual world paradigm (VWP) have shown how listeners simultaneously combine linguistic and visual cues to predict upcoming linguistic input (for a review of the VWP see Huettig, Rommers, & Meyer, 2011). Based on the idea that prediction is a unifying principle of the human mind, a large body of psycholinguistic research, in the VWP, as well as employing EEG, has been examining the role of prediction in language processing (see Huettig, 2015; Huettig & Mani, 2016). Anticipatory eye-movements collected in the VWP have reliably shown that people predict upcoming referents based on the previous linguistic material (e.g. Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003), as well as based on the visually presented events (Knoeferle, Crocker, Scheepers, & Pickering, 2005).

Our present work examines the influence of the visual modality on processing linguistic information by specifically investigating speaker gaze, as an inseparable part of the visual context in situated communication, and its influence on prediction making. We hypothesized that speaker gaze to the upcoming referent helps constrain the set of possible targets and thus, by increasing the predictability of the cued object reduces the cognitive load induced by its linguistic referent. In addition, we examined whether any cost reduction on the referent would be accompanied by a cost increase on the gaze cue, effectively spreading the cognitive load across the two modalities, namely, the gaze cue and the language.

The Gaze Cue Gaze has been shown to play an important role in situated communication. Listeners inspect objects they anticipate will be mentioned next (Altmann & Kamide, 1999), and they fixate the mentioned object 200 - 300 ms after the speaker started referring to it (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Speakers also fixate the relevant object 800 - 1000 ms before mentioning it (e.g. Griffin & Bock, 2000). But, how does speaker's gaze add to the listener's prediction?

Previous research on gaze and language has established the gaze cue to be utilized and proven helpful while processing linguistic material (e.g. Hanna & Brennan, 2007) even when the speaker is a robot (Staudte & Crocker, 2011) or a virtual agent (Staudte, Crocker, Heloir, & Kipp, 2014). The conclusions about the effects of gaze following are drawn on the basis of participants' eye-movements and responses to the task at hand. The eye-movement data give insight into the shifts of visual attention indicating whether the gaze cue was considered. In addition, the effect of considering the gaze cue on the comprehension of linguistic material is assessed by reaction times, comprehension and production tasks. Importantly however, no direct effect of speaker gaze on cognitive effort required for language processing has, to our knowledge, been examined yet. This paper sets out to investigate just that by combining eye-movement data with an online pupillary measure of cognitive load, which enabled us to measure this cost / benefit directly, and independent of a specific task.

We examined both the process of creating and (dis)confirming predictions in one and the same setting. To this end we employed a novel measurement of cognitive load - the Index of Cognitive Activity (ICA; Marshall, 2000) which allows for an experimental design that combines eye-movements, shedding light on the predictive processes, while simultaneously measuring cognitive load at different stages of sentence processing. The results revealed that referent gaze is followed and that it indeed adds to the predictability of the upcoming referent such that the spoken reference induces less cognitive load. Note that whether this effect could also be induced by arrows or other visual pointers is considered irrelevant at this point.

The Index of Cognitive Activity (ICA) The two experiments presented in this paper are conducted in the VWP. We considered the eye-movements and the cognitive load reflected in pupil size. While the traditional eye-movement analysis helps reveal any patterns of anticipation of potential target objects, the ICA allowed us to simultaneously also measure cognitive load both on the gaze cue and on the referent noun, i.e. the cognitive load induced by creating and (dis)confirming one's predictions.

Pupil dilation happens in consequence of either changes in light or cognitive activity. Two groups of muscles are responsible for pupil size: circular muscles that make the pupil contract, and radial muscles that make it dilate. Pupil size changes due to light and due to cognitive activity employ different activation and inhibition processes, the dilation due to cognitive load being shorter and more abrupt (Beatty, 1982).

The ICA measurement disentangles the two types of change in pupil size by performing a wavelet analysis on the pupil dilation record and removing the large oscillations, while considering only the quick pupil jitter that is related to cognitive activity (Marshall, 2000). Such events of small abrupt changes in pupil size are referred to as the ICA events.

The ICA has been shown to reflect changes in cognitive load in a variety of different studies since its appearance (e.g. Marshall, 2002, 2007). However, only recently has it been examined with cognitive load induced by linguistic processing (Demberg & Sayeed, 2016). Demberg and Sayeed present a series of seven experiments showing that the ICA indexes linguistic processing difficulty for both reading and auditory presentation of linguistic stimulus. In addition, the ICA proved to be robust with respect to eye-movements making it a valid measure of processing difficulty in the VWP.

