# Construal in language

Divjak, Dagmar; Milin, Petar; Medimorec, Srdan

[Link to publication on Research at Birmingham portal](Link to publication on Research at Birmingham portal)

Dagmar Divjak[†,*], Petar Milin[†] and Srdan Medimorec

# Construal in language: A visual-world approach to the effects of linguistic alternations on event perception and conception

**Abstract:** The theoretical notion of 'construal' captures the idea that the way in which we describe a scene reflects our conceptualization of it. Relying on the concept of ception – which conjoins conception and perception – we operationalized construal and employed a Visual World Paradigm to establish which aspects of linguistic scene description modulate visual scene perception, thereby affecting event conception. By analysing viewing behaviour after alternating ways of describing *location* (prepositions), *agentivity* (active/passive voice) and *transfer* (NP/PP datives), we found that the linguistic construal of a scene affects its spontaneous visual perception in two ways: either by determining the order in which the components of a scene are accessed or by modulating the distribution of attention over the components, making them more or less salient than they naturally are. We also found evidence for the existence of a cline in the construal effect with stronger expressive differences, such as the prepositional manipulation, inducing more prominent changes in visual perception than the dative manipulation. We discuss the claims language can lay to affecting visual information uptake and hence conceptualization of a static scene in the light of these results.

**Keywords:** construal, ception, preposition, voice, dative, visual world paradigm, eye-movements

[†] Joint first authors

**\*Corresponding author: Dagmar Divjak,** Departments of English Language & Linguistics and Modern Languages, University of Birmingham, Edgbaston, Ashley Building, Birmingham B15 2TT, United Kingdom of Great Britain and Northern Ireland, E-mail: d.divjak@bham.ac.uk
https://orcid.org/0000-0001-6825-8508
**Petar Milin,** Department of Modern Languages, University of Birmingham, Birmingham, United Kingdom of Great Britain and Northern Ireland, E-mail: p.milin@bham.ac.uk
https://orcid.org/0000-0001-9708-7031
**Srdan Medimorec,** Department of Psychology, Teesside University, Middlesbrough TS1 3BA, United Kingdom of Great Britain and Northern Ireland, E-mail: s.medimorec@tees.ac.uk
https://orcid.org/0000-0001-7552-5432

# 1 Introduction

Language provides a variety of ways to express events. For example, we are at liberty to choose between the active and passive voice to describe a scene. The linguistic packaging is thus not dictated by the scene itself but reflects the speaker's conceptualization of it. Language or, more specifically, *communicative intentions* seem to work independently from *perceptual constraints*, at least to a certain degree. Cognitive linguistics uses the theoretical concept of 'construal' to account for the choice between alternating expressions. The two grammatical possibilities for expressing one and the same situation are two different ways of describing and thereby 'construing' that situation. The lexical and syntactic choices that speakers make reflect a specific framing of their experience and a certain commitment to how that experience will be communicated between interlocutors. Construal is one of the fundamental notions in the cognitive linguistic approach to language (Section 1.1) and we set out to test to what extent it invokes changes in how a scene is perceived and conceived across a cline of constructions (Section 1.2).

## 1.1 Construal: When language affects mental imagery

Construal plays a prominent role in what constitutes the core of the cognitive linguistic approach to language. For cognitive linguists, meaning resides in cognition, rather than in the relationship between language and world. While certain linguistic traditions consider the relation between language and world as a static fact that can be adequately described with truth conditions, cognitive linguists recognize that *meaning is made*. The language user plays an important role in making meaning as the one who negotiates the experience, its perception and its description. Therefore, meaning cannot be captured satisfactorily by an analysis of the properties of the object of conceptualization alone; instead, it requires the inclusion of the subject of conceptualization (Verhagen 2007) alongside the properties of the code used to describe the object.

Experience is so rich that there is no single way to represent a situation. The grammar of a language provides users with a range of constructions, which differ in meaning and satisfy varying semiotic and interactive goals. Construal thus (re)distributes attention in a specific way, or (re)directs attention towards certain aspects of the situation and reflects the user's ability to adjust the focus of attention by altering what tends to be called the *mental imagery* associated with a situation. Both Langacker (1987) and Talmy (1988) have proposed

detailed and largely overlapping classifications of construal phenomena (these classifications were later revised in Talmy 2000; Langacker 2007). Cognitive Grammar (Langacker 1987), crucially, depends on the notion of 'profiling', a construal operation which makes some entity stand out in profile against a background. That entity is called the Figure: it receives prominence as pivotal entity around which the scene is organized. Crucially, the Figure/Ground organization is not predetermined for a given scene and it tends to be possible to structure the scene around different Figures. Cognitive linguists have argued that language plays a pivotal role in this process as an "attention-directing device": linguistic choices reveal and support the need to bring one element or event rather than another to the listener's attention.

The Figure/Ground segmentation stems from research on visual perception within the tradition of Gestalt Psychology. Pioneering work on this particular problem was done by Rubin (1921), and Talmy (1978) was the first to introduce it in Cognitive linguistics. In Gestalt psychology, the whole (*die Gestalt*) is more than the sum of its parts, and Gestalt Formation has been invoked to explain why the mind would prefer a unit as a whole rather than its parts (Von Ehrenfels 1890). The Figure/Ground segmentation, in particular, represents the principle of perceptual grouping and, essentially, describes the process of simplifying visual input for efficient perception (cf., Wever 1927). The Figure/Ground segmentation has been fruitfully applied to interpret a wide range of language-related phenomena (Croft and Cruse 2004).

Language can be seen as a promoter or a demoter of the salience of various situational cues, which modulates how we attend to those cues. The aim of this study is to investigate the effect that the different ways to describe a scene linguistically have on that scene's perception and conception across three types of alternations: the locational or prepositional alternation (*the poster is above the bed* vs *the bed is below the poster*), the voice alternation (*the policeman arrested the thief* vs *the thief was arrested by the policeman*) and the dative alternation (*the boy gave the girl a flower* vs *the boy gave a flower to the girl*) in English. These three alternations were chosen because they represent a cline with respect to the extent to which they implement the construal operation of perspective or prominence (Langacker 1987, Langacker 2007). Moreover, each alternation has a more frequently used variant, and a less frequently used one, with the more frequent version accounting for roughly 80% of all instances of occurrence.

The locational or prepositional alternation is a textbook example of perspective or prominence, and clearly implements the Figure/Ground idea from Gestalt Psychology. Both *the poster is above the bed* and *the bed is below the poster* denote one and the same spatial configuration but, in each case, a

different element is selected as Ground with respect to which the Figure is located. In the formulation *the poster is above the bed* it is the bed that serves as Ground against which the Figure, i. e., the poster, is situated. But the poster serves as Ground for the bed in *the bed is below the poster*. In this alteration, spatial configuration is key; ultimately, the construction is about locating two entities (e. g., two objects) with respect to each other. It is typical for speakers to consider the larger entity as the Ground, and the smaller, more easily moveable entity as Figure (Shank and Walker 1989).

Voice alternations have received ample attention in linguistics and psycholinguistics (for a summary see Thompson 2012). Different from the locational alternation, the voice alternation is not about a spatial configuration but about the relations between animate agents in an event: the voice alternation thus implements a Figure/Ground distinction in a scene with animate participants. In the active voice, *the policeman arrested the thief*, the Agent is mentioned first and treated as Figure, while in the passive voice, *the thief was arrested by the policeman*, the Patient is named first and treated as Figure. In passive sentences, the Agent can remain implied or can be omitted altogether, which supports interpreting the role of the Agent as (back)ground for the Patient. It is typical for speakers to select the active construction, where the Agent is the Figure, with active sentence more frequent than passive sentences (Roland et al. 2007).

