

## Visual attention-capture cue in depicted scenes fails to modulate online sentence processing

**Juliane Burmester**

JULIANE.BURMESTER@UNI-POTSDAM.DE

*Empirical Childhood Research  
University of Potsdam  
Karl-Liebknecht-Straße 24-25  
14476 Potsdam, Germany*

**Katharina Spalek**

KATHARINA.SPALEK@HU-BERLIN.DE

*Department of German Language and Linguistics  
Humboldt-Universität zu Berlin  
Unter den Linden 6  
10099 Berlin, Germany*

**Isabell Wartenburger**

ISABELL.WARTENBURGER@UNI-POTSDAM.DE

*Cognitive Sciences, Dept. Linguistics  
University of Potsdam  
Karl-Liebknecht-Straße 24-25  
14476 Potsdam, Germany*

**Editor:** Vera Demberg

Submitted 06/2019; Accepted 12/2019; Published online 12/2019

### Abstract

Everyday communication is enriched by the visual environment that listeners concomitantly link to the linguistic input. If and when visual cues are integrated into the mental meaning representation of the communicative setting, is still unclear. In our earlier findings, the integration of linguistic cues (i.e., topic-hood of a discourse referent) reduced discourse updating costs of the mental representation as indicated by reduced sentence-initial processing costs of the non-canonical word order in German. In the present study we aimed to replicate our earlier findings by replacing the linguistic cue by a visual attention-capture cue that directs participants' attention to a depicted referent but is presented below the threshold of conscious perception. While this type of cue has previously been shown to modulate word order preferences in sentence production, we found no effects on sentence comprehension. We discuss possible theory-based reasons for the null effect of the implicit visual cue as well as methodological caveats and issues that should be considered in future research on multimodal meaning integration.

**Keywords:** linguistic vs. visual salience, accessibility, discourse processing, ERP

### 1 Introduction

Everyday communication is multimodal, comprising linguistic as well as extra-linguistic (e.g., visual) information. A growing branch of psycholinguistic research highlights effects of extra-linguistic cues (e.g., eye gaze or gestures) and the visual environment on language processing (e.g., Crocker, Knoeferle, & Mayberry, 2010; Nappa & Arnold, 2014; Spevack, Falandays,

Batzloff, & Spivey, 2018; Staudte, Crocker, Heloir, & Kipp, 2014). By contrast, traditional models of sentence comprehension do not explicitly account for the role of visual attention during the comprehension process (e.g., Bornkessel & Schlesewsky, 2006; Friederici, 2002; Marslen-Wilson & Tyler, 1980). However, discourse models (or situational/mental models) go beyond sentence-level processing. Discourse models propose that during communication, interlocutors build a non-linguistic mental representation of relevant discourse referents and events based on the incoming linguistic and visual perceptual input amongst multiple other factors (e.g., Bower & Morrow, 1990; Gernsbacher, 1991; Grosz & Sidner, 1986; Johnson-Laird, 1980; Van Dijk & Kintsch, 1983; Zwaan, 2004; Zwaan & Radvansky, 1998). Therein, referents of high attentional state are assumed to be mentally represented with a higher degree of mental accessibility and/or a higher activation level compared to referents of low attentional state (e.g., Arnold, 2010; Arnold & Lao, 2015; Givón, 1988; Gundel, Hedberg, & Zacharski, 1993). We will refer to those referents of high attentional state as being more salient. For the present study, we differentiate between *linguistic salience* which is verbally induced by, for instance, subject-hood or topic-hood of a referent vs. *visual salience* which is induced by, for instance, *exogenous* visual cues to a depicted referent. Exogenous visual cues initiate a reflexive attention shift of the addressee to the location of a stimulus. In our study we used the terms *implicit* vs. *explicit* visual cues to enable a more precise distinction of exogenous visual cues, which were used in previous studies (analogous to the distinction by Myachykov, Thompson, Garrod, & Scheepers, 2012, p. 3). Implicit visual cues are presented below the threshold of perception (i.e., subconsciously). Explicit visual cues are presented above the threshold of perception (i.e., consciously) (for an overview about neuronal modulations by stimulus-driven (i.e., sensory cue-based) visual attention mechanisms, see Corbetta & Shulman, 2002). With the present study, we aim to test if visual salience induced by an implicit visual attention-capture cue to a depicted referent impacts online sentence-initial processing in a similar way as it has previously been shown for linguistic salience (Burmester, Spalek, & Wartenburger, 2014). Hence, we raise the underlying question if the accessibility degree of mentally represented discourse referents is affected by this type of implicit visual cue or if this is limited to linguistic cues.

In the linguistic domain, information structure is used to make certain entities of the discourse more salient. For instance, topic or aboutness topic is an information structural concept describing the entity (e.g., a referent) the sentence is about, that is, topic is attributed to that part of information about which the speaker intends to increase the listener’s knowledge (Gundel, 1985; Reinhart, 1981). Hence, topic is ascribed not solely a formal linguistic but also a cognitive concept that activates the listener’s mental representation at the beginning of a sentence (Portner, 2007). In the majority of languages, salient information –in terms of the grammatical subject and/or topic of the sentence– dominantly occupies the sentence-initial position, because subjects and topics own a higher accessibility degree compared to their complements, that is, objects and comments (e.g., Bock & Warren, 1985; Dryer, 2013; Tomlin, 1995). German is a language with a strong subject-first preference (e.g., Hemforth, 1993; Weber & Müller, 2004): The canonical word order in German main clauses is subject-verb-object (SO) (see example sentence (1)). Morphological case marking at the respective noun phrases enables the identification of the grammatical function of subject (via nominative case (NOM)) and object (via accusative case (ACC)) for masculine nouns. (Note that in the example sentences (1) and (2), the nouns “*Wal*” [whale] and “*Hai*” [shark] lack overt case affixes, while the determiners are overtly case marked, which nevertheless allows the unequivocal identification of subject and object.)

(1) SO: Der Wal streichelt den Hai.

[the<sub>[NOM]</sub> whale<sub>[NOM]</sub>]<sub>subject</sub> [strokes]<sub>verb</sub> [the<sub>[ACC]</sub> shark<sub>[ACC]</sub>]<sub>object</sub>.

*The whale strokes the shark.*

(2) OS: Den Hai streichelt der Wal.

[the]<sub>[ACC]</sub> shark<sub>[ACC]</sub>]<sub>object</sub> [strokes]<sub>verb</sub> [the]<sub>[NOM]</sub> whale<sub>[NOM]</sub>]<sub>subject</sub>.

*The shark, the whale strokes.*

Despite the strong subject-first preference in German, information structural characteristics allow reordering of sentential constituents such that the object can precede the subject (see example sentence (2) for a non-canonical object-verb-subject (OS) main clause). However, OS sentences in German are much less frequent than SO sentences (e.g., Bader & Häussler, 2010) and need a suitable context which increases the salience of the sentence-initial object. In our previous work, for instance, linguistic salience in short, fictitious stories of two animals was induced by a topic question (i.e., “*What about the shark?*”), which revealed one of two previously mentioned (i.e., discourse given) referents as the topic of the scene (Burmester et al., 2014). Compared to a neutral cue not indicating topic-hood but a wide focus (i.e., “*What exactly is going on?*”), subsequent online sentence-initial processing of OS sentences is eased. This facilitating impact of linguistic salience (i.e., topic-hood) is reflected in the event-related potentials (ERPs) in the form of a sentence-initial Late Positivity, which is attributed to reduced discourse updating costs (e.g., Schumacher & Hung, 2012).

In 2018 we directly compared linguistic and visual salience cues (Burmester, Saueremann, Spalek, & Wartenburger, 2018): Visual salience induced via an *explicit* gaze-shift of a virtual person to a depicted referent speeds sentence-initial reading times of German SO and OS sentences similar to linguistic salience induced via a topic cue. Hence, the sentence-initial processing ease was evident 1) independent of whether salience was induced linguistically or visually compared to a preceding neutral cue, and 2) independent of whether the salient referent is mentioned as the sentence-initial subject or object (Burmester et al., 2018). This is line with other studies supporting the view that utterance comprehension is facilitated when the speaker’s gaze increases the visual salience of depicted referents (e.g., Hanna & Brennan, 2007; Knoeferle & Kreysa, 2012; Staudte & Crocker, 2011). However, not only speakers’ eye gaze, which provides explicit information about referential intentions (henceforth: intentional information), influences utterance comprehension, but also various other visual salience cues. Staudte et al. (2014) showed that listeners benefit from an explicit (non-gaze) arrow cue (henceforth: attentional information) during utterance comprehension similar to eye gaze. Both the arrow and the gaze cue effectively direct listeners’ visual attention to a depicted object, to finally anticipate this salient object for an upcoming verbal reference (Staudte et al., 2014). Arnold and Lao (2015) showed that another abstract type of visual attentional cue (i.e., a black rectangle with a size of approximately 1.0° x 1.0° of visual angle<sup>1</sup> presented for 200 ms at the target referent’s location) together with the position of the referent in the visual display manipulates listeners’ trial-initial attention in depicted scenes. Still, when listeners interpret a subsequent pronoun, their trial-initial attention only secondarily influences which antecedent they select as the most accessible referent in discourse. Instead, pronoun interpretation is primarily driven by the linguistic cue of sentence-initial mention. Overall, such studies provide evidence that *explicit* visual attentional cues affect sentence- and discourse-level processes, although to a different extent than linguistic cues.

Evidence in favour of the impact of *implicit* visual cues comes from language production studies. Here, implicit similar to explicit visual cues effectively manipulate speakers’ attention to referents in a depicted scene. This manipulation of the speakers’ attention is reflected in the

<sup>1</sup> Note that in order to establish comparability of visual cues published in previous research, we calculated the visual angle of the cues used by Arnold and Lao (2015) and Myachykov et al. (2012) post hoc as these studies did not report the visual angle. Calculations of the visual angle account for the visual cue’s size and distance from participants’ eyes. For instance, for Arnold and Lao (2015) the calculation was based on a screen distance of 650 mm (22 - 34 inches reported), screen size width of 390 mm, screen resolution width of 1280 pixels, and cue size of 38 pixels:  $Visual\ angle = degrees(arctan(((38/2)/(650 * 1280/390)))) * 2 = 1.0^\circ$

sentential structure they choose in picture descriptions: Gleitman, January, Nappa, and Trueswell (2007) used an implicit visual cue by means of a black rectangle with a size of approximately  $0.5^\circ \times 0.5^\circ$  of visual angle presented for about 60 - 75 ms at the location of one of two subsequently depicted referents. Other sentence production studies used explicit visual cues by means of a black arrow (Tomlin, 1995), red dot or referent preview (e.g., Myachykov et al., 2012; Turner & Rommetveit, 1968) followed by the presentation of referents that are performing a simple transitive action. As a result, these implicitly or explicitly cued referents are more salient or accessible<sup>2</sup> than other, uncued referents as reflected in a greater likelihood of salient referents being mentioned sentence-initially as the grammatical subject and/or topic of the sentence (e.g., Arnold, 1998, 2010; Tomlin, 1997). Both cue types even lead to production of otherwise disfavoured linguistic structures (in English). For instance, in cases where the patient of the transitive action is cued, speakers produce the less frequent passive voice with salient referents (i.e., the patient) in sentence-initial position.