Hence, employing the ICA in present experiments allowed for simultaneous assessment of both visual attention and cognitive load.

Current Questions and Predictions Two studies were set out to examine whether gaze is considered as part of the context determining the predictability of the subsequent referent. We examined if the gaze cue actually helps reduce the cognitive load of a linguistic referent online and whether cognitive load is in fact spread across gaze and spoken reference such that the gaze cue itself then induces higher cognitive load.

Experiment 1 made use of the gaze cue that was always fitting (the previous linguistic context) and congruent (cuing the object to be referred to linguistically). We manipulated the **existence** of the gaze cue in order to answer the following research questions:

a) Does the gaze cue influence the predictability of a linguistic referent?

b) If so, how does it influence the cognitive load induced by the referent?

c) Can we measure cognitive load on the gaze cue itself?

Experiment 2 also made use of congruent gaze, while manipulating the **fit** of the referent (thus, also the fit of the gaze cue) with the previous linguistic context. This was done in order to answer the following research questions:

a) Does the gaze cue help reduce the cognitive load on the linguistic referent even when they both do not fit the previous linguistic context?

b) Does the gaze cue to a mismatching object itself induce higher cognitive load?

We expected mismatching gaze to be surprising and thus, more costly, which would as its consequence have a reduction in cognitive load on the corresponding linguistic referent.

Experiment 1

This study aimed to examine whether the online measure of cognitive load also supports previous findings that the gaze cue is actively considered in language processing, by quantifying how its existence modifies the cost induced by the linguistic referent. In addition, we were interested in measuring the potential cost of gaze perception.

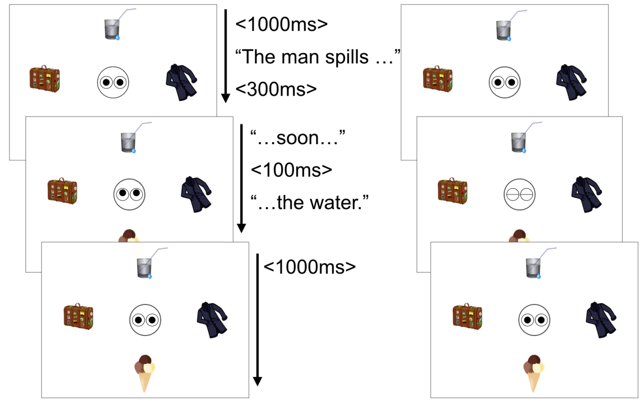

Figure 1: A trial timeline example (from Exp.1) - referent gaze condition (left); and no-gaze condition (right).

Method

The study made use of 2x2x2 mixed factorial design. The independent variable Gaze was a between subjects variable, i.e. half the participants were presented with the version of the experiment where all items included the gaze cue, while the rest saw the version with items never having the gaze cue. Fillers balanced the gaze conditions to the ratio of 1:1. In addition, four linguistic conditions were created with the two within subjects variables, Constraint and Plausibility. Constraint was manipulated by verb restrictiveness (*spill* vs. *order*), and Plausibility by noun fit (*spill*: *water* vs. *ice-cream*).

Participants 64 students of Saarland University took part in this study (45 women) and were monetarily reimbursed for their partaking. Their age ranged from 18 to 34 years old (M = 24.16). Participants were all native speakers of the German language with normal or corrected to normal vision.

Materials and Design Each participant was presented with 20 items and 30 fillers, both consisting of visual and auditorily presented linguistic stimulus. In addition, visual displays included a face-like object forming the gaze cue.

Note that the gaze cue used in our studies, since being an always congruent visual pointer, is arguably not different from an arrow. This is true, but, currently irrelevant, since the differences between a visual pointer and a gaze cue are potentially to be expected in cases of manipulated congruence.