The dative alternation, finally, is the most intricate with respect to perspective or prominence as it leaves the relation between the two interacting animate beings in the scene unaffected, and instead focuses on the relation between the Object and the Recipient. Corpus analysis has suggested that the difference between the noun phrase dative (NP) *the boy gave the girl a flower* and the prepositional phrase dative (PP) *the boy gave a flower to the girl*, are determined by fine contextual differences such as giveness, animacy and pronominality (Bresnan et al. 2007). This hypothesis was later experimentally confirmed (Bresnan and Ford 2010). The NP, also known as the *double object* dative, is more frequent than the PP and accounts for 79% of all occurrences in a corpus of spoken language and for 62% of all occurrences in written language (Bresnan et al. 2007).

Although construal is one of the fundamental notions in the cognitive approach to language, the extent to which it invokes changes in how a scene is perceived has not yet been tested explicitly. In previous research, the conceptualization accompanying a particular linguistic choice has typically been determined by the analyst. In this paper, we rely on eye-tracking to chart how differences in describing static scenes affect the way in which language users view them.

## 1.2 Ception: Between perception and conception

Construal is a multi-faceted phenomenon (Croft and Cruse 2004) and several dimensions of construal can be activated in one linguistic expression (Verhagen 2007). In addition to the above-mentioned adjustments that construal makes in terms of perspective or prominence, it also affects dynamicity. Langacker's dynamicity concerns the development of a conceptualization through processing time rather than through conceived time. It is connected to the inherent temporal nature of linguistic utterances: presenting elements of a conceptualization in a different order results in a different meaning (Verhagen 2007: 53–54). Research on sequential viewing confirms this: eye fixations follow the order of naming and elements are fixated in the order in which they will be named in a sentence just before producing a sentence (Griffin and Bock 2000; Myachykov et al. 2013).

In this paper, we continue on the path of scene viewing. In order to test to what extent a different linguistic construction of a scene by the speaker suggests a different take on that situation to the hearer, we investigate the extent to which the way in which a scene is perceived differs depending on how the scene is described. Involving perception in a study on construal is a logical step for two reasons. Involving perception is methodologically advantageous because perception can be accurately measured, as will be explained in Section 2. It is also theoretically justified, as perception and conception are conjoined within Cognitive linguistics through Talmy's (2000) notion of *ception*. Talmy (2000) defined ception as a conjunction of the domains of perception and conception, representing both in a single continuous domain, to cover "all the cognitive phenomena, conscious and unconscious, understood by the conjunction of perception and conception" (p. 139). This combination of perception and conception makes ception a suitable theoretical starting point for an empirical investigation of construal, and for the following question in particular: does any difference in linguistic encoding affect the way in which events are perceived and is this effect consistent in the larger population?

To test construal behaviourally, we embed our study of construal in the larger framework of studies on visual perception and language-guided visual perception in particular. The way in which our eyes sample the environment constrains what is available for further processing (Desimone and Duncan 1995; Egeth and Yantis 1997), and the distribution of visual attention depends both on the properties of the stimulus and the observer's goals (Bacon and Egeth 1994; Egeth and Yantis 1997; Langton et al. 2008; Parkhurst et al. 2002). That language might play a role in marking salient cues and, thus, in constraining attention has long been known. Studies investigating which elements in a static scene attracts attention revealed that (verbal) instructions affect viewing patterns

(Buswell 1935; Yarbus 1967). Later research has revealed more specific properties of the language-perception link: individuals' visual attention can be mediated by the unfolding language input (Cooper 1974; Tanenhaus et al. 1995) in that their eye movements follow the order of objects mentioned in sentences closely (Allopenna et al. 1998; Dahan et al. 2001) and anticipate objects before they are mentioned in a sentence (Altmann and Kamide 1999).

The interaction of perception and language remains a hotly debated domain of interdisciplinary research (Huettig et al. 2011; Lupyan 2012; Lupyan and Lewis 2017) and the empirical evidence that has accrued focuses on the most robust correlations between language and perception, i. e., between naming and viewing. However, the effect of more subtle linguistic differences, such as the prepositional, the voice and the dative alternations, on perception *and* conception remains understudied.[1] Complementary insights are available from a psycholinguistic tradition that investigates how attentional resources are implicated in language *production* (Tomlin and Myachykov 2015). These studies examine how the salience of the elements in a scene and the distribution of attention over the elements in a scene influence the order in which the elements are named and the grammatical roles they are assigned in a visually situated spoken sentence across a range of different languages (e. g., English: Tomlin 1995; Russian: Myachykov and Tomlin 2008; Finnish: Myachykov et al. 2011; Korean: Hwang and Kaiser 2009).

## 1.3 This study

By relying on the concept of ception (Talmy 2000) to link conception to perception, we can operationalize construal in such a way that it sheds light on the mutually co-implicative domains of language (and language-encoded conceptualisation in particular) and perception. While previous findings have established a global language-perception link, the current study elucidates the linguistic specifics of that link. Using data from a visual world eye-tracking study, we investigate which differences in linguistic encoding affect scene perception across three constructions that represent a cline with respect to the extent to which they implement the construal operation of perspective or prominence. This cline is expected to be reflected both in the number of eye-movement measures that show an effect and in the strength of these effects: the preposition

---

**1** A notable exception here is work on spatial language (Altmann and Kamide 1999; Coventry et al. 2010; Lindsay et al. 2013), which focuses on the fine-grained differences in linguistic expression and how these differences affect scene inspection and associated event comprehension.

alternation is expected to show stronger effects across more measures, while the dative alternation is expected to show weaker effects across fewer measures. Not observing any relationship between language and perception at all would lead to the conclusion that construal is an expressive device that is informative about discourse preferences, but does not affect the distribution of attention over elements of a scene. This would limit the claims it can lay to affecting visual information uptake and hence conceptualization of a static scene.

# 2 Materials and methods

## 2.1 Visual world paradigm

We used a Visual World Paradigm to investigate to what extent linguistic encoding affects the way in which events are perceived, and thus potentially conceived. The Visual World Paradigm, henceforth VWP, is an eye-tracking method often used in the context of spoken language processing (Cooper 1974). In this task, participants are presented with images on a screen, while simultaneously listening to spoken stimuli. In an attempt to understand how the linguistic description of a scene affects scene perception and conception, we opted for a consecutive processing task by introducing a slight delay between the sentence presentation offset and the image presentation onset.

## 2.2 Participants

Sixty students and staff (46 female; mean age = 27.4, age range: 18–57) from the University of Sheffield (UK) participated in the experiment in exchange for £7. All participants were native speakers of English and had normal or corrected-to-normal vision. Six further participants were excluded because they either failed to complete the experiment or were non-native English speakers.

## 2.3 Study design

We ran a cross-modal visual world eye-tracking study in which we recorded the gaze of participants as they viewed scenes and listened to a description (e. g., active or passive sentences). We used a two-level (natural image viewing vs. language-and-image viewing) within-subject design.

### 2.3.1 Stimuli and apparatus

Visual stimuli consisted of 48 full-coloured photographs, with resolution of $1024 \times 768$ pixels. The photographs were presented on a 21-inch monitor (refresh rate: 60 Hz), 70 cm away from participants' eyes, subtending visual angles of 22.9° horizontally and 17.2° vertically. All images depicted naturalistic events and were downloaded from the internet under the creative commons licence.