While production and/or eye-tracking data indicate shifts in the addressee's attention, ERPs allow us to investigate whether and when during the course of sentence processing increased effort is needed. Numerous ERP studies have provided insights into underlying discourse-level mechanisms elicited by different types of linguistic cues during online sentence processing (e.g., Bornkessel, Schlesewsky, & Friederici, 2003; Burkhardt, 2006; Burmester et al., 2014; Kaan, Dallas, & Barkley, 2007). Based on specific neural correlates, the *Syntax-Discourse Model* (Schumacher & Hung, 2012), as an instance of a neurocognitive account of discourse processing, specifies two temporally distinct processing mechanisms of meaning computation: discourse linking (N400) and discourse updating (Late Positivity). In Burmester et al. (2014), the facilitative impact of the linguistic salience cue (i.e., topic-hood) elicits a reduced Late Positivity around 500 - 700 ms time-locked to the sentence-initial position of OS sentences, but not of SO sentences. In line with the assumptions of the Syntax-Discourse Model, the reduced Late Positivity in the non-canonical OS word order is attributed to reduced processing costs for updating the current discourse model following the linguistic topic cue compared to the neutral cue. This interpretation of the Late Positivity as an index for integration and updating processes of mental representations is further supported by recent studies (e.g., Delogu, Drenhaus, & Crocker, 2018; or within the *neurocomputational model* of language comprehension by Brouwer, Crocker, Venhuizen, & Hoeks, 2017). However, the assumptions of the Syntax-Discourse Model as well as of other discourse models (Hagoort & Van Berkum, 2007) go beyond the impact of purely sentential context on meaning computation, but include situational context information. Even more explicitly, the *Coordinated Interplay Account* (Crocker et al., 2010) highlights the role of visual attention for listeners' mental representations. This account assumes closely temporally synchronized stages of visual and linguistic information processing during sentence comprehension as supported by multiple "visual world" eye-tracking studies (e.g., Knoeferle & Kreysa, 2012) and also ERP studies (e.g., Knoeferle, Habets, Crocker, & Münte, 2007). For instance, *visual* cues reduced online processing costs of OS sentences: Facilitating cues included explicit, intentional, speech-aligned (beat) gesture cues indicating a specific sentence part as salient (Holle et al., 2012), or explicit visual presentations of the visually depicted event of the target sentence (Knoeferle et al., 2007). To the best of our knowledge it has not been reported so

---

<sup>2</sup> Note that in contrast to *mental* accessibility (e.g., Ariel, 1988) which relates to discourse-level processing, these sentence production studies attribute the term accessibility to the lemma/conceptual level which corresponds to the retrieval of a referent's mental representation from memory or "representing potential referents in thought" (Bock & Warren, 1985, p. 47). Accordingly, the lemma of a more accessible referent is earlier retrieved from memory and hence mapped to a more prominent syntactic role (Levelt, 1989). However, we use the term accessibility from both the side of discourse and from the side of sentence production, since for both sides referent-related features such as animacy, linguistic or visual prominence contribute to a referent's high accessibility degree (e.g., Ariel, 1988; Prat-Sala & Branigan, 2000). Disentangling different concepts of accessibility is not within the scope of the present study.

far how implicit visual cues that purely direct the addressee's attention to depicted referents impact online sentence processing.

Using ERPs for investigating the impact of implicit visual cues on sentence comprehension might contribute to our understanding of the underlying neurophysiological mechanisms during sentence processing which might be comparable to those evoked by linguistic cues. Our study aims to answer the question whether –parallel to our earlier findings concerning linguistic salience– a referent in sentence-initial (i.e., topic) position is easier to process if visual salience is induced via an implicit attention-capture cue. Hence, by using an implicit visual cue in the present study we intend to conceptually replicate our earlier ERP-findings, that is, the sentence-initial Late Positivity modulation evoked by an (explicit) linguistic cue (Burmester et al., 2014). The implicit visual cue of the current study was presented for 66 ms analogously to the Gleitman et al. (2007) study in which a similar type of cue significantly manipulated speakers' attention in depicted scenes, and hence, modulated what speakers mentioned first during sentence production. In accordance with the earlier findings concerning linguistic salience, we predict that visual salience of a depicted referent induces modulations of the Late Positivity at sentence-initial position of subsequent OS sentences. Besides the Late Positivity, the linguistic topic cue in Burmester et al. (2014) elicited an early perceptual repetition effect due to word repetition in the topic but not in the neutral condition. This effect was reflected in a reduced early positivity around 200 ms at sentence-initial position of both SO and OS sentences. In the present visual cueing paradigm, no word repetition occurs. Therefore, we do not expect any modulations of this early positivity. In addition, the Burmester et al. (2014) study revealed a word order effect in terms of generally greater processing costs for OS than SO sentences that we expect to replicate in the present study.

## 2 Materials and methods

### 2.1 Participants

Thirty-one native speakers of German participated after giving informed consent. Except for one participant, participants were right-handed as assessed by a German version of the Edinburgh Handedness Inventory (Oldfield, 1971). All had normal or corrected-to-normal vision and had no reported neurological disorder. Participants were reimbursed or received course credits for participation. Data of two participants were excluded from further analysis, that is, one participant due to left handedness, and one participant due to a technical error during recording the electroencephalogram (EEG). The analysed group consisted of 29 participants (15 female, mean age 24.8 years, age range 19.4 - 25.2 years).

### 2.2 Design and material

In the present study (analogous to Burmester et al., 2014) participants were presented with short stories of two animals that were going to perform a fictitious transitive action (e.g., a whale and a shark, one of which is going to stroke the other) while an EEG was recorded to investigate ERPs during online sentence processing. In contrast to Burmester et al. (2014), stories were additionally depicted by pictures of the two animals and the action instrument (cf. Figure 1). The study used a 2 x 2 within-subject design with the fully crossed factors CUE (TOPIC vs. NEUTRAL) and WORD ORDER (SO vs. OS sentences), resulting in four conditions: TOPIC SO, NEUTRAL SO, TOPIC OS, NEUTRAL OS. A total of 160 different stories (40 per condition) was created based on coloured pictures of 40 animals (monomorphemic nouns of masculine gender which were 1-syllabic ( $n = 18$ ) or 2-syllabic ( $n = 22$ )) and 10 actions (which were monomorphemic, 2-syllabic transitive, and accusative-assigning verbs). For 90% of nouns, NOM and ACC case were overtly marked only at the determiner. In the remaining 10 % of nouns, NOM case was overtly marked only at the determiner, but ACC case was overtly marked at the determiner *and* noun (e.g., “*den Löwen*” [the<sub>[ACC]</sub> lion<sub>[ACC]</sub>]<sub>object</sub>). Nouns and verbs were controlled for normalised written lemma

and type frequency values according to the dlex database (Heister, Würzner, Bubenzer et al., 2011). Moreover, other semantic and discourse factors such as animacy and discourse-givenness of sentential arguments impact referent accessibility and hence the ordering principles at the sentence-level (e.g., Clark & Clark, 1977; Grewe, Bornkessel, Zysset et al., 2006). We controlled for these factors by exclusively choosing animate referents that were explicitly mentioned in the lead-in sentence.