We made use of simple German sentences (Subject - Verb - Adverb - Object) that included a restrictive (*spill*) and a non-restrictive (*order*) verb and two object nouns (*water* vs. *ice-cream*) of differing semantic fit in relation to the restrictive verb. The chosen nouns were controlled for frequency and two pretests have shown that in the context of *order*, *water* was more predicted (cloze probability of 13.67%; plausibility rating of 1.12 on a 7-point Likert scale[1]), than *ice-cream* (cloze

[1]The scale ranged from "very plausible" (1), to "not plausible,

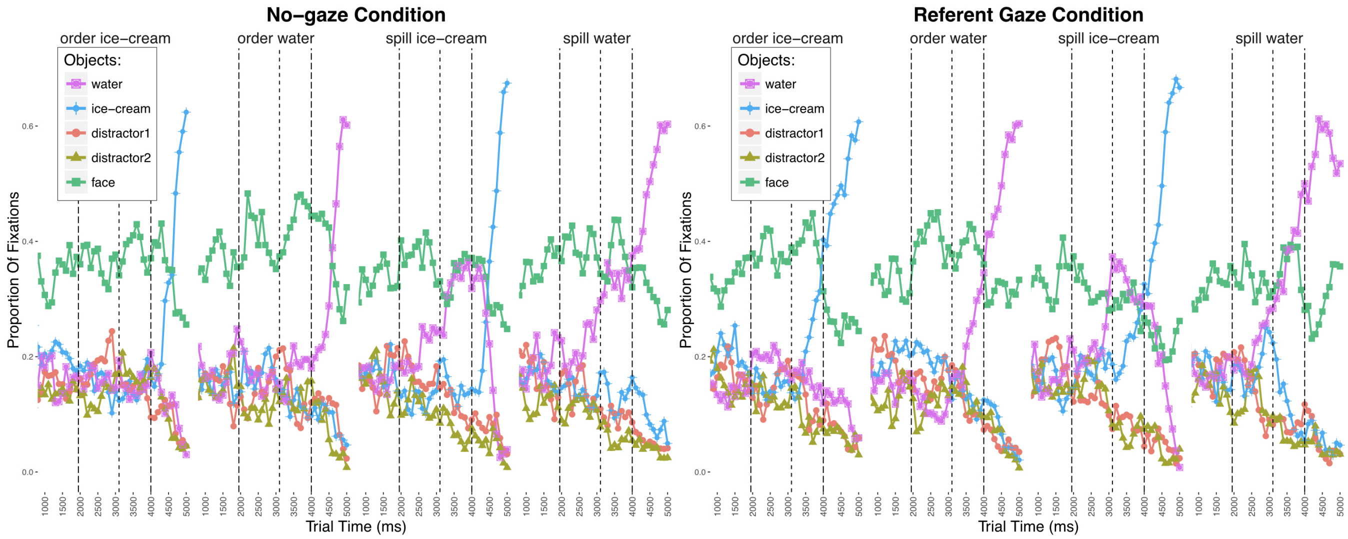

Figure 2: (Exp.1) Proportion of fixations in the four linguistic conditions without the gaze cue (left) and with ref. gaze (right). The verb onset, gaze and noun onset are shown averaged across trials, and aligned to the 100 ms bins within which they fall.

probability of 0.16%; plausibility rating of 2.76). The same adverb, neutral in meaning (*gleich* - ENG: *soon*), was used for all items and served the purpose of a spillover region.

In addition, visual displays with four concrete objects[2] were presented. Two of the four objects fit the category introduced by the restrictive verb (*spill*: *water*, *ice-cream*), while all four fit the non-restrictive verb (*order*: *water*, *ice-cream*, *suitcase*, *coat*). The referent noun was always fitting with the previous linguistic context and the gaze cue was always congruent, that is, cuing the object that is about to be mentioned. The main manipulation of the study was the presence (vs. absence) of the gaze cue which was presented before the target object was referred to verbally.

Figure 1 illustrates a trial timeline. The visual scene with open eyes was presented 1000 ms prior to sentence onset. The gaze cue (or closed eyes) was introduced 300 ms after the verb, i.e. from adverb onset to sentence end. Finally, the eyes would look straight for another 1000 ms.

Fillers Fillers included the same visual setting, but differed in the structure and complexity of the linguistic stimulus and the number of objects that fit the verb category. 30 fillers were used, 25 of which had the opposite and 5 the same gaze condition as the items (ratio of 1:1 for gaze and no-gaze). 19 fillers were followed by simple yes/no comprehension questions that were answered on a key-press. The questions were related exclusively to the linguistic content. This was done in order not to inspire extensive inspection of the visual scene, but rather so that participants consider it only optionally and freely in addition to the linguistic information.

