Multimodal Surprisal in the N400 and the Index of Cognitive Activity

Christine S. Ankener (ankener@coli.uni-saarland.de)
Heiner Drenhaus (drenhaus@coli.uni-saarland.de)
Matthew W. Crocker (crocker@coli.uni-saarland.de)
Maria Staudte (masta@coli.uni-saarland.de)
Department of Language Science and Technology;
Saarland University, Germany

Abstract

A word's predictability or surprisal, as determined by cloze probabilities or language models (e.g. Frank, Otten, Galli, & Vigliocco, 2015) is related to processing effort, in that less expected words take more effort to process (e.g. Hale, 2001). A words surprisal, however, may also be influenced by the non-linguistic context, such as visual cues: In the visual world paradigm (VWP), for example, anticipatory eye movements suggest that comprehenders exploit the scene to predict what will be mentioned next (Altmann & Kamide, 1999). How visual context affects word surprisal and processing effort, however, remains unclear. Here, we present evidence that visually-determined probabilistic expectations for a spoken target word predict graded processing effort for that word, in both pupillometric (ICA) and ERP (N400) measures. These findings demonstrate that the non-linguistic context can immediately influence both lexical expectations, and surprisal-based processing effort.

Keywords: N400; Language processing; EEG; Index of Cognitive Activity; Prediction; Eye-tracking

Introduction

The use of information-theoretic concepts of *surprisal* (Hale, 2001; Levy, 2008) and *entropy* (Shannon, 1949) to describe the predictability of information content have gained much attention in recent psycholinguistic research since they correlate with measures of processing effort (Smith & Levy, 2013; Frank et al., 2015) and allow for quantitative predictions in language processing. One recent approach is to derive surprisal from language models or cloze probabilities in order to quantify the amount of information conveyed by each word. The surprisal values are then typically used as a predictor of processing effort experienced by the comprehender upon encountering these words. This approach to computing and assessing information content, however, inherently neglects the listener and the situation at a particular point in time. After all, seeing an object in the immediate surroundings can make the corresponding noun less surprising and more predictable than it would be according to the linguistic context. Comprehenders could predict the target word in a probabilistic way, based additionally on all copresent target objects. Processing would then be facilitated proportional to its probability in the multimodal context. Alternatively, target word processing could be equally hard, depending only on the linguistic experience and context, but not on the exact number of co-present alternatives in the visual context.

We assess this question, using two different measures of effort (a pupillometric measure (ICA) and ERPs), and present evidence that the processing effort induced by a word is indeed dependent on the referential context provided by a copresent visual display.

Surprisal in the Visual World Paradigm Manipulating or even just establishing a comprehender's knowledge and expectations as she attempts to comprehend and interpret an utterance is indeed very difficult. The Visual World Paradigm (VWP) has provided a powerful tool to overcome these difficulties. Multimodal information can be presented in such a way that it allows both 1) the introduction of situational aspects to language processing and 2) the control and monitoring over referential uncertainty and predictability of a referent noun. That is, listeners' anticipatory eye movements reliably indicate what and when objects are considered as potential verb arguments (Altmann & Kamide, 1999). It seems likely that the anticipation of objects affects the surprisal and therefore the processing effort for a corresponding noun (cf. Tourtouri, Delogu, & Crocker, 2017; Sekicki & Staudte, 2017, for related work), or, that anticipation itself, reflecting lower or greater referential uncertainty prior to the noun, might affect processing effort for the verb.

We employed the VWP with sentences containing transitive verbs and depicted verb arguments (as in Altmann & Kamide, 1999). We further assessed processing effort during anticipation and during reference resolution (verb and object noun regions) using the Index of Cognitive Activity (ICA), which reliably indexes processing effort despite eye movements. An additional EEG experiment tested the effects of visual world information on word processing as reflected in the N400.