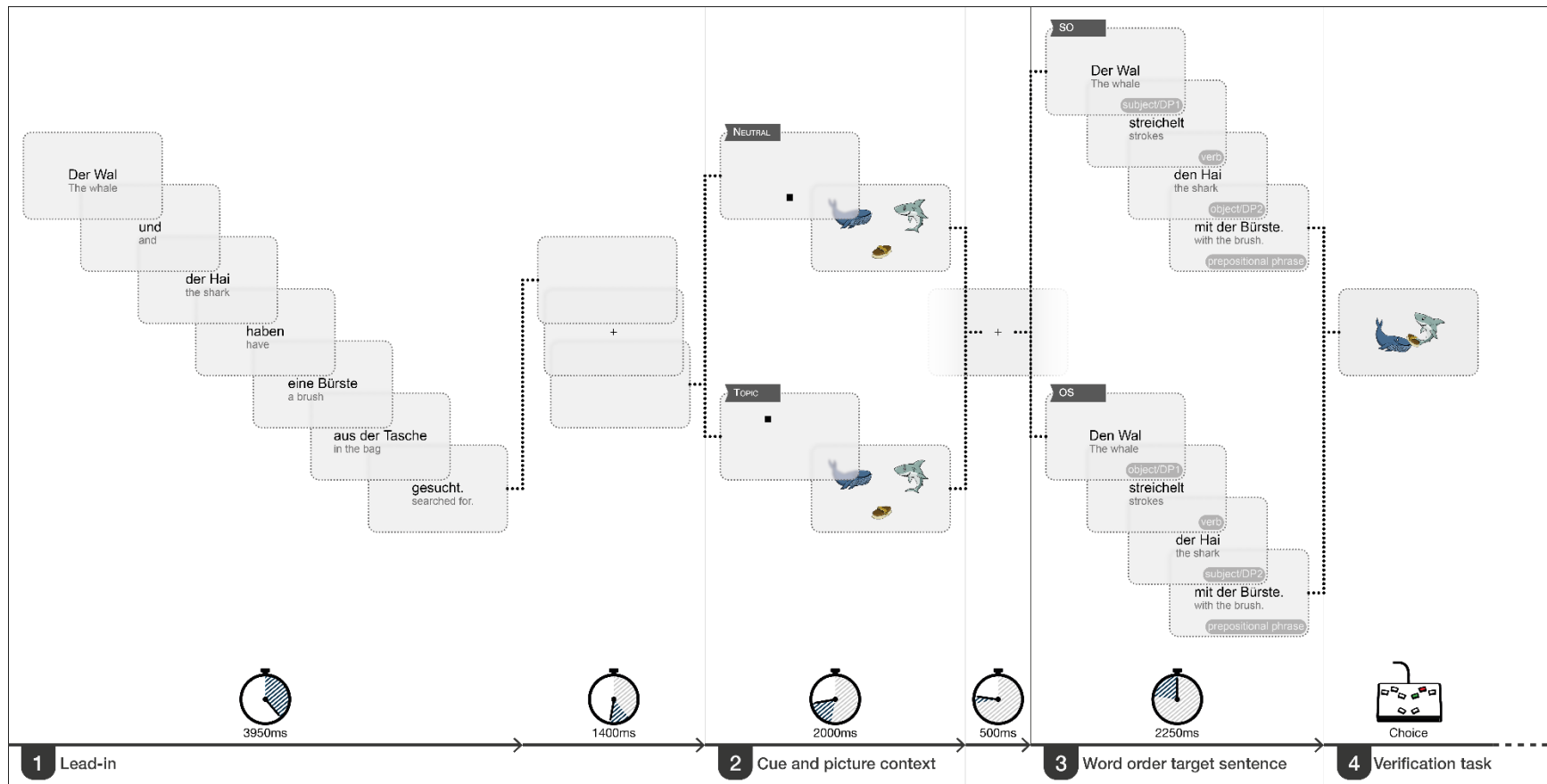
Each trial started with a red fixation cross signalling the beginning of a new story. Afterwards a blank screen for 500 ms was followed by a phrase-wise presented lead-in sentence (see Figure 1 (1)) introducing the two relevant animals of the scene, the action instrument and a corresponding prepositional phrase (e.g., the place where the animals were finding the action instrument). With regard to information structure, the lead-in revealed both animals as discourse-given (Prince, 1981) and the action as inferable based on the mentioned instrument (Prince, 1992). The following visual context (2) consisted of the implicit visual attention-capture cue located at one of three picture positions: (i) the upper left or (ii) upper right animal (i.e., TOPIC CUE, respectively), (iii) the bottom centre position of the action instrument (i.e., NEUTRAL CUE). The visual cue was presented in the form of a black square with approximately  $0.3^\circ \times 0.3^\circ$  of visual angle presented against a light greyish background colour for a duration of 66 ms. The cue was immediately followed by the pictures at the respective positions similar to previous visual cueing paradigms (Gleitman et al., 2007; Myachykov et al., 2012). Hence, either one of the two animals was cued in order to direct participants' attention to the topic referent (TOPIC CUE), or the action instrument was cued in order to direct participants' attention to a wider scope of the scene, the to-be-performed action of the two animals (NEUTRAL CUE). The coloured pictures of the animals and actions were presented with  $9.3^\circ \times 9.3^\circ$  of visual angle. After a black fixation cross, the target sentence (3) was presented phrase-wise either in SO or OS WORD ORDER describing the thematic role relations of the depicted animals (i.e., who is performing the action with whom), followed by a blank screen for 200 ms. The target sentence consisted of a first determiner phrase (DP1), verb, second determiner phrase (DP2), and a prepositional phrase specifying the animals' location or the action instrument. DP1 was either both subject and agent of the action or object and patient/undergoer of the action. DP2 always carried the inverse syntactic and thematic role of DP1. Note here that in our study syntactic and thematic role always coincided. With the closing prepositional phrase (e.g., "with the brush") we aimed to prevent that processing of DP2 is contaminated by "wrap up" effects typically occurring at the end of sentences (e.g., Just & Carpenter, 1980). The phrase-wise presentation durations were chosen in analogy to previous studies (Bornkessel et al., 2003): DPs and prepositional phrases presented for 500 ms, respectively; conjunctions, auxiliary verbs, and main verbs presented for 450 ms, respectively; with a 100 ms interstimulus interval. In 20 % of trials a sentence-picture-verification task (4) probed participants' attentive reading of the stories. For this, 32 pictures (eight per condition) depicting the content of the preceding target sentence were created -half with correct, half with exchanged (i.e., incorrect) thematic role assignments (e.g., *shark stroking whale* vs. *whale stroking shark*). Pictures of the sentence-picture-verification were presented for 2 s before participants had to press the corresponding button within a 2 s time window. The verification task was followed by a blank screen for 500 ms. The experimental items were identical to the ones in Burmester et al. (2014) except that the lead-in sentence was not presented in a self-paced-reading manner but automatically and that the linguistic context question was replaced by the presentation of the visual context.

Due to the fictive character of the stories as in children's books, both animals could be plausible agents or patients of the action. In the visual context, animals of a story were always facing each other. In the target sentence, animals equally often occurred as the agent or patient of an action. Animals were distributed equally across conditions and were always performing the action with a different animal. Introducing the animal first or second in the lead-in sentence as well as the presentation of the animals on the left or right side of the screen was counterbalanced

across conditions. To avoid possible effects of structural priming (e.g., Scheepers & Crocker, 2004), trials were presented in pseudo-randomized order with maximally two consecutive trials of the same condition and word order in the target sentence. Preferences of thematic role assignment or topic continuity due to preceding trials were minimized by at least five intermediate trials before an animal was repeated. There were four lists of 160 trials each. Lists were created such that within each list each item (i.e., animal pair and action) occurred once and across the four lists each item occurred in each condition. We presented each participant with one of these lists of 160 trials. These lists did not include any filler trials to arrange the experimental session in an appropriate time frame for participants' motivation and concentration ability, and hence, minimize artefacts and alpha waves in the EEG signal.

### **2.3 Procedure**

Participants were tested individually, seated in a sound-attenuated booth with 80 cm distance to a computer screen (1680 x 1050 pixels screen resolution). After the preparation for EEG recording, participants were visually presented (on screen) with all pictures used in the subsequent experiment with their corresponding word forms to become familiar with the pictures, that is, the 40 animals and 10 actions. Afterwards participants received a written instruction in which they were asked to read each story attentively and silently and to answer the sentence-picture-verification task after some of the stories as accurately and fast as possible. Participants were asked to sit relaxed, and to avoid eye-movements, blinks, and other muscle movements. Participants had a button box (Cedrus® response pad model RB-830) on their lap and performed three practice trials to become familiar with the procedure. To answer the sentence-picture-verification task, the green and red response button (according to correct vs. incorrect pictures) were assigned to the right fore and middle finger (which was counterbalanced across participants). Participants were instructed that they will be presented with a new story as soon as they see a red fixation cross and they press the yellow button of the button box on which they should put their left thumb throughout the whole experiment. The experiment was visually presented by means of the Presentation® software (version 14.1; www.neurobs.com). The whole experiment included pauses after each 40 trials and lasted approximately 30 minutes. In a post-experiment questionnaire, participants were asked if they have an idea about the purpose of the study and if they noticed anything in the course of the experiment, for instance, any cues or disturbances during picture presentation.



**Figure 1.** Experimental design of sample trial with approximate English translation written in grey. (1) Participants read the phrase-wise presented lead-in sentence followed by a blank screen (200 ms), fixation cross (1000 ms), and blank screen (200 ms). (2) The implicit visual cue presented for 66 ms (TOPIC vs. NEUTRAL CUE) was directly followed by the pictures. After another fixation cross, (3) the SO or OS WORD ORDER target sentence was presented phrase-wise. In the TOPIC cue condition, the SO or OS sentence mentioned the cued referent (i.e., whale) first (i.e., as DP1). (4) In 20 % of trials participants had to answer a sentence-picture-verification task afterwards. *Abbreviations:* SO = subject-verb-object, OS = object-verb-subject, DP = determiner phrase.



## 2.4 EEG recording

The EEG was recorded using a 32 channel active electrode system (Brain Products, Gilching, Germany) with a sampling rate of 1000 Hz. The electrodes fixed at scalp by means of a soft cap included the following 29 scalp sites according to the international 10-20 system (American Electroencephalographic Society, 2006): F7/8, F5/6, F3/4, FC3/4, C5/6, C3/4, CP5/6, P3/4, P7/8, PO3/4, FPz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz. The electrooculogram (EOG) was monitored by electrodes above (position FP2) and below the right eye. The ground electrode was placed at FP1. Impedances were kept below 5 kOhm. The left mastoid served as the reference electrode online, whereas recording was re-referenced to bilateral mastoids offline.

## 2.5 ERP data analysis

For ERP data analysis, the Brain Vision Analyzer software (version 2.1, Brain Products Gilching, Germany) was used. To exclude slow signal drifts and muscle artifacts from the EEG raw data the Butterworth zero phase filter (low cutoff: 0.3 Hz; high cutoff: 70 Hz; slope: 12 dB/oct) was applied additional to the notch filter of 50 Hz. For the correction of artifacts caused by vertical eye movements the algorithm by Gratton, Coles, and Donchin (1983) was applied. We applied an automatic artifact rejection to reject blinks and drifts in the time window of -200 to 2150 ms relative to the onset of the target sentence as well as -200 to 500 ms relative to onset of the visual cue (rejection criteria: max. voltage step of 30  $\mu$ V/ms, max. 200  $\mu$ V difference of values in intervals, lowest activity of 0.5  $\mu$ V in intervals). On average 1.64 % of trials was rejected. ERPs were averaged for each participant and each condition within a 2150 ms time window time-locked to the onset of the target sentence and within a 500 ms time window time-locked to the onset of the visual cue, with a 200 ms pre-stimulus onset baseline, respectively.

For the statistical analysis, IBM SPSS Statistics (version 25.0) was used. The chosen parameters for ERP analysis were identical to the ones used by Burmester et al. (2014) to maintain comparability with the ERP results on the impact of linguistic context information on online sentence processing, in which the same sentences were presented to participants. Hence, based on previous psycholinguistic research, we analysed language-related ERP components of the target sentence in the following time windows time-locked to the onset of DP1, verb, and DP2, respectively: 100 - 300 ms (P200), 300 - 500 ms (N400), 500 - 700 ms (Late Positivity). Via computation of mean amplitudes of three electrodes, respectively, nine regions of interest (ROIs, which were identical to the ones in Burmester et al., 2014) entered the statistical ERP analysis as the fixed factor ROI: LEFT FRONTAL (F7, F5, F3), LEFT FRONTO-CENTRAL (FC3, C5, C3), LEFT CENTRO-PARIETAL (CP5, P3, P7), RIGHT FRONTAL (F8, F6, F4), RIGHT FRONTO-CENTRAL (FC4, C6, C4), RIGHT CENTRO-PARIETAL (CP6, P4, P8), FRONTAL-MIDLINE (FPz, AFz, Fz), CENTRAL MIDLINE (FCz, Cz, CPz), PARIETAL MIDLINE (Pz, POz, Oz). For statistical ERP analysis, mean amplitude values of ERPs within each condition were analysed following a hierarchical schema (e.g., Bornkessel et al., 2003; Burmester et al., 2014). Firstly, we computed a fully crossed repeated measures analysis of variance (ANOVA) with the fixed factors CUE (TOPIC vs. NEUTRAL), WORD ORDER (SO, OS), and ROI (nine levels, see above) for each of the three time windows time-locked to the onset of DP1, verb, and DP2, respectively.

In addition to the analyses of the target sentence, we analysed early ERP components (i.e., N1, P2) relative to the onset of the visual cue, henceforth termed CUE POSITION (i.e., TOPIC LEFT, TOPIC RIGHT, and NEUTRAL BOTTOM), in the time window of 100 - 200 ms as well as of 250 - 350 ms (see e.g., Luck & Hillyard, 1994; Mangun, 1995 for attention-based early visual processing changes reflected in different early evoked potentials). Notably, the pictures followed the cue immediately, hence these time windows started 100 ms and 250 ms after cue onset but also 34 ms and 184 ms after picture onset.