The auditory stimuli were 96 recorded sentences (see Supplementary Materials A) describing the 48 images (events); each image was described in two different ways (e. g., active/passive). A major consideration when creating a stimulus was its imageability – ultimately, the sentence had to depict a scene for which a naturalistic image could be found. Given that the strength of association between a word and a construction depends on how it is calculated (compare here the dative NP vs PP preferences as reported in Bresnan et al. 2007 versus Gries and Stefanowitsch 2004) we settled for depictable scenes that were adequately described using either construction; lexical effects were controlled for statistically (see Section 3). Thus, there were 32 sentences in each of the three categories: Preposition (16 typical/16 atypical sentences with typical sentences locating the more easily moveable item with respect to the less easily moveable one), Voice (16 active/16 passive sentences), and Dative (16 noun phrase/16 prepositional phrase sentences). The stimuli were recorded in a sound-proofed room by a female native speaker of British English who was a professional radio broadcaster. The average sentence duration was 2600 ms (range: 1877–3672 ms) across the data; for Preposition the range was 2014–3672 ms, for Voice 1877–2713 ms and for Dative 2172–3654 ms The auditory stimuli were presented to participants through Sennheiser HD 280 pro headphones.

The task was implemented using OpenSesame (Mathôt et al. 2012). The eye movement data were collected using an EyeLink Portable Duo eye tracker (SR research, ON, Canada), tracking at a sampling rate of 500 Hz in the head-stabilized mode. Participants were calibrated by the 9-point calibration type. Tracking was monocular, using participants' dominant eye. Ocular dominance was determined using a variation of the distance hole-in-the-card test. In this test, participants held a card (210 × 297 mm) with a 30 mm diameter hole in the centre, with both hands extended. They were asked to visually align (with both eyes) the hole in the card with a target (diameter = 18 mm) at 2 m distance. Next, participants were instructed to close the left eye and asked if they could still see the target. The procedure was repeated with the closed right eye. The eye that could see the target was the dominant eye. When neither eye was dominant, the right eye was used. Overall, the right eye was recorded for 75% of participants ($n = 45$).

### 2.3.2  Procedure

Participants were seated in front of the monitor and head position was controlled using a chinrest. The calibration was performed at the beginning of the experiment. Drift corrections were performed between blocks.

In the first block of the experiment, participants were instructed to look at the images in the absence of any linguistic guidance. The image-only condition would reveal the default or naturalistic viewing pattern for the scene and also eliminate the effect of any salient image properties from the language-mediated viewing conditions in Blocks 2 and 3. Each of the 48 images was presented for 3500 ms, and the order of presentation was randomized. Individual images were preceded by a 1000 ms central fixation point presented on a grey background.

In the following two blocks (Blocks 2 and 3) participants were instructed to listen to individual sentences and then look at the matching images. This would reveal the changes in viewing pattern due to linguistic construal of the scene. There were 48 trials, consisting of sentence/image pairs, per block. For each trial, a central fixation point was presented for 1000 ms, followed by a sentence, a 250 ms fixation point, and finally an image depicting the event described by the sentence. During the sentence presentation, the central fixation point remained on the screen. Images were presented for 3500 ms. The order of Blocks 2 and 3 was counterbalanced across participants, and the presentation order of sentence/image pairs within blocks was randomized. Two different sentences describing the same image always appeared in different blocks (Block 2 or Block 3) and in different categories (i. e., Preposition, Voice, Dative) with the corresponding subcategories evenly distributed between Blocks 2 and 3. The entire experiment took approximately 20 minutes to complete.

## 2.4  Data preparation

Our definition of interest areas (IAs) was content-dependent, i. e., each IA was defined empirically, based on the components of the scene. While the IAs were fixed to a scene component, the order in which the elements were mentioned changed according to the alternating construction. Details are presented in Table 1 below. Note that in Voice and Dative, A refers to the Agent, B refers to Patient or Recipient and C refers to the Action or Object. For Preposition, A captures the more easily moveable item, while B captures the less easily moveable item.

Custom IAs were created using the EyeLink Data Viewer software (SR Research, ON, Canada), and validated using fixation heat maps from 10

**Table 1:** Assignment of interest areas to constructional slots.

| Condition | Mode | Example | Interest Area | | | Order of mention |
|---|---|---|---|---|---|---|
| | | | **A** | **B** | **C** | |
| Preposition | Typical | The poster is above the bed | Poster | Bed | – | AB |
| | Atypical | The bed is below the poster | Poster | Bed | – | BA |
| Voice | Active | The policeman arrested the thief | Policeman | Thief | Arrest | ACB |
| | Passive | The thief was arrested by the policeman | Policeman | Thief | Arrest | BCA |
| Dative | NP | The boy gave the girl a flower | Boy | Girl | Flower | ABC |
| | PP | The boy gave a flower to the girl | Boy | Girl | Flower | ACB |

participants. Specifically, after aggregating heat maps for 10 participants using Data Viewer, the IAs were specified around hot fixation areas for each individual image. IAs were relevant if they corresponded to the events described in the sentences. Each image contained 2 or 3 IAs; the IA outlines were not visible to the participants. An example is provided in Image 1 for the sentence pair *The policeman arrested the thief/The thief was arrested by the policeman.* IA A would capture the face of the thief, IA B would cover the face of the policeman, and IA C would capture the action.



**Image 1:** Fixation heat map used to determine the outline of the Interest Areas.

Data pre-processing removed all data points without eye-movements, i. e., where the first saccade entering into the IA was not available; this eliminated 33.2% of data points. An additional 1% of data points was removed because the start time of the first fixation was not available or was very short ($\leq$60 ms, excluding 250 ms for fixation point presentation). These initial data trimming steps left us with 13,451 valid data points.

In the next step, we split the dataset into three independent (i. e., non-overlapping) datasets, one for each linguistic manipulation: Preposition, Voice, and Dative. Each of these three datasets also included the data from the corresponding naturalistic image viewing. In other words, for each participant and per image we combined data on eye-movements recorded during spontaneous image viewing, and during language-guided viewing (e. g., one for eye-movements after presenting the sentence in the active voice, and another one for eye-movements after presenting the sentence in the passive voice). The Preposition dataset consisted of 2,692 data points, the Voice dataset of 4,563, and the Dative of 5,997 data points.

Following Baayen and Milin (2010), we applied a minimal a-priori trimming strategy, removing only unambiguously discontinuous data points (i. e., those that are clear extremes, leaving a solid gap between themselves and the data mass). For the Preposition dataset this trimming resulted in an additional 0.59% of data loss. Similarly, minimal loss of data was incurred in the Voice and Dative datasets, 0.37% and 0.17% respectively. This left us with final datasets of sizes $N_{preposition} = 2669$, $N_{voice} = 4546$, $N_{dative} = 5987$, totalling 13,407 data points. These three datasets were subjected to statistical analyses. Appendix 1 provides summary statistics for our three main dependent variables across the three datasets.

# 3 Results

For modelling we used the functionality of the mgcv (Wood 2006, Wood 2011) and itsadug (van Rij et al. 2016) packages in the R software environment (R Core Team 2017). The three datasets (Preposition, Voice, and Dative) were submitted individually to mixed modelling, because their combined distribution is strongly non-Gaussian and violates model assumptions. Effectively, these models assessed the significance of differences between experimentally manipulated situations within construction: naturalistic static scene viewing versus viewing when auditory information (i. e., construal) preceded scene presentation. Hence, the reported test results remind of ANOVA results, but include additional random effects where statistically justified; for this reason, model diagnostics will not be standardly reported. We preferred this statistical approach over one that

would attempt to construct the most comprehensive (combined) model, because we targeted the relationship between eye movement measures and viewing mode *within* linguistic condition. Because experimental sentences and depicted scenes were consistently manipulated (controlled) within but not between construal types the design is thus, in essence, nested. This means that statistically testable comparisons across constructions (e. g., passive voice vs. dative NP) are not further considered as they are not theoretically justified. To avoid spurious effects, we used a link-function appropriate for the different types of dependent variables. Numeric dependent variables were transformed using Box and Cox (1964) power-transformation (see also Yeo and Johnson 2000) to better approximate a normal distribution and facilitate model fitting. We report the details of the full models below in the corresponding sections. We applied model criticism (cf., Baayen and Milin 2010) to all candidate final models. Because removing influential residuals did not affect the results, we report only the full models.