Procedure An EyeLink II head-mounted eye-tracker (SR Research, Ltd; Mississauga, Ont., Canada) was used to track both eyes at a sampling rate of 250 Hz.[3] Participants were instructed to listen carefully to the sentences while looking freely at the presented objects. They would advance the experiment on a button press after each trial. Other two buttons were used to answer the comprehension questions. The experimental session was preceded by a three-trial practice session. The experiment lasted for approximately 15 minutes.

Results

First, in order to gain insight into the patterns of prediction and visual attention we consider the proportion of fixations to the presented objects throughout a trial. Second, we analyse new inspections. Consecutive fixations to the same interest area are considered as one inspection. Since we are interested in the shift of attention inspired by a relevant stimulus, we analyse new inspections, i.e. the first inspection to an interest area that started after the linguistic or visual point of interest (as done in e.g. Staudte & Crocker, 2011). We consider new inspections from verb onset (showing linguistic predictions) and from gaze cue onset (showing if the gaze cue influenced visual attention). Finally, the ICA events are extracted from the pupil jitter, summed over a duration of a relevant time-window and statistically analysed. For the ICA analysis, we considered the gaze time-window and the referent noun window.

Variable Coding and Data Analysis In their VWP experiment, Demberg and Sayeed (2016) establish a time-window taken 600 - 1200 ms from the onset of the critical word to be an appropriate window size and timing for the analysis of the ICA events. Since our critical words differ in length across items, we correct this potential confound by taking a time-window that starts from the middle of a word[4], and con-

difficult to imagine" (7).

[2]Images used (for both experiments) were taken from an open source database (www.openclipart.org) and pretested for naming.

[3]This is the required setup for the subsequent extraction of the ICA events from EyeWorks Workload Module software (version 3.12).

[4]The middle of the referent noun was calculated by taking the audio duration of the whole word and using its half as the starting

sider the following 600 ms. In addition, we analyse the gaze window: 600 ms from the gaze cue onset.

The ICA events are extracted for both eyes separately. Since there is no clear theoretical reason why differences should be expected for the two eyes, we combine the two datasets by summing the ICA events for corresponding time-windows and conduct the analyses on the combined data.

All independent variables were contrast coded for the statistical analysis. New inspections, a binary dependent variable required the use of generalized mixed effects models of binomial type. On the other hand, the analysis of the ICA, a count variable, required the use of generalized mixed effects models with Poisson distribution. All models included a maximal converging random structure for both Item and Subject. The analyses were conducted in R programming environment (R Core Team, 2013) and using the *lme4* package.

Proportion of Fixations Figure 2 illustrates the proportions of fixations to all presented objects during a trial. The first dashed line presents verb onset; second line - gaze onset (not relevant for no-gaze); the third line - referent noun onset. It is apparent that the restrictive verb (*spill*) shifts the focus of visual attention to one particular object (*water*). The less fitting object (*ice-cream*) is considered only upon being referred to linguistically (no-gaze), or earlier, at the point of the gaze cue (referent gaze), confirming that the visual attention is not only influenced by linguistic content but also by the gaze cue.

New Inspections We conducted a statistical analysis of new inspections to an object (*water*, *ice-cream*, distractors) from verb onset (to verb offset). In addition, we consider the new inspections to both *water* and *ice-cream*, from gaze onset (to adverb offset).

Considering the verb window, three identical models were run for the inspections to the three relevant objects.[5] *Looks to water*: A main effect of Constraint (β = -0.361, SE = 0.124, z = -2.904, p = 0.004) suggests that more new inspections to *water* occurred upon hearing *spill* (vs. *order*). *Looks to ice-cream*: No effect of Constraint (p = 0.406) suggests *ice-cream* was looked at with no significant difference in the contexts of both verbs. *Looks to the distractors*: A marginal effect of Constraint (β = 0.153, SE = 0.081, z = 1.89, p = 0.059) suggests that there were somewhat more new inspections to the two distractors in the non-restrictive context of *order*.

Considering the gaze window, we analysed new inspections to both *water* and *ice-cream* together as TargetInspections and considered the effects of gaze on the looks to these two objects.[6] We find a main effect of Gaze (β = 0.427, SE = 0.141, z = 3.022, p = 0.003) confirming that the objects were more readily looked at with the gaze cueing them compared to the absence of gaze, i.e. that the gaze cue caused an immediate shift in visual attention.

point (for each word individually).