We report Greenhouse and Geisser (1959) corrected *F*- and *p*-values, the original degrees of freedom (*df*) in brackets, and the Greenhouse and Geisser epsilon ( $\epsilon$ ) factor for non-sphericity

adjustments of the original  $df$  according to Jennings & Wood (1976) (only for  $F$ -tests with more than one  $df$  in the numerator). Statistically significant effects (i.e.,  $p < .05$ ) involving an interaction with ROI were resolved by computing post hoc paired  $t$ -tests to reveal the topographical distribution of the effect. We controlled for the Type I error due to multiple pairwise  $t$ -tests of levels of the fixed effects in the nine ROIs by adjusting the significance level according to the Bonferroni correction. Thus, for post hoc  $t$ -tests the following Bonferroni adjusted  $p$ -values (two-tailed) were considered as statistically significant at  $\alpha = .05$ :  $p < .006$  to resolve the WORD ORDER x ROI interaction, and  $p < .002$  to resolve the CUE POSITION x ROI interaction. For presentation purposes the displayed ERPs in Figure 2, 3, and B-1 in Appendix B are 10 Hz low-pass filtered.

## 2.6 Behavioural data analysis

For the statistical analysis of the response accuracy of the sentence-picture-verification task, logit mixed models fitted by the Laplace approximation were calculated using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) provided by the *R* environment (version 3.5.1, R Core Team, 2013). To analyse the binary distributed response accuracy data (correct vs. incorrect) with the logit mixed models, CUE, WORD ORDER, and the interaction of both were defined as fixed effects, and Participants and Items were defined as random effects. Fixed effects were coded as  $+/- .5$  to resemble the contrast coding of traditional ANOVA analyses. Model fitting started with the simple model (i.e., the two fixed effects and their interaction, and Participants and Items as random intercepts). In a step-wise manner, slope-adjustments were included if they significantly improved the explanatory power of the simpler model without that slope adjustment as revealed by log-likelihood tests (e.g., Baayen, 2008). The statistics of the fixed effects of the final models are reported with estimates ( $b$ ), standard errors ( $SE$ ),  $z$ - and  $p$ -values.

## 3 Results

As reported in the post-experiment questionnaire, participants did not notice any manipulation of their visual attention suggesting that the cue was not consciously perceived and hence, truly implicit. In this Section, we first describe the ERP results with respect to initial and subsequent processing of the target sentence, before reporting the results of an additional analysis of the present ERP data following the visual cue together with the published data following the linguistic cue (Burmester et al., 2014). Secondly, we present the ERP results with respect to the onset of the visual cue. Thirdly, we report the behavioural results of the probe sentence-picture-verification task.

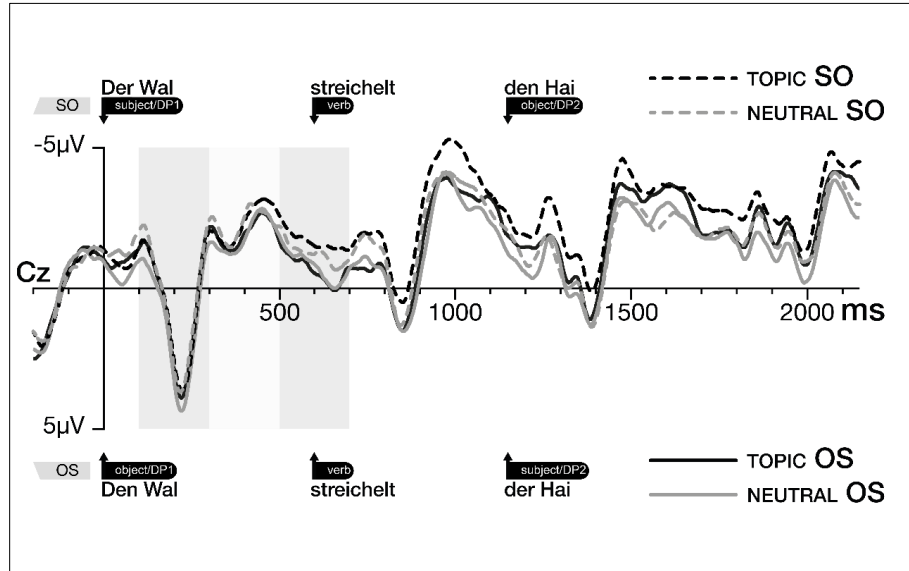
### 3.1 ERP results of the target sentence

Figure 2 illustrates the grand average ERPs at one representative electrode time-locked to the onset of the target sentence (i.e., DP1) followed by the subsequent sentence positions (i.e., verb and DP2) for both CUES (TOPIC and NEUTRAL) and both WORD ORDERS (SO and OS). For grand average ERPs at selected electrodes of each ROI, see Figure B-1 in Appendix B illustrating the CUES *within* each WORD ORDER.

#### 3.1.1 ERP results of sentence processing following the implicit visual cue

For sentence-initial processing, statistical analyses in the time windows of 100 - 300 ms and 300 - 500 ms time-locked to the onset of DP1 neither revealed any statistically significant main effects of CUE (TOPIC vs. NEUTRAL) or WORD ORDER (SO vs. OS), nor significant interactions of CUE, WORD ORDER and/or ROI [ $p > .1$ ] (see Appendix A for the complete statistical output). The analysis in the following time window of 500 - 700 ms revealed a statistically significant main effect of WORD ORDER [ $F(1, 28) = 6.254, p = .019$ ] and a significant interaction of WORD ORDER x ROI [ $F(8, 224) = 3.004, p = .029, \epsilon = 0.424$ ], but no statistically significant effects or

interactions of the factor CUE. Separate post hoc analyses for SO and OS sentences (averaged across the cue conditions) within each ROI yielded a statistically significant enhanced positive-going ERP for OS sentences compared to SO sentences in the LEFT CENTRAL ROI [ $t(28) = -3.605$ ,  $p = .001$ ] and in the MIDLINE CENTRAL ROI [ $t(28) = -3.369$ ,  $p = .002$ ].



**Figure 2.** Grand average ERPs (with baseline correction) at one representative electrode of the MIDLINE CENTRAL ROI for the factors CUE (TOPIC vs. NEUTRAL) and WORD ORDER (SO vs. OS) time-locked to the onset of the target sentence, that is, the first Determiner phrase (DP1) followed by the verb and second Determiner phrase (DP2): TOPIC SO [dotted black] vs. NEUTRAL SO [dotted grey], TOPIC OS [solid black] vs. NEUTRAL OS [solid grey]. Negativity is plotted upwards. Grey shades indicate the three time windows of DP1.

For subsequent sentence positions, the statistical ERP analysis for the time windows 100 - 300 ms, 300 - 500 ms, and 500 - 700 ms post verb onset neither revealed a statistically significant main effect nor interaction involving the factor CUE, but revealed statistically significant effects involving the factor WORD ORDER: In the time window 100 - 300 ms post verb onset a significant main effect of WORD ORDER [ $F(1, 28) = 10.856$ ,  $p = .003$ ] and a significant interaction of ROI x WORD ORDER [ $F(8, 224) = 3.359$ ,  $p = .013$ ,  $\epsilon = 0.485$ ] was reflected in an enhanced positive-going ERP for OS compared to SO sentences in the LEFT CENTRAL [ $t(28) = -3.356$ ,  $p = .002$ ] and LEFT POSTERIOR [ $t(28) = -4.507$ ,  $p < .001$ ] ROI, the RIGHT POSTERIOR ROI [ $t(28) = -3.090$ ,  $p = .004$ ], and the MIDLINE CENTRAL [ $t(28) = -3.854$ ,  $p = .001$ ] and MIDLINE POSTERIOR [ $t(28) = -4.192$ ,  $p < .001$ ] ROI. In the following time window of 300 - 500 ms post verb onset a significant interaction of ROI x WORD ORDER [ $F(8, 224) = 8.101$ ,  $p < .001$ ,  $\epsilon = 0.447$ ] was similarly reflected in form of a significantly enhanced positive ERP for OS compared to SO sentences in the LEFT POSTERIOR ROI [ $t(28) = -3.448$ ,  $p = .002$ ] and in the MIDLINE POSTERIOR ROI [ $t(28) = -4.602$ ,  $p < .001$ ]. Furthermore, in the time window of 500 - 700 ms post verb onset the significant interaction of ROI x WORD ORDER [ $F(8, 224) = 6.157$ ,  $p < .001$ ,  $\epsilon = 0.558$ ] was reflected in an enhanced positive ERP for OS compared to SO sentences in the MIDLINE POSTERIOR ROI [ $t(28) = -3.114$ ,  $p = .004$ ]. Statistical analyses of ERPs post onset of DP2 neither showed any statistically significant main effects, nor interactions of CUE or WORD ORDER in any of the calculated time windows (i.e., 100 - 300 ms, 300 - 500 ms, and 500 - 700 ms) [ $p < .1$ ].

In summary, the ERP results of all three sentence positions (i.e., DP1, verb, DP2) did not show any statistically significant modulation by the preceding visual cue. An impact of the

varying word order with enhanced positive-going ERPs for OS compared to SO sentences was evident in multiple time windows time-locked to the onset of DP1 (i.e., 500 - 700 ms) and verb (i.e., 100 - 300 ms, 300 - 500 ms, and 500 - 700 ms).

### 3.1.2 ERP results compared to sentence-initial processing of the linguistic cue (Burmester et al., 2014)

Since the identical sentence material was used in the present study with the visual cue as in the study with the linguistic cue (Burmester et al., 2014), we aimed at directly comparing the impact of the visual vs. linguistic cue *modality* on sentence-initial processing. For this purpose, we computed additional comparisons of the published ERP data following the linguistic cue with the ERP data following the visual cue by adding the between-subject factor MODALITY (VISUAL vs. LINGUISTIC) while the within-subject-factors CUE, WORD ORDER, and ROI maintained as in the original analysis (reported in Section 2.5). The results as summarised in Appendix C show the following statistically significant effects of MODALITY: In the 100 - 300 ms time window, the analysis showed a significant interaction of MODALITY x ROI [ $F(1, 46) = 3.190, p = .038$ ] and of MODALITY x WORD ORDER [ $F(1, 46) = 5.793, p = .020$ ], which was reflected in a marginally significantly different processing of SO vs. OS sentences following the linguistic cue [ $t(18) = 2.092, p = .051$ ], which was absent following the visual cue [ $t(28) = -4.16, p = .680$ ]. In line with the separate analyses, there were no statistically significant effects of MODALITY in the 300 - 500 ms time window. For the 500 - 700 ms time window post onset of DP1, the joint analysis revealed a marginally statistically significant interaction of MODALITY x CUE x WORD ORDER [ $F(1, 46) = 3.963, p = .052$ ]. This confirms the results of the two separate analyses of the linguistic and visual cue: the presence of a statistically significant interaction of CUE x WORD ORDER x ROI [ $F(8, 144) = 4.15, p < .05$ ] following the linguistic cue and its absence following the visual cue [ $F(8, 224) = 1.605, p = 0.189; \epsilon = 0.417$ ] (cf. Figure 2 and B-1 in Appendix B) in the present study and Figure 2 and Table 3 in Burmester et al., 2014).