As explained above (Section 2.3), our experiment consisted of three blocks. The first block established the eye-movement patterns during the visual uptake of a static scene under naturalistic conditions (i. e., image only viewing). The second and the third blocks measured the eye-movement patterns during the visual uptake of a static scene under language-guided viewing conditions. These blocks were counterbalanced per participant and the presentation order of the items was randomized within each block. Counterbalancing ensured random order of exposure to the *canonical* construal (typical preposition, active voice, and noun phrase for dative), and the atypical construal (atypical preposition, passive voice, and prepositional phrase for dative). This approach safeguards against a differential interference effect, e. g., one might expect a stronger effect of the active sentence when it is heard before the passive, but a weaker effect if the passive sentence is heard before the active. In our statistical models, a binary indicator, CanonicalFirst (1/0), encoded whether the canonical construal was presented before or after the non-canonical one. For example, when *CanonicalFirst* = 1 for voice, the active voice sentence appeared in Block 2, before the passive sentence that appeared in Block 3. Conversely, if *CanonicalFirst* = 0, the active sentence was presented in Block 3, after the passive sentence was presented in Block 2. The CanonicalFirst variable was used in all statistical models to keep the possibility of a repeated exposure effect under explicit statistical control. Furthermore, we allowed for the possibility that participants differ with respect to their 'sensitivity' to whether they first heard the canonical or the non-canonical construal – i. e., that the differential interference might vary across participants. To account for this, we included additional by-participant adjustments for CanonicalFirst.

Our analysis consists of two parts. The first part (Section 3.1) presents a pre-analysis that establishes whether our experimental manipulation was successful.

To this end we modelled the length of the gaze path (GazePathLength), which determines the amount of focus needed for information uptake between experimentally manipulated conditions, and the average pupil size per interest area (AvgPupilSize), which provides a general measure of cognitive effort. The second part (Sections 3.2 through 3.4) contains the main analyses, conducted on 3 measures, that test our hypotheses. Underlying eye-tracking research is the so-called "eye-mind hypothesis": gaze duration reveals cognitive effort, i.e., the expense incurred by processing information. We selected three indicators derived from eye-movement measurements to answer our question regarding the effect of linguistic construal on perception and conception. The order in which the Interest Areas were accessed (i.e., OrdOfAccess; Section 3.1) reveals whether interest areas were accessed in a different order, depending on the linguistic construction used to describe the event. First run gaze duration (FirstGazeDur; Section 3.2) and total gaze duration (TotalGazeDur; Section 3.3) reveal language-induced differences in salience between the IAs across experimentally manipulated situations, as described above. The selected measures occupy a different place on the scale of early versus late information uptake, with OrdOfAccess being an early measure, and TotalGazeDur a late measure, revealing the effort needed to integrate information (cf., Boston et al. 2008; Kuperman and Van Dyke 2011; Rayner 2009). FirstGazeDur falls in between the two other measures (OrdOfAccess and TotalGazeDur) and is indicative of the initial effort spent as viewing commences.

## 3.1 Pre-analysing the effect of the experimental manipulation

The analyses of GazePathLength[2] and AvgPupilSize reveal interesting trends across both models, showing an interaction of construal type (Condition) and the order in which the typical vs. atypical constructions were presented (CanonicalFirst). Figure 1 depicts the differences: (1) there is a strong within-construal consistency in the trends for both GazePathLength (downward) and AvgPupilSize (less steep and upward), and (2) there are major differences between the three types of constructions (Preposition, Voice, and Dative). From this we can conclude that there is no evidence for any differential interference effect; i.e., the general effects remain unaffected by the order in which the participants were

---

**2** To calculate gaze path length, we considered all fixations in every IA for each participant and per image. We retrieved those fixations' coordinates on the X and Y axes and calculated the Euclidean distance between two succeeding fixations. For 2D problems this essentially requires solving the Pythagorean theorem: $\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$. The total length of the gaze path (GazePathLen) is then the sum of all distances between consecutive fixation points.
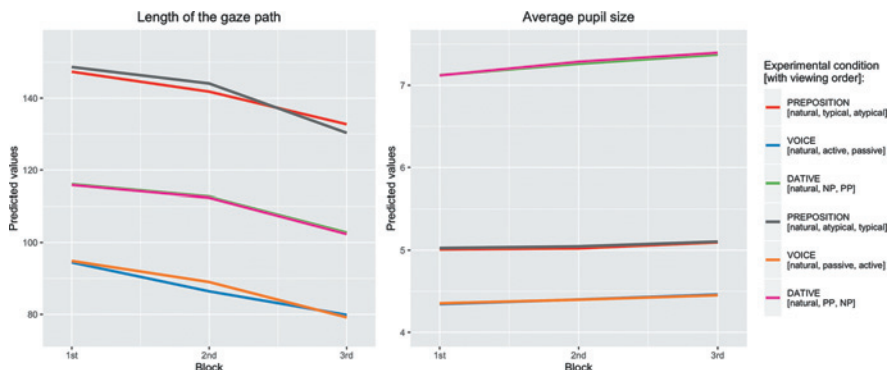
**Figure 1:** Effects of the experimental manipulation on GazePathLength and AvgPupilSize.

exposed to the typical or atypical construal of an event. Thus, our experimental design did not induce effects of repetition that are different across conditions.

Analysing the repetition in a bit more detail, we observe that the length of the gaze path shortens across experimental blocks (Figure 1, left panel). This pattern signals that participants' visual exploration becomes more efficient: since the images are repeated across blocks, the participants can take advantage of being familiar with them. This does, however, not imply that the repetition affected proper engagement with the language-mediated conditions. The average pupil size confirms this (Figure 1, right panel): here, we observe larger pupils as the experiment unfolds, which implies that participants are working harder to integrate linguistic information with their respective static scenes. Finally, there is no evidence of task habituation or fatigue towards the end of the experiment that would negatively affect our results.

## 3.2 Order of access

To determine the extent to which different types of construal affect order of access, we compared the order in which the main elements in a scene were accessed, depending on the way in which they were described. Recall that the main elements in a scene constituted an IA of their own (see Section 2.4 above). The order in which each of the IAs was accessed was calculated using the time elapsed since trial onset: the start time of the first fixation in an interest area was extracted and used as first IA access time. First entrances into each IA were then rank-transformed to reflect the IA access order per participant on a given trial.

For statistical modelling we used the ordered categorical link-function which essentially assumes a latent variable following a logistic distribution,

expressing the probability that the latent variable lies between certain cut-points (i. e., for the ordered categorical variable to be in the corresponding category; for details see Wood et al. 2016). For each alternation, a separate ordinal model was fitted predicting the order in which the IAs were accessed.