The Index of Cognitive Activity (ICA) The ICA was deployed for light-insensitive online assessment of processing effort while participants could move their eyes freely. That is, the measure has been shown to be robust with respect to eye movements and changes in luminocity (Demberg & Sayeed, 2016). In addition to the already established connection between pupil dilations and cognitive effort with respect to language processing (Engelhardt, Ferreira, & Patsenko, 2009), the ICA distinguishes task evoked dilations from contractions caused by light reflex. This is based on the assumption that the eye's circular muscles contract when reacting to light, while radial muscles are activated, when reacting to mental effort (S. P. Marshall, 2000). Effort related dilations are hence tiny, very short and abrupt movements of less than 0.5 mm in extent (Beatty, 1982), which are also less likely to be affected by artefacts due to gaze position in relation to the tracker lense. In order to get the ICA, larger light induced oscillations are discarded while abrupt discontinuities in the pupil signal remain, referred to as ICA events (for a description, see S. Marshall, 2002). The Index then returns the times at which such events occur during trials with a resolution of only 100 ms. For analysis, the number of those events is then counted within a time period of interest.

Originally introduced as a measure of cognitive load in interactions with a visual display, the ICA has recently been proven to be reliable and responsive to cognitive effort induced by processing of language in different contexts (Sekicki & Staudte, 2017). Generally, high ICA values (i.e., more ICA events in a given time window) reflect higher cognitive effort, while low ICA values suggest comparatively less effort. In order to obtain ICA events, we used binocular eye-tracking at 250 *Hz* on an Eye-Link II tracker. The calculation of rapid small dilations from the tracker data was conducted in the *EyeWorks Workload Module* software (Version 3.12).

Experiment 1 - Multimodal surprisal in the Index of Cognitive Activity

Design This experiment was an eye-tracking study in the VWP, designed to observe the effect of visual context on word surprisal and processing effort as reflected by a pupillometric measure. It featured a 1x4 design in which the same utterance was presented in four different visual contexts where either **0**, **1**, **3**, or **4** objects matched the verb ("spillable" objects). Target word surprisal hence only differed when visual information was processed in combination with the sentence, as the linguistic surprisal itself was identical across conditions. All sentences were presented auditorily and always simultaneously with the visual displays.

Visual Materials All scenes presented consisted of four simple pieces of clip art, arranged around the screen center. Among those four objects, we manipulated the number of objects matching the verb constraints, such that either **0**, **1**, **3**, or all **4** objects could be potential target referents (see Fig. 1, from left to right and top to bottom). The scenes were counterbalanced between two items in such a way that for example a **0** target condition picture for one item served as a **4** target condition in another item. Positions of targets, competitors and distractors in the scenes were rotated. None of the clip art items in a scene depicted the highest-cloze noun of the matching sentence. Filler trials introduced variation in terms of the number of categories displayed (i.e., edible, wearable, or driveable objects, etc.). A pre-test assessed whether the clip art used was indeed matched with the verb.

Linguistic Materials All linguistic stimuli were plausible German sentences such as *Der Mann verschüttet gleich das Wasser (the man spills soon the water)*. The sentences were constant across all conditions, while only the simultaneously presented visual contexts differed. The plausibility of all linguistic stimuli was validated offline, prior to the studies.

Figure 1: Example for a visual display as used in **Experiment 2** in all four conditions. **Experiment 1** used a similar type of set up. From left to right and top to bottom: 0, 1, 3 and 4 possible targets, given the sentence *The man spills soon the water in the kitchen*.

Predictions We sought to replicate patterns of anticipatory eye movements towards possible target referents (Kamide, Altmann, & Haywood, 2003). Simultaneously, the ICA assessed processing effort on the verb where anticipatory eye movements typically occur as well as on the referent noun. Anticipatory eye movements may reflect the decision of excluding objects, that do not match the verbal restrictions, from a mental set of potentially upcoming referents – or vice versa, they may indicate which and how many objects still *are* considered to be named next. Thus, a greater reduction in uncertainty about upcoming referent nouns could, for instance, induce increased processing effort (more ICA events) on the verb. If the combination of visual and linguistic information affects surprisal, we expected higher processing effort on the target word in conditions where it was less predictable, i.e. more surprising due to more competitors (i.e. "spillable" objects) displayed and highest effort when no spillable objects was shown.