[5]NewInspections ~ Constraint + (1 + Constraint | Subject) + (1 + Constraint | Item), family = "binomial"

[6]TargetInspections ~ Gaze + (1 + Gaze | Subject) + (1 + Gaze | Item), family = "binomial"

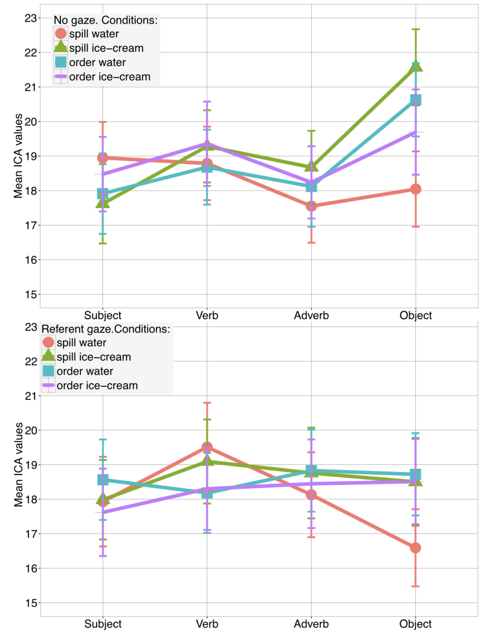

Figure 3: (Exp. 1) ICA events at the four time-windows of a sentence presented for the no-gaze (above) and the ref. gaze (below) conditions separately. (95% CI error bars)

The Index of Cognitive Activity The analysis of the gaze time window did not yield significant results.

Considering the referent noun window[7], we found a main effect of Gaze (β = -0.116, SE = 0.051, z = -2.26, p = 0.024), suggesting that the presence of the gaze cue led to the reduction of cognitive load on the subsequent referent. Moreover, we found a significant Constraint:Plausibility interaction (β = -0.184, SE = 0.042, z = -4.37, p < 0.001), as well as a main effect of Plausibility (β = 0.057, SE = 0.029, z = 2.01, p = 0.045). Further comparisons show a main effect of Plausibility in the subset of *spill* (β = 0.152, SE = 0.035, z = 4.37, p < 0.001), suggesting that *spill water* induced less cognitive load than *spill ice-cream*. No such effect was found in the non-constraining subset (p = 0.249), suggesting no difference between *order water* and *order ice-cream*. Figure 3 illustrates these findings (note: Adverb - gaze window; Object - referent window). Finally, to rule out an effect of experiment part found in the second study, experiment Half was included in the fixed effects structure. We found no Half:Gaze interaction, but a main effect of experiment Half (β = -0.047, SE = 0.013, z = -3.57, p < 0.001), since the second part of the experiment induced less cognitive load.

Discussion

The results show that the gaze cue inspired fixations to the cued object even when it was not predicted by the verb. Moreover, the presence of gaze led to the reduction of cognitive load on the referent noun in all conditions, while preserving the preference for the item that best matched the verb. Interestingly,

[7]ICA ~ Constraint*Plausibility + Half*Gaze + (1 + Constraint*Plausibility || Subject) + (1 + Constraint*Plausibility || Item), family = poisson (link = "log")

the existence of the gaze cue did not in itself induce additional cost on cognitive load.

Hence, we saw that the gaze cue influences predictability of the linguistic referent and subsequently reduces the cognitive load it induces. Interestingly, on the cue itself, no differences in cognitive load were induced either by its mere existence, or by whether it was cuing an object already anticipated based on the linguistic context.

Experiment 2

The second study aimed at examining, firstly, whether the gaze cue helps reduce cognitive load on the linguistic referent even when they are both mismatching with the previous linguistic context, and, secondly, whether the cue to such an object is in itself more costly, since unexpected.

Method

Participants 36 students of Saarland University (23 female) took part in the study and were monetarily reimbursed for their partaking.[8] Their age ranged from 18 to 34 years ($M = 23.36$). Two students were excluded from the analysis due to technical issues, and two because their mother tongue was established to be Luxemburgish. Thus, 32 participants, German native speakers, were included in the analysis.

Materials and Design This experiment made use of 2x2 experimental design, combining Gaze (no-gaze/referent gaze) and referent noun Fit (fitting/mismatching). Only restrictive verbs were used (*spill*), combined with either a thematically fitting (*water*) or mismatching referent noun (*sausage*). 20 items were created, half of which were anomalous[9]. Note that the gaze cue was, again, always congruent (cuing the object subsequently referred to linguistically). When the referent noun did not fit the previous linguistic context, that made the gaze cue to the object in question mismatching as well. The same procedure was implemented as in Experiment 1.