For Preposition, IA access order was predicted from one fixed effect, the IA label itself, and one random effect for intercept adjustments for items (images). For Voice and Dative, IA access order was, as with Prepositions, predicted from the IA itself, but the structure of the random effects became somewhat more complex. Voice and Dative further required by-participant adjustments for the location of the IAs in the scene and for CanonicalFirst. This revealed a complex pattern of (random) individual differences in how participants engage with the image after receiving specific priming from the auditorily presented description, in the particular order of construal option presentation (canonical/non-canonical or vice versa). Importantly, however, order of presentation did not contribute to predicting OrderOfAccess systematically (i. e., its parametric effect remained non-significant) for any of the models. For simple and higher-order comparisons we applied the t-test for differences between proportions, given that the predicted values are expressed as probabilities. To remain conservative, we made use of the combined standard error (as proposed by Baker and Nissim 1963) and of Bonferroni's correction. That is, we computed the product of raw p-value and number of comparisons: $p_{\text{Bonferroni}} = p \times m$, where $m$ represents the number of comparisons (Dunn 1961). Note that we corrected for the number of theoretically justified comparisons, not the total number of possible comparisons.

Figure 2 shows that the predicted probabilities of IAs A (more moveable item) and B (less moveable item) being accessed first or second were similar across the three viewing manipulations in the Prepositional condition. Across viewing modes, the probability of the relatively more moveable element (A) to be accessed first is higher than that of the relatively less moveable element (B); this difference is significant across all three modes, as detailed in Table 2.

**Table 2:** Predicted probabilities of order of access to IAs in the preposition condition.

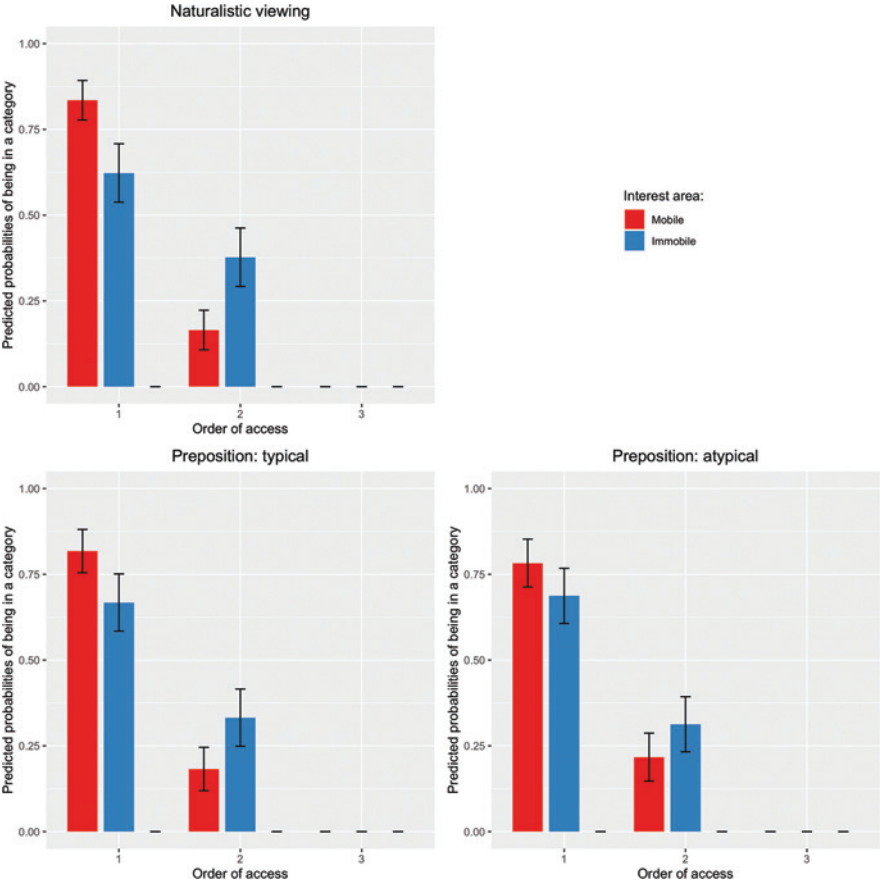| Viewing mode | Access Order | | t [p / $p_{\text{Bonferroni}}$] |
|---|---|---|---|
| | **1st** | **2nd** | |
| Natural | A | B | 8.074 [< 0.001 / < 0.001] |
| Typical Description | A | B | 5.614 [< 0.001 / < 0.001] |
| Atypical Description | A | B | 3.525 [0.001 / 0.003] |

**Figure 2:** Predicted probabilities of order of access as a function of viewing manipulation for preposition.

Second-order comparisons of the differences between the probabilities of accessing an IA sooner or later confirm these results. First, we calculated the difference between the probabilities of accessing A and B in natural viewing, typical and atypical mode ($Pr_{A\text{-}B \text{ Difference}}$). Next, we calculated the average standard error, the second-order difference, for the two differences using their respective standard error estimates. Finally, we ran a t-test for proportions to establish whether these second-order differences between the probabilities across the three conditions were significant themselves. As shown in Table 3, the A-B differences reached significance only between naturalistic viewing and viewing in atypical mode. Overall, the typical formulation appears to align with naturalistic scene viewing, strengthening the preference for the more moveable item to be accessed first. An atypical prepositional

**Table 3:** Prepositions – second-order comparisons for the preposition condition.

| Viewing Mode | Pr$_{\text{A-B Difference}}$ | t [p / p$_{\text{Bonferroni}}$] | |
| --- | --- | --- | --- |
| | | Typical Description | Atypical Description |
| Natural | 0.212 | 2.388 [0.023 / 0.069] | 4.404 [<0.001 / <0.001] |
| Typical Description | 0.150 | – | 2.023 [0.052 / 0.155] |
| Atypical Description | 0.096 | – | – |

mode neutralizes this viewing preference, as reflected in the two-fold drop, from 0.212 to 0.096, in the second-order difference ($Pr_{\text{A-B difference}}$).

Figure 3 illustrates the differences between the predicted probabilities of A (Agent), B (Patient) and C (Event) being accessed first, second or third for all
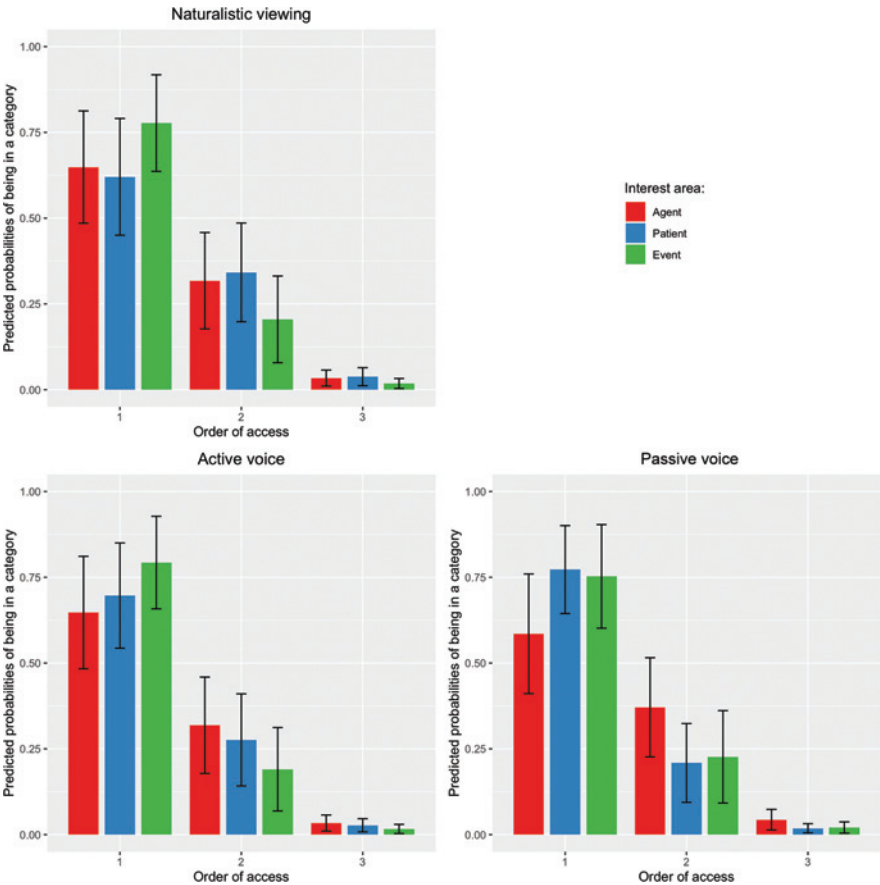


**Figure 3:** Predicted probabilities of order of access as a function of viewing manipulation for voice.

three viewing manipulations in the Voice condition. Overall, which element is most likely to be accessed first is the inverse (mirror image) of which element is most likely to be accessed second. There is little variation in which element is accessed third across viewing modes. Details of the statistical tests are presented in Table 4.