In summary, the visual cue had no impact on sentence processing in the present study. Therefore, the visual cue was not effectively increasing the salience of the cued referent (i.e., TOPIC CUE). To examine whether the visual cue per se modulated participants' processing, we computed a further ERP analysis time-locked to the onset of the visual cue. If yes, this should be reflected in differential, especially early sensory-evoked potentials in dependence of the CUE POSITION on screen (i.e., TOPIC LEFT, TOPIC RIGHT, and NEUTRAL BOTTOM).

### 3.2 ERP results of the visual cue per se

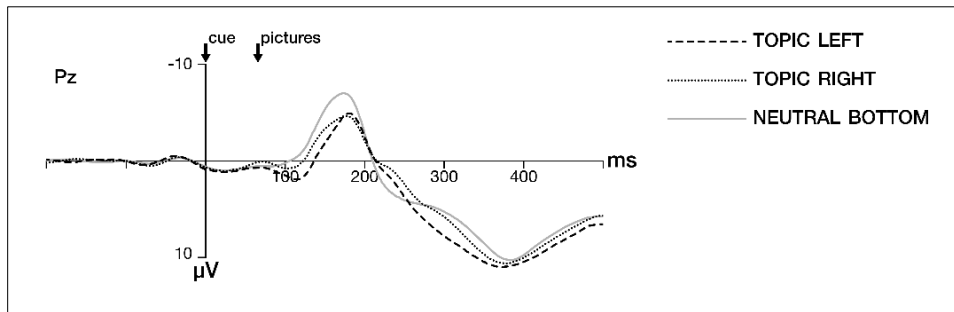
With the following ERP analysis we aimed to test if the implicit visual cue modulated participants' early sensory-evoked potentials (i.e., N1, P2) time-locked to the onset of the visual cue. The visual cue was presented for 66 ms and was directly followed by the pictures. However, since the timing and the position of the pictures were always the same, early processing differences are likely to be related to the different prior CUE POSITIONS. Therefore, ERP analyses were calculated to assess the impact of CUE POSITION (i.e., TOPIC LEFT, TOPIC RIGHT, and NEUTRAL BOTTOM) in the N1 (100 to 200 ms) and P2 (250 to 350 ms) time window post onset of the visual cue and its topographical distribution by the factor ROI.<sup>3</sup>

As can be seen in Figure 3, the ERPs time-locked to the visual cue show an early negativity around 100 - 200 ms followed by a positive deflection around 250 - 350 ms. Statistical analysis in the time window of 100 - 200 ms revealed a statistically significant difference in the form of a

<sup>3</sup> Note that for the ANOVA and post hoc *t*-tests reported in Appendix D, we report the statistical analysis based on the grand average ERPs of all available trials (i.e., 40 trials each for TOPIC LEFT vs. TOPIC RIGHT cue positions and 80 trials for the NEUTRAL BOTTOM cue position). However, conducting the analysis with a similar number of trials across conditions with random sampling of half of the NEUTRAL BOTTOM cue position trials (i.e., 40), revealed a similar pattern of results.

main effect of CUE POSITION [ $F(2, 56) = 9.477, p < .001, \epsilon = 0.966$ ] and of an interaction of CUE POSITION x ROI [ $F(16, 448) = 7.337, p < .001, \epsilon = 0.351$ ]. Post hoc comparisons show that the negativity around 100 - 200 ms was strongest for the cue at the NEUTRAL BOTTOM position compared to the TOPIC LEFT and TOPIC RIGHT position. This difference was present in multiple ROIs (see Appendix D for the complete post hoc statistical results of the respective ROIs). Regarding the positivity around 250 - 350 ms, statistical analyses revealed a statistically significant main effect of CUE POSITION [ $F(2, 56) = 7.628, p < .001, \epsilon = 0.949$ ] and a statistically significant interaction of CUE POSITION x ROI [ $F(16, 448) = 3.675, p = .007, \epsilon = 0.255$ ]. Post hoc analyses show that the TOPIC LEFT cue elicited a significantly more pronounced positive deflection compared to the NEUTRAL BOTTOM cue ( $p < .001$ ). Note, we cannot clearly disentangle the response to the cue and the response to the picture, as the pictures were always presented 66 ms after the cue. The modulation by CUE POSITION might therefore either reflect the direct effect of the cues themselves or their impact on the processing of the subsequently presented pictures.

In both cases, the results indicate that participants processed the implicit visual cues. But still, given that none of the participants noticed the presence of the cues nor was sentence processing influenced by the cue, we can assume, that the cues were processed only subconsciously.



**Figure 3.** Grand average ERPs of one representative electrode for the factor CUE POSITION (TOPIC LEFT, TOPIC RIGHT, NEUTRAL BOTTOM) time-locked to the onset of the implicit visual cue (~66 ms cue duration directly followed by pictures). Negativity is plotted upwards.

### 3.3 Behavioural results

In the sentence-picture-verification task (in 20 % of trials) participants showed a high response accuracy across conditions indicating that participants were attentive throughout the experiment: TOPIC SO:  $M = 0.91$  ( $SE = 0.03$ ), NEUTRAL SO:  $M = 0.91$  ( $SE = 0.02$ ), TOPIC OS:  $M = 0.82$  ( $SE = 0.03$ ), NEUTRAL OS:  $M = 0.81$  ( $SE = 0.02$ ). Logit mixed models analyses of participants' response accuracy did not reveal any statistically significant differences of the fixed effects CUE ( $b = 0.112, SE = 0.222, z = 0.503, p > .1$ ), WORD ORDER ( $b = 0.490, SE = 0.339, z = 1.443, p > .1$ ), or the interaction of CUE x WORD ORDER ( $b = 0.226, SE = 0.222, z = 1.020, p > .1$ ).

## 4 Discussion

We aimed at answering the question if an implicit visual cue to a depicted referent impacts sentence-initial processing similarly to a purely linguistic cue as revealed by ERPs (Burmester et al., 2014). With regard to the linguistic cue, the indication of the aboutness topic referent, which was subsequently mentioned in sentence-initial position, reduced the Late Positivity during online processing of OS sentences. In the present study, the linguistic topic cue was replaced by an implicit visual cue to a depicted referent in a visual scene. However, the impact of the linguistic cue on sentence-initial processing was not replicated. In fact, none of our analyses revealed any

statistically significant effects of the visual cue on sentence processing. The findings of the linguistic cue (Burmester et al., 2014) were interpreted within the Syntax-Discourse Model (Schumacher, 2014) which highlights the role of context information for meaning computation during sentence comprehension. Within this model, the impact of context information (i.e., including sentential and situational context, amongst others), is reflected in modulations of processing costs for discourse linking and updating the hitherto built mental representation of the listener. Following this model, the impact of a referent's topic-hood of the sentence-initial referent in OS sentences was attributed to reduced discourse updating costs as –in contrast to a neutral cue– the linguistic topic cue indicated one referent more salient amongst others which rendered this referent more likely to be mentioned sentence-initially. Obviously, the visual topic cue used in the present study did not increase the salience of the cued referent to a similar extent and hence did not elicit a facilitative effect on sentence processing in terms of reduced discourse updating costs.

However, the long-time neglected role of extra-linguistic (e.g., visual) information in traditional accounts of sentence comprehension (e.g., Frazier & Fodor, 1978; Friederici, 2002; Marslen-Wilson & Tyler, 1980) has been complemented by recent models underlining the close temporal integration of multimodal (e.g., visual *and* linguistic) information in the listener's current mental representation (e.g., Bower & Morrow, 1990; Zwaan, 2004). For instance, as supported by neuroscientific methods, a one-step model of sentence comprehension –integrating concomitant information from different modalities within one step– has been suggested (Hagoort & Van Berkum, 2007). More specifically, this model postulates that linguistic (for instance, sentential structure, semantics) *and* pragmatic information (from prior discourse, extra-linguistic information such as speaker's gestures or the visual world) is immediately processed by the same brain regions (namely the left inferior frontal gyrus) in order to directly map all information onto a discourse model as the basis for sentence interpretation. Similarly, the *Coordinated Interplay Account* by Crocker et al. (2010) explicitly outlines the integration of visual scene information during sentence processing, especially in cases where visual information becomes highly relevant for the interpretation and disambiguation of spoken sentences. By linking to the so called “blank screen paradigm” in which attention shifts during sentence comprehension even occurred when depicted objects were no longer presented (Altmann, 2004), Crocker et al. (2010) suggest that mental representations of a previously presented scene still influence the comprehension process. However, the weighting of (competing) visual and linguistic information and the role of specific inherent features of visual salience cues on the strength of impact during language comprehension is not specified in any of these models.

In the following, we will discuss our findings concerning the absent visual cue impact on sentence processing against the background of previous research using different visual cues while raising possible issues with respect to the present study design. Afterwards, we will briefly discuss the word order effect, which we replicated (Section 4.2).

#### **4.1 Null effect of visual cue on sentence-initial processing**

Against the background of previous research investigating the interaction of visual cues with linguistic processing, we discuss the absent cue effect of the present study 1) with respect to different aspects of informativity of the visual scene for the comprehension process, and 2) with respect to the different impact of visual cue types (implicit vs. explicit) on sentence comprehension and production.