Methods

40 item sentences were combined with 160 different visual displays and parted into four lists, using latin square design. This way, each participant saw each item in only one condition. Additional 40 filler sentence-display combinations, followed by yes/no comprehension questions (such as: "Did the man spill the lemonade?") were added to keep participants focused. Visual displays were presented from 1000 ms before sentence onset and during the whole sentence in order to avoid (visual) memory effects. ICA events were extracted from the eye-movement data. 32 native speakers of German (all students of Saarland University), aged between 18 to 32 years (M = 24.56) were tested and monetarily reimbursed.

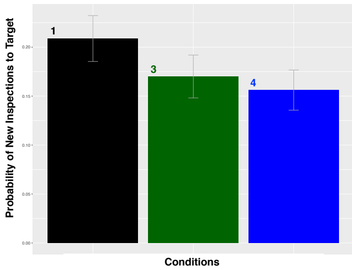

Figure 2: Probability of verb-driven new inspections of target object before target word onset in all possible conditions of **E1**.

Results

Eye-movement Data All analyses were conducted using the R statistical programming environment (RCoreTeam, 2013). For presentation purposes, the overall fixation distribution across an averaged trial in all conditions is plotted in Fig. 3. They reveal an increase in fixations towards objects matching the verb from the onset of the verb onwards (left dashed line) when the visual scene allowed for anticipation of potential target nouns. In conditions **1** and **3**, this indicates a discrimination between those objects who matched the verb and those who did not. No increase in fixations was found when the contexts did not allow for further discrimination (conditions **0** and **4**).

Statistical significance was assessed by analysing differences in new inspections (i.e., the first in a series of fixations towards a region during the time periods of interest) between conditions. That is, for the verb time window, we compared probabilities of verb-driven (i.e., occurring between verb start and article start) attention shifts towards the actual target object. New inspections of the different objects in a visual scene were encoded as a binary dependent variable and were analysed using generalised mixed effects models with a Poisson distribution.

Analogous to the fixation distribution, Fig. 2 shows how new inspections to the target object increase upon hearing the verb in condition **1** ($M = 0.21$, $SD = 0.41$), compared to **3** ($M = 0.17$, $SD = 0.38$) ($\beta = -.221, SE = .099, z = -2.21, p < .05$) and to **4** ($M = 0.16$, $SD = 0.36$) ($\beta = -.293, SE = .099, z = -2.97, p < .01$). In the noun region, comprehenders were significantly more likely to inspect the mentioned object compared to any other object.

Theses results are in line with previous VWP findings (Kamide et al., 2003). We could replicate verb-driven anticipatory eye movements in our set up, even when more than one possible target object was displayed. This hints at more (one matching object) or less (three matching objects) strong anticipation of the target noun. Whether these anticipations alter surprisal and processing effort for either the verb or the target noun, was simultaneously assessed by the ICA.

ICA Event counts, that is, the number of abrupt changes in pupil size within a time window, were analysed. Both time windows for ICA analysis were non-overlapping and 600 ms in length, starting from the middle of the critical word's duration, as previously established for this measure (see e.g., Sekicki & Staudte, 2017, Demberg & Sayeed, 2016). Fig. 4 shows how the number of effort related pupil changes (ICA events) on the verb is similar, while differences between conditions appear on the noun, in reaction to our manipulation. That is, noun processing was facilitated when fewer competitors were displayed, making the noun less surprising.

Since the dependent variables of ICA events obtained within the two time windows were count variables, generalised mixed effects models of Poisson type were used. Contrast coded differences between the conditions (e.g., **0** versus **3** possible target objects displayed) were entered into the model as fixed factors and the maximal converging random structure was included (Barr, Levy, Scheepers, & Tily, 2013).

In line with the plot, analysis of the verb window revealed no significant differences between the conditions. This even holds for the linguistic - visual mismatch condition **0**, suggesting that anticipatory eye movements, although verb-driven, do not elicit differences in surprisal and processing effort on the verb itself [3].