**Table 4:** Predicted probabilities of order of access to IAs in the voice condition.

| Viewing mode | Access Order | | t [p / p_Bonferroni] | |
|---|---|---|---|---|
| | **1st** | **2nd/3rd** | | |
| Natural | C | B/A | 2.328 [0.027 / 0.159] (A-C) | 2.781 [0.008 / 0.050] (B-C) |
| Active | C | B/A | 2.689 [0.011 / 0.065] (A-C) | 1.848 [0.072 / 0.434] (B-C) |
| Passive | B[1]/C | A | 3.392 [0.001 / 0.008] (A-B) | 2.845 [0.007 / 0.042] (A-C) |

[1]In passive viewing mode, the probability of the Patient (B) being accessed first is significantly higher than in natural viewing mode ($p_{Bonferroni}$ = 0.026).

In natural viewing mode, depicted in the leftmost panel of Figure 3, the predicted probabilities of the Agent (A) or Patient (B) being accessed first are significantly lower (before Bonferroni correction), than those of the Event (C) being accessed first. A similar situation is observed in the active voice, depicted in the middle panel but here the difference is weaker. In the passive voice, depicted in the rightmost panel, the predicted probability of an element being accessed first if it is the Patient or the Event is significantly higher than if it is the Agent, while the difference between the Patient and the Event is not significant. In passive viewing mode, the probability of the Patient (B) being accessed first is also significantly higher than in natural viewing mode.

The second-order comparisons of the A-B differences, presented in Table 5, confirm these findings: the second order differences between naturalistic viewing and passive-primed conditions are significant, even after Bonferroni correction.

**Table 5:** Second-order comparisons for the voice condition.

| Viewing mode | Pr_A-B Difference | Passive |
|---|---|---|
| | | t [p / p_Bonferroni] |
| Natural | 0.028 | 2.764 [0.009 / 0.026] |
| Active | 0.049 | 2.467 [0.019 / 0.057] |
| Passive | 0.187 | – |

Those between active and passive voice are likewise significant, and remain marginally significant after Bonferroni correction.

Figure 4 shows the differences between the predicted probabilities of IAs Agent (A), Recipient (B) and Object (C) being accessed first, second or last between the naturalistic and language-mediated viewing manipulations in the Dative condition; Table 6 presents the test results. In natural viewing mode, depicted in the left-most panel, the predicted probability that the Agent (A) or the Recipient (B) will be accessed first is significantly higher than for the Object (C).[3] The Dative PP mode
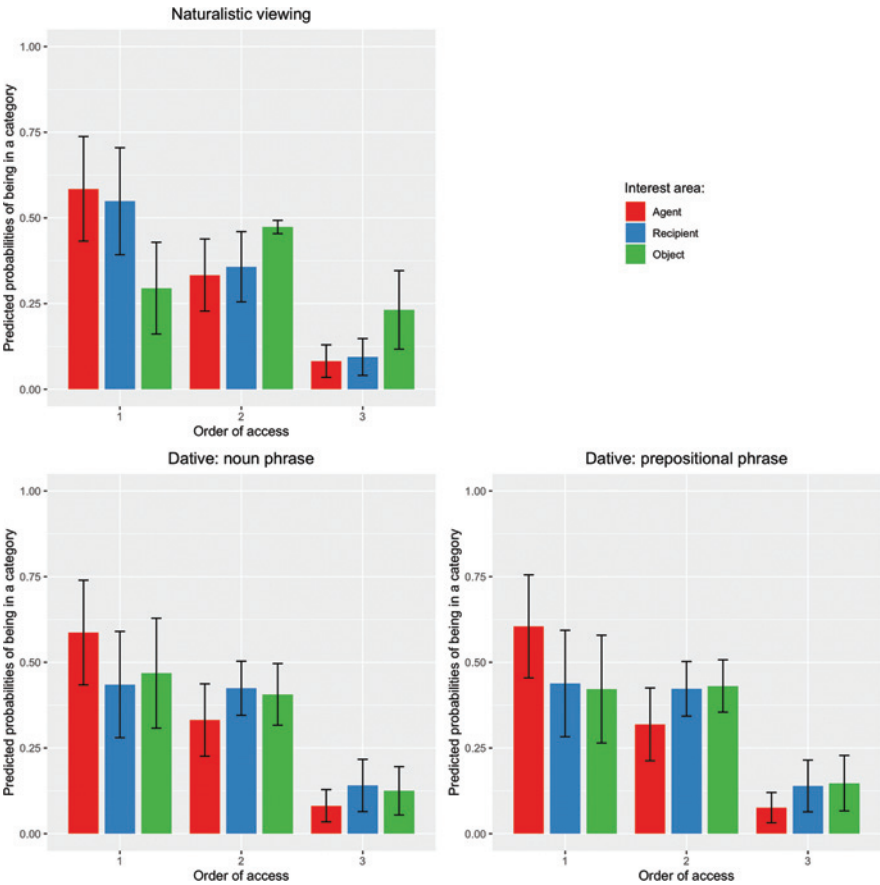


**Figure 4:** Predicted probabilities of order of access as a function of viewing manipulation for dative.

---

**3** Mirroring this pattern of effects, we see that the Object (C) is significantly more likely to be accessed second than the Agent (A) ($t = 5.143$, $p < 0.001$, $p_{Bonferroni} < 0.001$) and the Recipient (B) ($t = 4.362$, $p < 0.001$, $p_{Bonferroni} < 0.001$).

**Table 6:** Predicted probabilities of order of access to IAs in the dative condition.

| Viewing mode | Access order | | $t$ [$p$/$p_{Bonferroni}$] | |
|---|---|---|---|---|
| | 1st | 2nd/3rd | Comparison 1 | Comparison 2 |
| Natural | A/B | C | 5.592 [<0.001 /<0.001] (A-C) | 4.895 [<0.001 /<0.001] (B-C) |
| PP | A | B/C | 3.026 [0.004 / 0.041] (A-B) | 3.300 [0.002 / 0.017] (A-C) |
| NP | A/B/C | – | – | – |

is depicted in the right-most panel. The predicted probabilities for the Agent (A) to be accessed first are significantly higher than those for the Recipient (B) and for the Object (C). Within the Dative NP mode, depicted in the middle panel, the probabilities of A, B and C being accessed first are not significantly different (with the difference between Agent and Recipient being marginal: $p_{Bonferroni}$ = 0.095).[4]

There is no significant difference in order of access between the two dative modes on second-order comparison, as shown in Table 7. Second-order comparisons for first access between IAs B and C (Recipient and Object), which change position of mention in NP vs. PP, revealed a significant difference between naturalistic viewing and viewing after NP and after PP but not for viewing after NP vs. PP. We do again observe a decreasing trend, this time from naturalistic over NP to PP condition ($Pr_{A-C\ difference}$ = 0.254, 0.033, 0.016), revealing a degree larger difference in naturalistic than in language-mediated conditions.