With respect to the informativity of the visual scene, some previous studies emphasise the need for relevance of visual information for meaning computation during language processing. For instance, listeners use depicted events –similarly to linguistic cues (i.e., case marking)– for syntactic reanalyses of locally structurally ambiguous German sentences as reflected in a reduced P600 (Knoeferle et al., 2007). Further evidence from the field of referential processing shows that visual information impacts accessibility of discourse referents only if the linguistic context is

moderate, uninformative, or ambiguous (e.g., Nappa, Wessel, McEldoon et al., 2009; Vogels, Krahmer, & Maes, 2013). With respect to the present study design, the visual scene might not have added crucial information to the comprehension process, for instance, to assign thematic roles in the subsequent target sentence. Compared to visual scenes used in production studies (e.g., Gleitman et al., 2007; Myachykov et al., 2012), visual scenes in the present study did *not* depict thematic role relations for the following reasons: We aimed at minimizing confounding effects of prominence-related factors known to affect participants' gaze fixations of depicted transitive events (e.g., agent-directed fixations followed by patient-directed fixations, Ganushchak, Konopka, & Chen, 2017) as well as linear ordering preferences of sentential constituents (e.g., agent precede patient theta roles, Jackendoff, 1972). In short, we tested how a single implicit cue indicating the subsequent sentence topic would affect referent accessibility and therefore we eliminated all additional factors that could have masked this single cue. Indeed, previous research shows that the more (additional) information is conveyed by the prior discourse –by the linguistic context or by depicted scenes– the more predictable are specific upcoming words as reflected in an immediate ease of sentence processing (e.g., Burmester et al., 2018; Otten & Van Berkum, 2008; Van Berkum, Brown, Zwitserlood et al., 2005). More specifically, in an experimental design rather similar to the one in the present study, we could show that depicted thematic role information of the salient referent boosted the cue-based ease of sentence processing compared to non-predictable thematic role information (Burmester et al., 2018). Moreover, for linguistic context information, Otten and van Berkum (2008) found greater priming effects of the exact discourse message than of the mere word primes. Drawing the parallel to our earlier findings, this could speak in favour of predictive processing mechanisms following the linguistic cue explicitly indicating the upcoming sentence-initial topic (Burmester et al., 2014), while with the visual cue we rather manipulated accessibility in a way similar to a word prime. Hence, maybe further depicted information such as thematic role relations would have activated further semantic features in order to constrain a more precise and coherent discourse context, and finally support predictive processing.

Moreover, with respect to the informativity of the visual scene, the simultaneous visual presence of multiple referents has been argued to reduce referent accessibility (as revealed by reduced pronoun use in, for instance, Arnold & Griffin, 2007 or Fukumura, Van Gompel, & Pickering, 2010). In the present study, two possible referents were simultaneously depicted, which –parallel to the preceding argument– could have caused a competition of referent accessibility. This competition of accessibility was not the case in our ERP study in which the linguistic topic cue exclusively increased the accessibility of *one* referent while not mentioning the other (i.e., “*What about the ‘topic referent’?*”; Burmester et al., 2014). However, in our reading time study (Burmester et al., 2018) multiple referents were presented simultaneously and nevertheless an explicit visual (gaze) cue increased the accessibility of *one* amongst three depicted referents as reflected in sentence-initial processing ease similar to a linguistic topic cue. But, in contrast to the present study, participants in our reading time study were already familiarised with the visual scene by a multimodal lead-in *before* the gaze cue was presented. Therefore, participants might have had greater attentional capacities at the moment of processing the gaze cue than during processing the implicit cue with the subsequently presented pictures in the here presented study. Alternatively, the *type of the cue* matters, as gaze cues are more explicit and intentional in nature than the here used implicit abstract cue.

Taking previous studies using different visual cue types into account, we can assume that visual cues indeed modulate meaning computation during sentence comprehension while they seem to differ with respect to their impact on referent accessibility. As just mentioned, one type of visual cues that has been shown to clearly influence language processing, is eye gaze. This cue is a strong social-communicative cue signalling shared-attention of speaker and listener, it is hence an *intentional cue*. Crucial evidence provided by a few studies that compared both attentional and intentional cues supports the importance of the intentional component of visual

cues for listeners: Similar to speaker's gaze or pointing gesture, a visual cue (i.e., black square presented for 50 ms at the location of a possible referent) influences listener's pronoun interpretation, but only if the listener was previously instructed that this abstract visual cue is intentionally created by the speaker (Nappa & Arnold, 2014). However, the same visual cue but without the previous instruction of being intentionally created by the speaker does *not* reveal an impact. Analogously to this finding, Holle et al. (2012) showed that an intentional, conversational gesture co-occurring with the speaker's speech facilitates comprehension of German ambiguous SO and OS sentences: This short hand movement (i.e., beat gesture) emphasising the subject of the sentence reduces additional processing costs for disambiguation towards the OS word order as indicated by a reduced P600. In contrast to this gestural cue, listeners do not make use of an explicit visual attentional cue, that is, a moving point. Hence, the intentional component of visual cues plays a significant role in meaning computation during the comprehension process. This significance could be explained by assumptions of Pickering and Garrod's (2004) interactive alignment model. Accordingly, for a successful dialogue, interlocutors are assumed to develop aligned situational models. Thus, maybe intentional cues such as eye gaze offer a window into the interlocutors' mental representations in dialogue that might trigger other attentional mechanisms than those elicited by purely attentional, visual cues. However, the present study is a first step to better understand the impact of implicit attentional visual cues on sentence-initial processing, which demonstrates to us the subtle differences and difficulties of adapting study designs of sentence comprehension (Burmester et al., 2014) and sentence production (Gleitman et al., 2007).

Indeed, studying the impact of implicit visual cues on sentence comprehension might engender some caveats weakening the measurable outcome during later sentence processing. For the visual cue of the present study, we used a black square with  $0.3^\circ \times 0.3^\circ$  of visual angle against coloured pictures similar to the one by Gleitman et al. (2007), who used a slightly bigger cue (i.e.,  $0.5^\circ \times 0.5^\circ$  of visual angle) against full-colour clip-arts. Moreover, Myachykov et al. (2012) used a red dot with  $0.7^\circ \times 0.7^\circ$  against black-white-line drawings that –similar to the cue by Gleitman et al. (2007)– influenced sentence-initial mention during production. So maybe the visual cue in our study was too subliminal and hence, its impact too short-lived to be measurable during subsequent sentence processing. Following this line of thought, modulations of participants' attention by the implicit visual cue might be less strong and less long lasting with respect to their impact on referent accessibility compared to the high accessibility degree indicated by the explicit mention of the topic in the linguistic context. This explanation is in accordance with eye-tracking data reported by Arnold and Lao (2015): Listener's trial-initial attention to depicted referents was indeed modulated by multiple visual scene-based factors, that is, different visual attentional cues, the order of the depicted referents from left to right, and listener's idiosyncratic biases. But, the attentional cues themselves did not significantly predict subsequent pronoun interpretation and instead, the linguistic cue of sentence-initial mention was the strongest predictor. In addition, in an experimental task that requires sentence understanding, linguistic cues such as in Burmester et al. (2014) might less likely be ignored by the reader than visual cues and picture contexts such as in the present study. Hence, it is important to check whether visual cues and depicted scenes are indeed processed by the addressee of the stimuli.

However, based on the absent impact of the implicit visual cue on sentence processing in the present study, we suggest that this type of visually induced salience of discourse referents elicits a different, less strong, impact on the accessibility degree of mentally represented discourse referents compared to linguistically induced salience via topic-hood. This train of thought is supported by ERP and behavioural studies showing that linguistic stimuli (e.g., sentences, words) more strongly affect the accessibility of entities compared to pictures (e.g., Bögels, Schriefers, Vonk, & Chwilla, 2011; Fukumura et al., 2010). Moreover, a different neural processing of linguistic and visual stimuli has been suggested (e.g., Brandon & Andrew, 2007; Zhang, Begleiter, Porjesz, & Litke, 1997) which, for instance, has been explained by the more efficient



semantic access and/or memory retrieval of words compared to pictures (Dorjee, Devenney, & Thierry, 2010). However, words and pictures revealed a very similar time course of the N400 (congruity) effect with differences only in the topographical distribution (Ganis, Kutas & Sereno, 1996).

In summary, reasons for the null effect of the implicit cue in our study might, on the one hand, be traced back to a weaker, less long lasting impact on referent accessibility compared to intentional or linguistic cues. On the other hand, it might to some degree be related to methodological differences to production studies in which implicit cues modulated sentence-initial mention. Disentangling these reasons and testing the impact of other types of visual attentional cues is left open for future research.

#### **4.2 Word order effect**

In the present study, word order effects are reflected in sustained positive deflections (at DP1 and verb position) which were more pronounced for OS compared to SO sentences across multiple ROIs. Hence, we replicated differential processing costs for SO and OS sentences found in Burmester et al. (2014). A body of neurocognitive research concerning the processing of German sentences with varying word order demonstrates increased processing costs for OS sentences compared to their canonical (SO) counterpart as reflected in different ERP components, time windows and across different sentence positions (e.g., Holle et al., 2012; Knoeferle et al., 2007; Matzke, Mai, Nager et al., 2002; Schlesewsky, Bornkessel, & Frisch, 2003). For this paper, the word order effect is not of primary relevance and we take it just as a “sanity check” of our ERP-data. In line with the previous literature, we argue that the processing differences between OS and SO sentences are engendered by the subject/nominative-first-preference in German leading to increased processing demands for the non-canonical and less frequent OS word order. Note that in our study the impact of word order might be confounded with the ordering of theta roles, as both grammatical role and theta role coincided in the sentence constituents, that is, subject and agent, object and patient. The increased online processing difficulties were, however, not visible in the behavioural (accuracy) results of the subsequent probe sentence-picture-verifications. In summary, we argue that the replicated word order effect during online sentence processing confirms the validity of the design: The replication of the word order effect shows that we are not dealing with a replication failure per se but rather that the different findings can be traced back specifically to cue modality.

### **5 Conclusion**

All in all, previous findings speak in favour of the integration of visual and linguistic cues into listeners’ mental representations –although with a different magnitude of both cue modalities. In our study, the *implicit* visual cue to a depicted referent followed by subsequent sentence-initial mention did *not* influence online processing of German SO and OS sentences. Hence, the impact of the linguistic topic cue on the identical sentence material could not be replicated with the present study design, although a similar type of cue was effective in previous production studies (Gleitman et al., 2007). It therefore remains an open question if comprehension and production are influenced by similar underlying processes. Hence, we conclude that the role of visual, purely attention-directing cues for meaning computation during sentence processing needs further clarification. Future research needs to shed more light on the role of different visual cues and their interaction with intentional and attentional aspects in guiding information packaging preferences during utterance comprehension in order to disentangle experimental task specific effects.

## Acknowledgements

This work was supported by the German Research Foundation (DFG) under Grant SFB 632 'Information structure'. We thank Franziska Machens and Tobias Busch for assistance in material preparation and data collection as well as Jan Ries for his help in preparation of the Figures.