For the noun window, models revealed a significant processing facilitation (fewer ICA events) when three possible targets were shown ($M = 19.37$, $SD = 8.17$), compared to the unhelpful condition **0** ($M = 20.90$, $SD = 8.12$) ($\beta = -.08, SE = .03, z = -2.40, p < .05$). Further, a significant facilitation occured when only one possible target was shown ($M = 17.40$, $SD = 7.79$), compared to three targets ($\beta = -.11, SE = .04, z = -2.57, p < .05$). Differences in processing effort between condition **3** and **4** ($M = 20.13$, $SD = 8.45$) did not reach significance. This shows a direct effect of multimodal information on target word surprisal and, linked to that, on the effort needed to process the noun.

Discussion

We replicated previous findings of verb-driven anticipatory eye movements towards matching objects (Kamide et al., 2003), also in conditions with more than one possible target object. This suggests that comprehenders exploit the visual information in combination with the verb in order to anticipate the noun with more or less certainty.

Although eye movements showed clear patterns of anticipation during the verb, the ICA did not show effects on the word's processing effort. Thus, comprehenders shifted attention towards possible target objects based on the verb information, but possibly refrained from ultimately excluding distractors. Alternatively, a reduction of referential uncertainty (as reflected by the eye movement patterns) does not induce any additional effort. Lastly, the ICA might not be sensitive towards this sort of effort.

Obtained ICA events on the referent noun, however, differed between conditions **0**, **3** and **1**, showing that visual con-

[3] We additionally analysed a 600 ms time window starting from trial onset, in order to examine whether possible effects of grouping of the displayed objects could be seen prior to hearing the verb. No significant differences were found in this region.

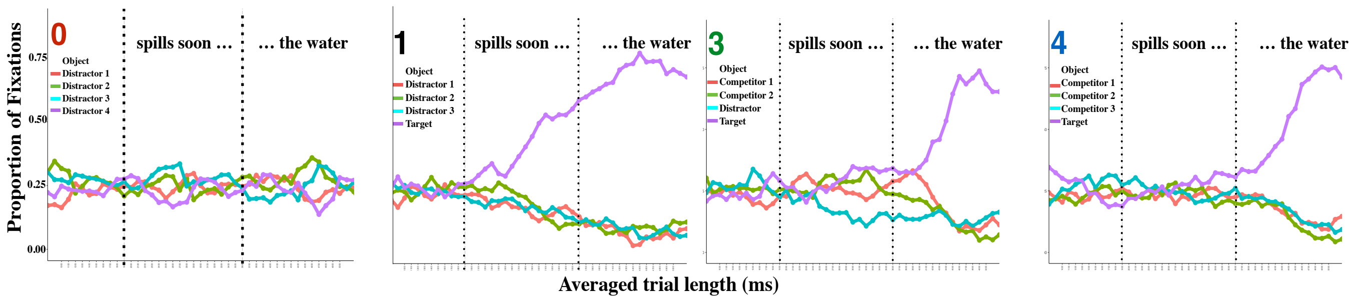

Figure 3: Proportion of fixations across averaged trial length (in 100 ms bins) in all conditions of **Experiment 1**.

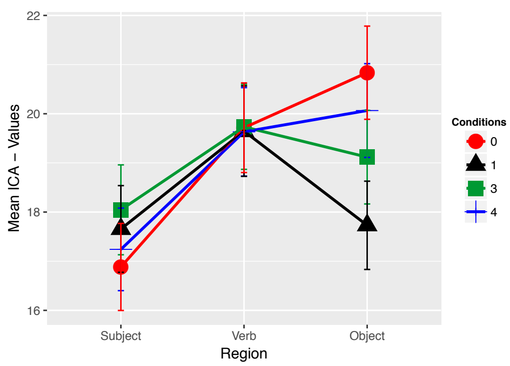

Figure 4: ICA results separated by the experimental conditions. Error bars reflect 95% confidence intervals (CI).

text information directly affected the surprisal and processing effort on the target word. Conditions **3** and **4** did not differ significantly from each other.

These results can be interpreted as being probabilistic, that is, analogous to the probability of a target object in the visual display to correspond to the actual target word coming up: The noun was least surprising and easiest to process if the displayed object would correspond to the noun with 100% certainty (**1**) and more surprising when this correspondence was only 33% certain (**3**). Differences between conditions **3** and **4** not reaching significance might then be attributable to either a lack of power, or three and four being too similar to result in measurable differences. A follow-up EEG experiment (**Experiment 2**) aimed to assess the sensitivity explanation in the verb window and tested whether differences on the noun between **3** and **4** occur in the N400.