**Table 7:** Second-order comparisons for the dative condition.

| Viewing mode | $Pr_{B-C\ Difference}$ | NP | PP |
|---|---|---|---|
| | | $t$ [$p$ / $p_{Bonferroni}$] | $t$ [$p$ / $p_{Bonferroni}$] |
| Natural Viewing | 0.254 | 4.030 [<0.001 /<0.001] | 4.362 [<0.001 /<0.001] |
| NP | 0.033 | – | 0.299 [0.381 / 1.0] |
| PP | 0.016 | – | – |

## 3.3 First run gaze duration

In order to test whether and how linguistic construal can affect and modulate which elements attract attention in a static scene, first run gaze durations were compared across viewing modes for each condition. The first run gaze duration

---

4 Interestingly, however, even though the Object (C) is named last, it is significantly more likely to be accessed first in NP mode than after naturalistic viewing ($t$ = 3.251, $p$ = 0.002, $p_{Bonferroni}$ = 0.020).

is the summation of all fixations that occurred before the viewer moved out of an interest area for the first time. We expect the length of the first run gaze duration to reflect the extent to which the linguistic description of the event promotes or demotes elements in the scene. Separate mixed effect models were fit to each of the three datasets. The same two independent variables, ViewingMode and InterestArea, were considered as fixed effects for all three conditions but different modelling solutions were retained, as described below. All models contained two random effects: image and smooths of participants across experimental trials.

For Preposition, the model containing an interaction between ViewingMode and InterestArea turned out to be the most robust. A Wald test revealed a complex pattern of significant contrasts, visualized in Figure 5. Overall, first looks to A and to B are significantly longer in the language-mediated viewing modes than in naturalistic viewing. First looks to the more moveable item A are significantly longer after typical description than in naturalistic viewing (*Chi-sq.* = 31.089, $p < 0.001$, $p_{Bonferroni} < 0.001$) or after atypical description (*Chi-sq.* = 14.800, $p < 0.001$, $p_{Bonferroni} < 0.001$). First looks to the more moveable item A are also significantly longer than first looks to the less moveable item B following a typical description (*Chi-sq.* = 27.144, $p < 0.001$, $p_{Bonferroni} < 0.001$), but this difference is only marginally significant in naturalistic viewing (*Chi-sq.* = 6.129, $p = 0.013$, $p_{Bonferroni} = 0.080$). After an atypical description, where the more moveable item A is named last, A is looked at shorter than after a typical description (see above: *Chi-sq.* = 14.800, $p < 0.001$, $p_{Bonferroni} < 0.001$), and the less moveable item B is looked at longer than in the naturalistic viewing mode (*Chi-sq.* = 22.416, $p < 0.001$, $p_{Bonferroni} < 0.001$). Finally, across the language-mediated viewing modes and across IAs, the longest first looks are to the more
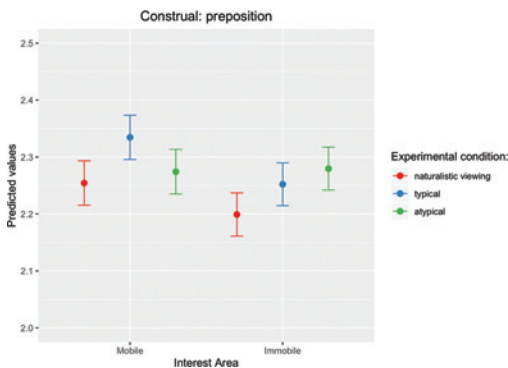


**Figure 5:** First run gaze duration for Preposition.

moveable item A after typical description (where A is named first) and to the less moveable item B after atypical description (where B is named first), and that difference is also significant (*Chi-sq.* = 12.005, $p < 0.001$, $p_{Bonferroni} = 0.003$).

The best model for Voice contained ViewingMode and InterestArea as main effects, without interaction; this model is depicted in Figure 6. The first run gaze duration differed significantly across all IAs within the naturalistic viewing mode (all $p_{Bonferroni} \leq 0.01$), and within both language-mediated viewing modes (all $p_{Bonferroni} \leq 0.01$). Within IAs but across viewing modes (naturalistic, active, and passive), differences were significant between naturalistic viewing and the language-mediated viewing modes (all $p_{Bonferroni} < 0.001$) but not between Active and Passive. Overall, the Event (C) attracted shorter first looks than the Patient (B) which attracted shorter looks than the Agent (A).
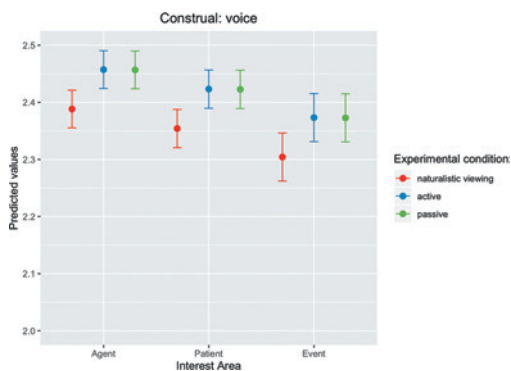


**Figure 6:** First run gaze duration for Voice.

Similar to the best model for Voice, the best Dative model contained ViewingMode and InterestArea as main effects, without interaction. The findings are presented in Figure 7. Here too, the first run gaze duration differed significantly across the three IAs, within the naturalistic viewing mode and within each of the two language-mediated viewing modes. For each of the three viewing conditions, the Object (C) was fixated longest and the Recipient (B) was looked at shortest on first run (all $p_{Bonferroni} < 0.001$). The difference between first fixation durations on the Agent (A) and the Object (C) was somewhat less pronounced but remained robustly significant (all $p_{Bonferroni} < 0.01$). Yet, there were no significant differences in first run gaze duration between viewing after NP versus PP description of the situation for any of the IAs. Both language-mediated conditions (NP and PP) did require significantly longer first gaze durations than did naturalistic viewing mode (all $p_{Bonferroni} \leq 0.001$).
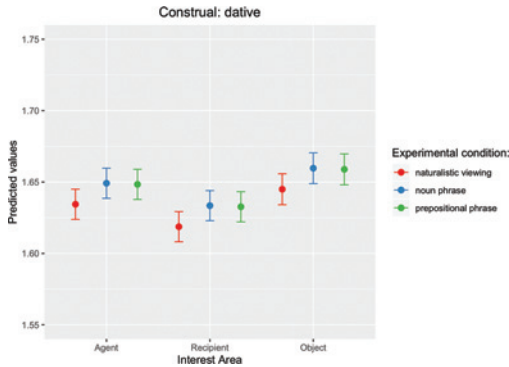
**Figure 7:** First run gaze duration for Dative.

## 3.4 Total gaze duration per interest area

A second way to establish the extent to which different types of linguistic construal modulate scene perception relies on total gaze duration per IA. Here too, more pronounced highlighting of scene components by the linguistic description is expected to extend the total gaze duration. The total gaze duration is the sum of the duration across all fixations in an IA, capturing the total length of time spent in an IA while viewing the scene. This measure is an indication of total processing effort per IA. The same two independent variables, ViewingMode and InterestArea, were considered as fixed effects for all three conditions but different solutions were retained. All models contained two random effects: image and a by-participant factorial smooth for TrialOrder.