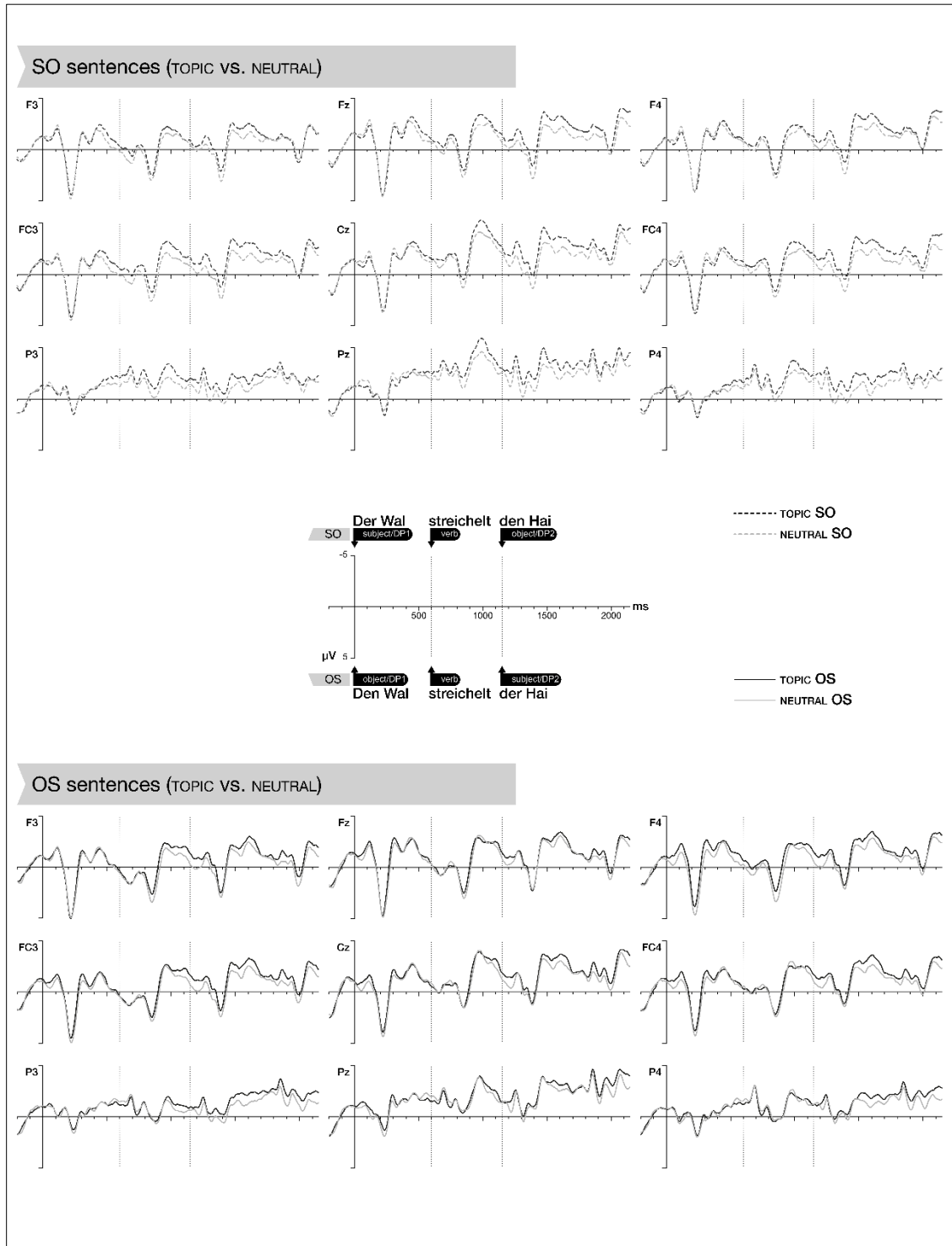
## Appendix A

Results of analysis of variance (ANOVAs) of the ERPs for the different time windows time-locked to the onset of the first determiner phrase (DP1).

	<i>df</i>	<i>F</i> -values		
		100 - 300 ms	300 - 500 ms	500 - 700 ms
CUE (TOPIC vs. NEUTRAL)	1, 28	0.669	0.266	1.020
CUE X ROI	8, 224	0.275 ( $\epsilon = 0.252$ )	0.575 ( $\epsilon = 0.337$ )	0.978 ( $\epsilon = 0.285$ )
CUE X WORD ORDER	1, 28	1.374	0.030	0.723
CUE X WORD ORDER X ROI	8, 224	1.760 ( $\epsilon = 0.411$ )	0.564 ( $\epsilon = 0.399$ )	1.605 ( $\epsilon = 0.417$ )
WORD ORDER (SO vs. OS)	1, 28	0.983	1.004	6.254*
WORD ORDER X ROI	8, 224	1.872 ( $\epsilon = 0.433$ )	1.425 ( $\epsilon = 0.378$ )	3.004* ( $\epsilon = 0.424$ )

*Note.* Greenhouse & Geisser (1959) corrected significance levels: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ . *df* = degrees of freedom.  $\epsilon$  = Greenhouse & Geisser epsilon factor for non-sphericity to adjust the original *df* according to Jennings and Wood (1976).

Appendix B



**Figure B-1.** Grand average ERPs (with baseline correction) at selected electrodes for the factors CUE (TOPIC vs. NEUTRAL) and WORD ORDER (SO vs. OS) time-locked to the onset of the target sentence, that is, the first Determiner phrase (DP1) followed by the verb and second Determiner phrase (DP2): upper panel: TOPIC SO [dotted black] vs. NEUTRAL SO [dotted grey], lower panel: TOPIC OS [solid black] vs. NEUTRAL OS [solid grey]. Negativity is plotted upwards.

## Appendix C

Results of overall analysis of variance (ANOVAs) of the ERPs for the different time windows time-locked to the onset of the first determiner phrase (DP1) of target sentences following both cue modalities (MODALITY), that is, the implicit visual cue (i.e., VISUAL: data of the present study) and the linguistic cue (i.e., LINGUISTIC: data published in Burmester et al., 2014).

	<i>df</i>	<i>F</i> -values		
		100 - 300 ms	300 - 500 ms	500 - 700 ms
CUE (TOPIC vs. NEUTRAL)	1, 46	1.970	0.065	0.000
CUE x ROI	8, 368	0.311 ( $\varepsilon = 0.311$ )	0.736 ( $\varepsilon = 0.428$ )	3.454* ( $\varepsilon = 0.373$ )
CUE x WORD ORDER	1, 46	4.162*	0.440	0.829
CUE x WORD ORDER x ROI	8, 224	2.306 ( $\varepsilon = 0.404$ )	2.088 ( $\varepsilon = 0.452$ )	4.429** ( $\varepsilon = 0.441$ )
WORD ORDER (SO vs. OS)	1, 46	1.429	0.139	0.136
WORD ORDER x ROI	8, 368	2.372 <sup>#</sup> ( $\varepsilon = 0.507$ )	1.666 ( $\varepsilon = 0.483$ )	1.435 ( $\varepsilon = 0.460$ )
MODALITY (VISUAL vs. LINGUISTIC) x ROI	8, 368	3.190*	1.266	1.378
MODALITY x CUE	1, 46	0.089	0.018	1.778
MODALITY x CUE x ROI	8, 368	0.538	0.592	1.891
MODALITY x WORD ORDER	1, 46	5.793*	0.567	3.717
MODALITY x WORD ORDER x ROI	8, 368	1.117	0.377	1.152
MODALITY x CUE x WORD ORDER	1, 46	0.285	0.200	3.963 <sup>#</sup>
MODALITY x CUE x WORD ORDER x ROI	8, 368	0.446	1.843	1.487

*Note.* Greenhouse & Geisser (1959) corrected significance levels: <sup>#</sup>  $p < .06$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ . *df* = original degrees of freedom.  $\varepsilon$  = Greenhouse & Geisser epsilon factor for non-sphericity to adjust the original *df* according to Jennings and Wood (1976).

**Appendix D**

Results of post hoc pairwise *t*-tests of the ERPs to resolve the ROI x CUE POSITION interaction in the N1 and P2 time window (i.e., 100 - 200 ms, 250 - 350 ms) time-locked to the onset of the visual cue.

ROI	CUE POSITION		TOPIC LEFT vs. TOPIC RIGHT				TOPIC LEFT vs. NEUTRAL BOTTOM				TOPIC RIGHT vs. NEUTRAL BOTTOM			
			100 - 200 ms		250-350 ms		100-200 ms		250-350 ms		100 - 200 ms		250 - 350 ms	
	<i>t</i> (28)	<i>p</i>	<i>t</i> (28)	<i>p</i>	<i>t</i> (28)	<i>p</i>	<i>t</i> (28)	<i>p</i>	<i>t</i> (28)	<i>p</i>	<i>t</i> (28)	<i>p</i>	<i>t</i> (28)	<i>p</i>
LEFT FRONTAL	0.753	0.458	1.153	0.259	3.780	0.001*	3.726	0.001*	1.659	0.108	1.767	0.088		
RIGHT FRONTAL	-1.076	0.291	-2.207	0.036	-0.344	0.734	0.909	0.371	0.914	0.368	3.131	0.004		
LEFT FRONTO-CENTRAL	1.512	0.142	2.616	0.014	4.178	<.001*	3.328	0.002	2.220	0.035	1.093	0.284		
RIGHT FRONTO-CENTRAL	0.237	0.814	0.289	0.775	2.573	0.016	2.907	0.007	2.247	0.033	2.755	0.010		
LEFT CENTRO-PARIETAL	0.568	0.575	1.333	0.193	4.885	<.001*	2.844	0.008	4.313	<.001*	2.023	0.053		
RIGHT CENTRO-PARIETAL	1.055	0.301	0.111	0.912	4.208	<.001*	1.962	0.060	3.467	0.002	2.369	0.025		
FRONTAL-MIDLINE	1.542	0.134	0.804	0.428	2.523	0.018	3.068	0.005	0.397	0.694	1.954	0.061		
CENTRAL MIDLINE	1.646	0.111	3.210	0.003	3.810	0.001*	4.503	<.001*	1.883	0.070	1.145	0.262		
PARIETAL MIDLINE	2.254	0.032	2.625	0.014	6.538	<.001*	3.994	<.001*	5.021	<.001*	1.348	0.188		

*Note.* Bonferroni corrected significance level (two-tailed) at  $\alpha = .05$ : \*  $p < .002$ .