Experiment 2 - Multimodal surprisal in the EEG

Experiment 2 used event-related potentials (ERPs), not only to validate the ICA method used in the previous experiment, but also to examine whether any differences in processing effort for the verb occur in a potentially more sensitive method, and to extend current (language-centric) ERP finding regarding surprisal: The ERP component N400 has been associated with surprisal on the word level (Frank et al., 2015) as well as to predictability effects in different modalities. Ganis, Kutas, & Sereno, 1996, for example, presented participants with written sentences either ending with normal words, or with line drawings instead of the word and found that words and drawings can elicit a similar N400. If the *combined* information, gained from integrating language and additional (as opposed to substituting) visual context also affects the N400, we expected similar differences in the target noun's amplitude as previously found in the ICA. If it is more sensitive to effort induced by anticipation or uncertainty reduction – a higher amplitude was expected on the verb as more distractors can be excluded as possible targets.

In order to keep results comparable, **Experiment 2** used the same manipulation and type of stimuli as **Experiment 1**.

Methods

96 stimuli of the same type as described in **Experiment 1** were used along with the same amount of fillers, parted and randomized as before. Conditions and task were identical to the ones used in **Experiment 1**. The only differences were the written presentation of the sentences, and the now quadrangular arrangement of the four pieces of clip art with an added sepia filter, in order to tone down salient colours.

36 right-handed native speakers of German (*M* age: 25,22; Age range: [19, 34]; *SD*: 3.47; Male: 7; Female: 29) with normal or corrected-to-normal vision took part in the ERP experiment. All Participants gave informed consent and were monetarily reimbursed for their participation.

Visual displays were again presented with a 1000 ms preview time, in which participants were allowed to move their eyes around in order to identify and inspect the objects. As soon as the fixation cross appeared for a jittered duration in the center of the display, participants were asked to keep their eyes focused on the phrases presented subsequently. They were further instructed to prevent blinking throughout the sentence. Sentences were presented in phrases and in the centre of the screen with a presentation period of 400 ms and a 100 ms ISI. The visual displays stayed on the screen for the entire trial time. Questions were presented after each trial and either concerned the visual (e.g.*Was the milk on the right?*), or the linguistic content (*Did the man spill the milk?*). Subjects were asked to answer on a button press. The distance between the participant and the screen was always 103 cm in order to keep all of the objects in a 5° visual angle from the center of the screen in order to minimize eye movements throughout the experiment.

Electroencephalographic recording and processing parameters The EEG was recorded by 24 Ag/AgCl scalp electrodes embedded in a cap (acti-CAP, BrainProducts) and amplified with a BrainAmp (Brain-Vision) amplifier. Electrodes were placed according to the 10-20 system (Sharbrough et al., 1995) and referenced online to the reference electrode (FCz) as well as re-referenced offline to the average of both mastoid electrodes. The ground electrode was located at AFz. Horizontal eye movements were monitored via bipolar recording of the electrooculogram (EOG) with an electrode on the outer canthus of the left and right eye. Blinks (Vertical EOG) were monitored with electrodes over the supraorbital and the infraorbital ridge of the left eye, referenced to the left mastoid. Electrode impedances were kept below 5 kΩ. The signal was sampled at 500 Hz, using an anti-aliasing low-pass filter of 250 Hz online during recording. Data were later band pass filtered offline at 0.01-40 Hz (Luck, 2014). All records were semi-automatically examined and marked offline for EOG and other artefactual contamination such as electrode drifts, amplifier blocking and excessive muscle activity. Artefactual trials were excluded with a rejection threshold of 20% per condition for participant rejection, resulting in the exclusion of 8 participants from the analysis.