Fillers The experiment included presenting 75 trials in total, 55 of which were fillers. 20% of the total number of sentences were anomalous (10 items, 5 fillers). The gaze cue was present in 2/3 of all trials (10 items, 40 fillers). Only 16% of all trials included an anomalous gaze cue, i.e. gaze that was cueing a mismatching object (5 items, 3 fillers).

Results

The same measures and analyses were conducted as in Experiment 1, except for the new inspections analysis where only the gaze window was considered, due to the differing experimental design of Experiment 2.

Proportion of Fixations Figure 4 shows the proportion of fixations to all presented objects during a trial. As previously, the first dashed line presents verb onset; the second line - gaze onset (not relevant for no-gaze conditions); and the third line - referent noun onset. We see that the verb shifts the focus of visual attention to one particular object (*water*). The mismatching object (*sausage*) is considered only upon being referred to linguistically (no gaze), or earlier, at the point of the gaze cue (referent gaze). Thus, we observe the same pattern as in Experiment 1, namely, of the gaze cue shifting visual attention, on a par with the linguistic information.

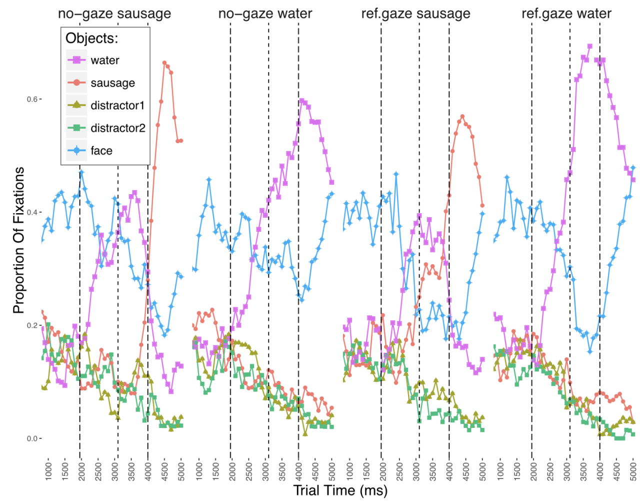

Figure 4: (Exp. 2) Proportion of fixations to presented objects in the four experimental conditions.

New Inspections Considering the gaze window we combined the new inspections to both *water* and *sausage* together as TargetInspections and examined the effect of gaze on the looks to these two objects.[10] A main effect of Gaze (β = 0.381, SE = 0.191, z = 1.995, p = 0.046) confirmed that the gaze was followed, as found in Experiment 1.

The Index of Cognitive Activity We first analysed the referent window[11] and found a main effect of Fit (β = 0.223, SE = 0.043, z = 5.21, $p < 0.001$), suggesting that the anomalous *spill sausage* required more cognitive load than *spill water*. Considering the effect of the gaze cue on the cost of the referent, a significant Gaze:Half interaction was observed (β = -0.126, SE = 0.062, z = -2.05, p = 0.040). Further analysis showed a marginal main effect of Gaze in the second half of the experiment (β = -0.091, SE = 0.047, z = -1.93, p = 0.054), suggesting that the referent gaze reduced the cognitive load on the referent noun in both linguistic conditions. No such effect was found in the first half of the experiment (p = 0.392).

Since gaze affected referent processing in each experimental half differently, we considered cognitive load on the cue itself (gaze window) for each half separately.[12] The first part of the experiment revealed a Gaze:Fit interaction (β = 0.179, SE =

[8]None of the students were familiar with Experiment 1.

[9]In order to counterbalance the referent nouns the experiment was run in two versions. Version a) included a verb fitting to one noun in the item (*spill water* vs. *sausage*), while the verb in version b) fit the other noun (*grill sausage* vs. *water*).