## References

- Gerry T.M. Altmann (2004). Language-mediated eye movements in the absence of a visual world: the ‘blank screen paradigm’. *Cognition*, 93(2): B79–B87.
- American Electroencephalographic Society (2006). Guideline 5: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 11: 111–113.
- Mira Ariel (1988). Referring and accessibility. *Journal of Linguistics*, 24(01): 65–87.
- Arnold (1998). *Reference form and discourse patterns*. PhD thesis, Stanford University, Stanford, California.
- Jennifer E. Arnold (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4): 187–203.
- Jennifer E. Arnold and Zenzi M. Griffin (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, 56(4): 521–536.
- Jennifer E. Arnold and Shin-Yi C. Lao (2015). Effects of psychological attention on pronoun comprehension. *Language, Cognition and Neuroscience*, 30(7): 832–852.
- R. Harald Baayen (2008). *Analysing linguistic data*. Cambridge University Press, Cambridge.
- Markus Bader and Jana Häussler (2010). Word order in German: A corpus study. *Lingua*, 120(3): 717–762.
- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1): 1–48.
- Kathryn J. Bock and Richard K. Warren (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1): 47–67.
- Sara Bögels, Herbert Schriefers, Wietske Vonk, and Dorothee J. Chwilla (2011). Pitch accents in context: how listeners process accentuation in referential communication. *Neuropsychologia*, 49(7): 2022–2036.
- Ina Bornkessel and Matthias Schlesewsky (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4): 787–821.
- Ina Bornkessel, Matthias Schlesewsky, and Angela D. Friederici (2003). Contextual information modulates initial processes of syntactic integration: The role of inter- versus intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5): 871–882.
- Gordon H. Bower and Daniel G. Morrow (1990). Mental models in narrative comprehension. *Science*, 247(4938): 44–48.
- A. Ally Brandon and E. Budson Andrew (2007). The worth of pictures: Using high density event-related potentials to understand the memorial power of pictures and the dynamics of recognition memory. *NeuroImage*, 35(1): 378–395.
- Harm Brouwer, Matthew W. Crocker, Noortje J. Venhuizen, and John C.J. Hoeks (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41: 1318–1352.
- Petra Burkhardt (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2): 159–168.
- Juliane Burmester, Antje Saueremann, Katharina Spalek, and Isabell Wartenburger (2018). Sensitivity to salience: Linguistic vs. visual cues affect sentence processing and pronoun resolution. *Language, Cognition and Neuroscience*, 33(6): 784–801.
- Juliane Burmester, Katharina Spalek, and Isabell Wartenburger (2014). Context updating during sentence comprehension: The effect of aboutness topic. *Brain and Language*, 137: 62–76.

- Herbert H. Clark and Eve V. Clark (1977). *Psychology and language: An introduction to psycholinguistics*. Harcourt Brace Jovanovich, New York.
- Maurizio Corbetta and Gordon L. Shulman (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3): 201–215.
- Matthew W. Crocker, Pia Knoeferle, and Marshall R. Mayberry (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112(3): 189–201.
- Francesca Delogu, Heiner Drenhaus, and Matthew W. Crocker (2018). On the predictability of event boundaries in discourse: An ERP investigation. *Memory & Cognition*, 46(2): 315–325.
- Dusana Dorjee, Lydia Devenney, and Guillaume Thierry (2010). Written words supersede pictures in priming semantic access: a P300 study. *NeuroReport*, 21(13): 887–891.
- Matthew S. Dryer (2013). Order of subject, object and verb. In: Matthew S. Dryer and Martin Haspelmath (Eds). *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, Available online at <http://wals.info/chapter/81>.
- Lyn Frazier and Janet D. Fodor (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4): 291–325.
- Angela D. Friederici (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2): 78–84.
- Kumiko Fukumura, Roger P.G. Van Gompel, and Martin J. Pickering (2010). The use of visual context during the production of referring expressions. *The Quarterly Journal of Experimental Psychology*, 63(9): 1700–1715.
- Giorgio Ganis, Marta Kutas, and Martin I. Sereno (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.
- Lesya Y. Ganushchak, Agnieszka E. Konopka, and Yiya Chen (2017). Accessibility of referent information influences sentence planning: An eye-tracking study. *Frontiers in Psychology*, 8: 250.
- Morton A. Gernsbacher (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *The Psychology of Learning and Motivation*, 27: 217–263.
- Thomas Givón (1988). The pragmatics of word-order: predictability, importance and attention. In: Michael Hammond, Edith A. Moravczik, and Jessica R. Wirth (Eds). *Studies in syntactic typology*. Typological Studies in Language. 17. John Benjamins Publishing Company, Amsterdam, 243–284.
- Lila R. Gleitman, David January, Rebecca Nappa, and John C. Trueswell (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4): 544–569.
- Gabriele Gratton, Michael G. H. Coles, and Emanuel Donchin (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55: 468–484.
- Samuel W. Greenhouse and Seymour Geisser (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2): 95–112.
- Tanja Grewe, Ina Bornkessel, Stefan Zysset, Richard Wiese, D.Yves von Cramon, and Matthias Schlesewsky (2006). Linguistic prominence and Broca's area: The influence of animacy as a linearization principle. *NeuroImage*, 32(3): 1395–1402.
- Barbara J. Grosz and Candace L. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3): 175–204.
- Jeanette K. Gundel (1985). 'Shared knowledge' and topicality. *Journal of Pragmatics*, 9: 83–107.
- Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2): 274–307.

- Peter Hagoort and Jos J.A. Van Berkum (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 362(1481): 801–811.
- Joy E. Hanna and Susan E. Brennan (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4): 596–615.
- Julian Heister, Kay-Michael Würzner, Johannes Bubenzer, Edmund Pohl, Thomas Hanneforth, Alexander Geyken, and Reinhold Kliegl (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, 62(1): 10–20.
- Barbara Hemforth (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. PhD Thesis, University of Bochum, Bochum.
- Henning Holle, Christian Obermeier, Maren Schmidt-Kassow, Angela D. Friederici, Jamie Ward, and Thomas C. Gunter (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, 3: 74.
- Ray Jackendoff (1972). *Semantic interpretation in Generative Grammar*. MIT Press, Cambridge, Mass.
- Richard Jennings and Charles C. Wood (1976). The  $\epsilon$ -adjustment procedure for repeated-measures analyses of variance. *Psychophysiology*, 13(3): 277–278.
- Philip N. Johnson-Laird (1980). Mental models in cognitive science. *Cognitive Science Proceedings*, 4: 71–115.
- Marcel Adam Just and Patricia A. Carpenter (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4): 329–354.
- Edith Kaan, Andrea C. Dallas, and Christopher M. Barkley (2007). Processing bare quantifiers in discourse. *Brain Research*, 1146: 199–209.
- Pia Knoeferle, Boukje Habets, Matthew W. Crocker, and Thomas F. Münte (2007). Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*, 18(4): 789–795.
- Pia Knoeferle and Helene Kreysa (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*, 3: 538.
- Willem J. M. Levelt (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, Mass.
- Steven J. Luck and Steven A. Hillyard (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31: 291–308.
- George R. Mangun (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, 32: 4–18.
- William Marslen-Wilson and Lorraine K. Tyler (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1): 1–71.
- Mike Matzke, Heinke Mai, Wido Nager, Jascha Rüsseler, and Thomas F. Münte (2002). The costs of freedom: an ERP – study of non-canonical sentences. *Clinical Neurophysiology*, 113: 844–852.
- Andriy Myachykov, Dominic Thompson, Simon Garrod, and Christoph Scheepers (2012). Referential and visual cues to structural choice in visually situated sentence production. *Frontiers in Psychology*, 2: 396.
- Rebecca Nappa and Jennifer E. Arnold (2014). The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology*, 70: 58–81.
- Rebecca Nappa, Allison Wessel, Katherine L. McEldoon, Lila R. Gleitman, and John C. Trueswell (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, 5(4): 203–234.



- R. Carolus Oldfield (1971). The assessment and analysis of handedness: The Edinburgh handedness inventory. *Neuropsychologia*, 9: 97–113.
- Marte Otten and Jos J.A. Van Berkum (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6): 464–496.
- Martin J. Pickering and Simon Garrod (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Science*, 27: 169–226.
- Paul Portner (2007). Instructions for interpretation as separate performatives. In: Kerstin Schwabe and Susanne Winkler (Eds.). *On information structure, meaning and form: Generalizations across languages*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Mercè Prat-Sala and Holly P. Branigan (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, 42(2): 168–182.
- Ellen F. Prince (1981). Topicalization, focus-movement, and Yiddish-movement: A pragmatic differentiation. In: *Proceedings of the seventh annual meeting of the Berkeley Linguistics Society*. Berkeley Linguistics Society. 249–264.
- Ellen F. Prince (1992). The ZPG letter: Subjects, definiteness, and information-status. In: Sandra A. Thompson and William C. Mann (Eds.). *Discourse description: Diverse analyses of a fund raising text*. John Benjamins Publishing Company, Amsterdam, 295–325.
- R Core Team (2013). *R: A language and environment for statistical computing*, Vienna, Austria.
- Tanja Reinhart (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27(1): 53–94.
- Christoph Scheepers and Matthew W. Crocker (2004). Constituent order priming from reading to listening: A visual-world study. In: Manuela Carreiras and Charles Clifton, JR. (Eds.). *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond*. Psychology Press, New York, 167–185.
- Matthias Schlesewsky, Ina Bornkessel, and Stefan Frisch (2003). The neurophysiological basis of word order variations in German. *Brain and Language*, 86(1): 116–128.
- Petra B. Schumacher (2014). Content and context in incremental processing: “the ham sandwich” revisited. *Philosophical Studies*, 168(1): 151–165.
- Petra B. Schumacher and Yu-Chen Hung (2012). Positional influences on information packaging: Insights from topological fields in German. *Journal of Memory and Language*, 67(2): 295–310.
- Samuel C. Spevack, J. Benjamin Falandays, Brandon Batzloff, and Michael J. Spivey (2018). Interactivity of language. *Language and Linguistics Compass*, 12(7): e12282.
- Maria Staudte and Matthew W. Crocker (2011). Investigating joint attention mechanisms through spoken human–robot interaction. *Cognition*, 120(2): 268–291.
- Maria Staudte, Matthew W. Crocker, Alexis Heloir, and Michael Kipp (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, 133(1): 317–328.
- Russell S. Tomlin (1995). Focal attention, voice, and word order: An experimental, cross-linguistic study. In: Pamela Downing and Michael Noonan (Eds.). *Word Order in Discourse*. John Benjamins Publishing Company, Amsterdam and Philadelphia, 517–555.
- Russell S. Tomlin (1997). Mapping conceptual representations into linguistic representations: the role of attention in grammar. In: Jan Nuyts and Eric Pederson (Eds.). *Language and conceptualisation*. Language, Culture & Cognition. 1. Cambridge University Press, Cambridge, 162–189.
- Elizabeth A. Turner and Ragnar Rommetveit (1968). Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 7(2): 543–548.
- Jos J.A. Van Berkum, Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman, and Peter Hagoort (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3): 443–467.

- Teun A. Van Dijk and Walter Kintsch (1983). *Strategies of Discourse Comprehension*. Academic Press, New York.
- Jorrig Vogels, Emiel Krahmer, and Alfons Maes (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, 28(9): 1323–1349.
- Andrea Weber and Karin Müller (2004). Word order variation in German main clauses: A corpus analysis. In: *Proceedings of the 20th international conference on computational linguistics*, Geneva, Switzerland, 71–77.
- Xiao Lei Zhang, Henri Begleiter, Bernice Porjesz, and Ann Litke (1997). Visual object priming differs from visual word priming: an ERP study. *Electroencephalography and Clinical Neurophysiology*, 102(3): 200–215.
- Rolf A. Zwaan (2004). The immersed experiencer: Towards an embodied theory of language comprehension. In: Brian H. Ross (editor). *The psychology of learning and motivation: Advances in research and theory*. Academic Press, New York, 35–62.
- Rolf A. Zwaan and Gabriel A. Radvansky (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2): 162–185.