Results

Single-participant data were averaged for each of the four experimental conditions within 800 ms windows from the onset of the verb and the target noun. The segments were aligned to a 200 ms pre-critical baseline and data sets for both critical time windows were then exported from the averaged ERP data, using BrainVision Analyzer's (Version 2.1) Area-Information export function. The grand average of all participants was then analyzed, time-locked to the onset of the critical words. Analyses were done using the *ez* package in R to perform repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser corrected p-values. In addition, F-values, as well as η^2 (generalised eta-squared, see Bakeman, 2005) values are reported as a measure of effect size. ANOVAs were performed on data sets including all electrodes, with ROIs for frontal (F3, Fz, F4), central (C3, Cz, C4) and posterior (P3, Pz, P4) distributions. We analyzed a typical N400 time window between 300 and 500 ms after onset of the verb and noun. Main effects were assessed by running omnibus ANOVAs with electrode site (frontal/central/parietal) and experimental condition (number of competitors matching the verb) as within factors.

Fig. 5 shows how only the mismatch condition **0** elicited an increased negativity at 400 ms after verb onset. On the noun region, all four conditions elicited a modulated ERP response to the more or less predictable target word. That is, the N400, peaking at 400 ms after onset of the critical word, differed in amplitude between conditions, although the linguistic context never changed and only the visual display varied.

The model assesses statistical significance of these effects, revealing a main effect for condition ($F(3,81) = 8,18, p < 0.05, \eta^2 = 0,06$) on the verb. Follow-up pairwise comparisons showed that significantly larger negativity was elicited by condition **0** ($M = -1.34\,\mu V$), compared to the baseline condition **1** ($M = -0.69\,\mu V$) ($F(1,27) = 8,49, p < 0.05, \eta^2 = 0,06$). Negativity was widespread across frontal, central and parietal regions, while being largest in the latter. However, conditions **3** ($M = -0.75\,\mu V$) and **4** ($M = -0.4\,\mu V$) did not yield significant differences in the N400 component, compared to **1**, suggesting a binary evaluation of whether the visual display matched the verb, rather then more detailed distinctions of the displayed options.

On the noun window, a modulation of the (surprisal) effect was found. Namely, further analysis of the significant main effect of condition ($F(3,81) = 7.74, p < 0.05, \eta^2 = 0.13$) revealed that condition **1**, in which the noun was most predictable, resulted in the lowest N400 amplitude ($M = 0.07\,\mu V$). Conditions **3** ($M = -0.8\,\mu V$) and **4** ($M = -0.79\,\mu V$), where the target noun could be expected with 33% and 25% certainty, resulted in a significantly higher amplitude (three: $F(1,19) = 13.3, p < 0.05, \eta^2 = 0.16$, four: $F(1,19) = 7.03, p < 0.05, \eta^2 = 0.13$). Condition **0** ($M = -1.3\,\mu V$), where none of the clip art items in the visual display could be used to predict the target noun, yielded the highest difference in the N400 amplitude, compared to **1** ($F(1,19) = 18.9, p < 0.05, \eta^2 = 0.26$).

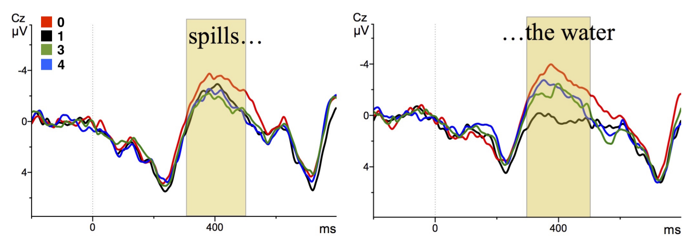

Figure 5: ERP time-locked to the verb (left) and noun (right) onset, separated by the experimental conditions. The reported region is highlighted. The data shows only the Cz electrode (unfiltered).

Discussion

We investigated whether the N400 (1) validated the ICA method, (2) is sensitive towards visually influenced surprisal of target words and, (3) reflects possible differences in processing effort for the verb, where visual uncertainty could be reduced.

Data on the verb additionally revealed an effect in the case of the mismatch condition **0**, compared to all three other conditions, which in **Experiment 1** had not elicited any effect in the ICA. The other three conditions did again not differ from each other.