[10]TargetInspections ~ Gaze + (1 + Gaze | Subject) + (1 + Gaze | Item), family = "binomial"

[11]ICA ~ Gaze*Fit + Half*Gaze + (1 + Gaze*Fit + Half*Gaze || Subject) + (1 + Gaze*Fit || Item), family = poisson (link = "log")

[12]ICA ~ Gaze*Fit + (1 + Gaze*Fit || Subject) + (1 + Fit | Item), family = poisson (link = "log")

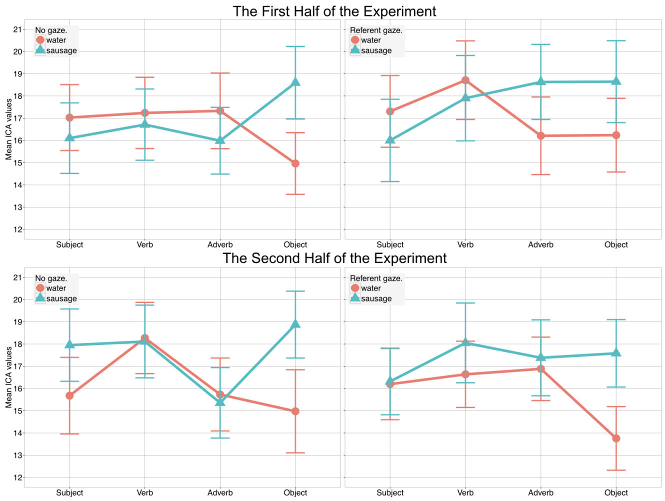

Figure 5: (Exp.2) ICA events at the four time-windows of a sentence in the first (above) and the second (below) half of the experiment. (95% CI error bars)

0.084, $z = 2.13$, $p = 0.033$). In the second half, a main effect of Gaze ($\beta = 0.099$, $SE = 0.045$, $z = 2.20$, $p = 0.028$) suggests that the referent gaze induced higher cognitive load than the no-gaze condition. The results are illustrated in Figure 5 (note: Adverb - gaze window; Object - referent window).

Discussion

The eye-movements data showed evidence of gaze following even when it was unpredicted, or worse, also mismatching. Regarding the cost of processing the referent noun, initially, the existence of gaze did not have an effect; but the cognitive load induced by the mismatching cue itself was higher than that induced by both fitting gaze and no gaze cue. However, in the second half of the experiment, cognitive load on the referent noun was marginally reduced due to the helpful gaze cue; while the cue itself (to both fitting and mismatching object) now induced higher cognitive load. This suggests that participants gradually adapted to and started relying on the surprising gaze cue (increasing load on the cue) and started making use of its informativity (lowering load on the noun).

Conclusions

Referent gaze is actively considered in the process of prediction making, shifting the visual attention to the cued object, and leading to the reduction of cognitive load on its linguistic referent. This holds even when the cue (and the corresponding referent) is mismatching with the verb. Gaze perception proved not to be costly unless mismatching with the verb.

Both studies included conditions with unpredicted but congruent gaze cue (Exp.1: *order water*, *order ice-cream*; Exp.2: *spill sausage*). Such a condition induced a higher processing cost on the gaze cue in Exp. 2, but not in Exp. 1. We argue that in Exp. 1 gaze is still processed naturally, as a gaze cue, due to its overall fit, while in Exp. 2, due to the mismatch with the verb, the cue became more salient, treated as a visual pointer, regardless if cuing a fitting or mismatching object. We interpret this as evidence against a spread of cognitive load between gaze and linguistic reference.

In sum, the gaze cue is exploited to predict the upcoming referent such that it can be processed with less effort. If verb selectional features direct visual attention to a particular object and the gaze cue (alternatively) introduces a different object, this creates a shift in visual attention but, does not negate the existing preference. Cognitive load is reduced as an effect of the gaze cue, but only when the cue is established as informative and reliable, and regardless of its contextual fit.

Acknowledgments

The reported research is funded by the Deutsche Forschungsgemeinschaft (DFG) under grant SFB1102, and by the Cluster of Excellence Multimodal Computing and Interaction (MMCI). We thank Christine Ankener, Laura Fraedrich and Luigi Palumbo for help and support throughout the process of working towards this paper.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Beatty, J. (1982). Task-evoked pupilary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276.

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS ONE*, *11*, e0146194.

Griffin, Z. M., & 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. *Journal of Memory and Language*, *57*, 596-615.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118-135.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? probably not. *Language, Cognition and Neuroscience*, *31*(1), 19-31.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151-171.

Kamide, Y., Altmann, G., & Haywood, S. (2003). Prediction and thematic information in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.*, *49*, 133-156.

Knoeferle, P., Crocker, M., Scheepers, C., & Pickering, M. (2005). The influence of immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95-127.

Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. *US Patent*, 6,090,051.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *proceedings of the 7th conference on human factors and power plants* (pp. 7-5–7-9). IEEE.

Marshall, S. P. (2007). Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine*, *78*, B165-B175.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*, 268-291.

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*, 317-328.

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