Results on the noun validated the differences previously found in the ICA and proved that the N400 is also sensitive towards multimodal surprisal. Condition **0** elicited a highly increased N400 amplitude on the noun, although the mismatch between the sentence and the scene had already been detected on the verb. The target noun's amplitude further showed a modulation according to our manipulation, namely it was lowest for condition **1**, where the target was most predictable, higher for **3** and **4**, while both again did not differ from each other, and highest for **0**. We interpret the effects to index participants' detection of a mismatch and a modulated predictability and surprisal of the target noun, acquired by integrating visual and linguistic information. Differences in the target noun's N400 reliably indexed the discrimination between one, many and no objects being possible target options, showing that the N400 is not only sensitive towards linguistic surprisal as shown by Frank et al., 2015, but also to multimodally derived surprisal.

General Discussion

In two experiments, deploying different measures, we sought to examine the effect of visual context on word surprisal (Hale, 2001) as well as on the processing effort for that word.

By manipulating multimodal surprisal of referent nouns while keeping linguistic surprisal constant across conditions, we ensured that any observed effects could be attributed to the influence of visual context specifics on surprisal. Apart from replicating anticipatory eye movements in the VWP **Experiment 1**, we observed that, although anticipation of the target noun did not elicit differences in surprisal and processing effort directly on the verb, it resulted in significant differences on the noun. The N400 in **Experiment 2** did not reveal additional effects of possible uncertainty reduction at the verb (apart from an increased amplitude in the case of the mismatch condition **0**). A reasonable conclusion might be that eye movements in anticipation of the noun(s) reflect comprehender's shift in attention to possible targets but do not lead to a decision for an exclusion of distractors based on the verb constraint. As the the ICA before, the N400 was further found to be sensitive to surprisal when visual and linguistic context information has to be actively combined.

The noun effect found in both studies can be interpreted as being attributable to participants only roughly deciding whether one, more than one, or none of the objects match the verb. Alternatively, results may reflect graded differences with respect to the specifics of the visual context, i.e. the probability profile of the different scene-sentence combinations. In this case, the null result between **3** and **4** could be attributed to a lack of power, or, simply to the fact that the difference between three and four objects was too small to elicit significant effects on processing effort.

In sum, we have demonstrated that the multimodal surprisal of a word – as modulated by the visual referential context – predicts both pupillometric (ICA) and ERP (N400) measures on online processing effort. These findings provide strong evidence that surprisal, and its associated processing effort, is not determined by the linguistic signal alone but rather reflects expectations derived online (at least) from the relevant visual environment in which language is used.

Acknowledgments

The reported research is funded by the Deutsche Forschungsgemeinschaft (DFG) under grant SFB1102 and the "Multimodal Computing and Interaction" Cluster of Excellence at Saarland University.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Bakeman, R. (2005, 01). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

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.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. (2009). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*, 639–645.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11.

Ganis, G., Kutas, M., & Sereno, M. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, *NAACL 01*, 1–8.

Kamide, Y., Altmann, G. T., & Haywood, S. (2003). Prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the 13th conference on empirical methods in natural language processing* (p. 234243). Waikiki, Honolulu.

Luck, S. J. (2014). An introduction to the event-related potential technique.

Marshall, S. (2002). The index of cognitive activity: Measuring cognitive workload. *Proceedings of the 7th conference on Human Factors and Power Plants*, *IEEE*, 75–79.

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

RCoreTeam. (2013). R: A language and environment for statistical computing. *Computer Software Manual. Retrieved from http://www.R-project.org/*.

Sekicki, M., & Staudte, M. (2017). The facilitatory effect of referent gaze on cognitive load in language processing. In *Proceedings of the 39th annual conference of the cognitive science society* (pp. 3107–3112). London, UK.

Shannon, C. (1949). Communication in the presence of noise. *Proceedings of the IRE*, *37*, 10–21.

Sharbrough, F., Chartrian, G. E., Lesser, R. P., Luders, H., Nuwer, M., & Picton, T. W. (1995). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200-202.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Tourtouri, E., Delogu, F., & Crocker, M. (2017). Specificity and entropy reduction in situated referential processing. In *Proceedings of the 39th annual conference of the cognitive science society* (pp. 3356–3361). London, UK.