    For Preposition, the model with an interaction between ViewingMode and InterestArea turned out to be the most robust. The Wald tests revealed a complex pattern of differences. Figure 8 shows that, for each of the three viewing modes, the more moveable item A attracts the viewer's overall gaze significantly longer than the less moveable item B (all $p_{\text{Bonferroni}} < 0.001$), but the difference between A and B is not significant after an atypical formulation (*Chi-sq.* = 3.7942, $p = 0.051$, $p_{\text{Bonferroni}} = 0.308$) when the more moveable item A is named last. More specifically, the more moveable item A is looked at longer after a typical formulation (when A is named first) than in naturalistic viewing (*Chi-sq.* = 7.3423, $p = 0.007$, $p_{\text{Bonferroni}} = 0.040$) while the less moveable item B is looked at longer after an atypical formulation (when B is named first) than in naturalistic viewing (*Chi-sq.* = 7.0819, $p = 0.008$, $p_{\text{Bonferroni}} = 0.047$). A is also looked at marginally longer after a typical than after an atypical description of the situation (*Chi-sq.* = 5.758, $p = 0.016$, $p_{\text{Bonferroni}} = 0.098$), i. e., when it is named first.
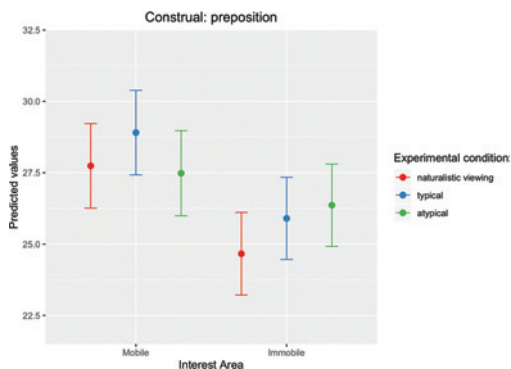
**Figure 8:** Total gaze duration for Preposition.

The best model for Voice contained ViewingMode as sole fixed effect, and is therefore not presented visually. The total gaze durations differ between natural viewing versus viewing after active (*Chi-sq.* = 4.748, *p* = 0.029) and after passive (*Chi-sq.* = 4.406, *p* = 0.036) formulations. In other words, language-guided viewing modes attract the viewer's gaze for longer overall.[5] There is, however, no significant difference between the active and passive viewing modes (*p* = 1.0).

The best model for the Dative contained ViewingMode and InterestArea as main effects, without interaction. Figure 9 shows the effects that emerge. Total gaze



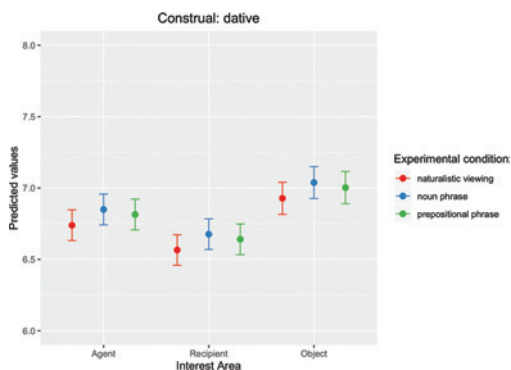**Figure 9:** Total gaze duration for Dative.

---

[5] The difference between naturalistic viewing and viewing after a passive formulation is the only one, across all models, which becomes marginally significant after trimming: *Chi-sq.* = 3.410, *p* = 0.065.

durations for the Agent (A), Recipient (B) and Object (C) are all, pair-wise, significantly different within naturalistic viewing mode (all $p_{\text{Bonferroni}} < 0.001$), after NP dative and after PP dative (all $p_{\text{Bonferroni}} < 0.001$). NP and PP datives do not differ significantly from each other in any of the IAs, however, after Bonferroni correction.

# 4 Discussion

Construal is one of the central hypothetical constructs of Cognitive Linguistics. It captures the flexibility that language offers in describing an event or a situation and it relates differences in linguistic description to differences in conceptualisation (i. e., representation) of the event or situation. In the present study we set out to confirm empirically whether and how a different construal invokes different conceptualisations of what is being communicated.

As part of a visual-world study, 60 participants viewed 48 static everyday scenes, first without description, and then again after hearing one of two variants of the Preposition, Voice or Dative alternation. Our prediction was that differences in construal will induce specific differences in eye-movements during the visual inspection of images: (a) generally, we expected observable differences between patterns of eye-movements in naturalistic viewing of a scene (the default experimental situation that was always the first block of trials) and in viewing the same scene following auditory presentation of a description of the scene; (b) more specifically, we expected observable differences in patterns of eye-movements between both verbal descriptions of the very same situation depicted in the images. Furthermore, different types of alternations should do so to different degrees: the theory predicts not just an effect but a cline of effects, ranging from a pronounced effect of construal in the prepositional alternation, via a more attenuated effect in the voice alternation, to a very subtle effect for the dative alternation. In other words, construal will trigger both qualitative and quantitative differences in conceptualization.

For this particular study we pre-selected three eye-movement measures to address our research questions and test our predictions. First, the order in which the Interest Areas (IAs) were accessed was used to establish the extent to which different types of construal affect perception. We used first run gaze duration and total gaze duration, typically taken to indicate 'integrative' efforts in information processing, to shed light on the language-induced change in salience of the elements of the visual scene, both between the IAs and across experimentally manipulated viewing modes. Together these three measures facilitate

analysing the complex and intricate pattern of effects that construal has on visual information processing. By extension, they contribute to understanding the fascinating interaction of perception and language.

Below, we will summarize and interpret the main findings with a view to presenting a unified account of how language construal profiles information uptake from a visually presented scene by making the elements that act as guiding cues more or less salient. Furthermore, we will argue that this change in salience and the way we process information is highly likely to affect what will be retained in memory and to what degree. This could then be, in the limit, an endless loop in which constructions, and our comprehension and/or production of them, will be affected by memory traces that are themselves affected by perceptual acts that are in turn affected by the plasticity of language, in describing experience and experience as expressed by construal. Such an endless loop, certainly, reminds us of how our "[mind] organizes the world by organizing itself" (Piaget 2013: 355), where adaptive pressures are realised through perpetual, on-going processes of assimilation (of new to existing information) and of accommodation (of existing to new information).

## 4.1 A cline of construal

We predicted a cline in the strength of the effect that the details of the linguistic description would have on the specifics of static scene perception. We expected a pronounced effect in the prepositional alternation, a more attenuated effect in the voice alternation, and a very subtle effect for the dative alternation. The cline would show in the number of attested eye-movement measures and in the magnitude of these effects: the preposition alternation would show stronger effects across more measures, while the dative alternation would show weaker effects across fewer measures. Our findings, summarized in Table 8, confirm these expectations. Supplementary Material B contains the details of a Logistic Generalised Additive Mixed Model that confirms the existence of a cline in the strength of the effects for Order of Access.

**Table 8:** Summary of attested differences between construal types across the three eye-tracking measures.

| Measure | Preposition | Voice | Dative |
| --- | --- | --- | --- |
| Order of Access | Attested | Attested | Not attested |
| First Gaze Duration | Attested | Not attested | Not attested |
| Total Gaze Duration | Attested | Not attested | Not attested |

### 4.1.1 Preposition

For Preposition, we established a modulatory effect of construal on visual information uptake. More specifically, an atypical construal makes the more easily moveable element less salient, as witnessed by changes in both early and late measures.

The order in which the IAs are accessed (i. e., OrderOfAccess), an early effect, is affected by construal: the atypical formulation modifies inspection of the picture. The more moveable item is significantly more likely to be visually inspected first in spontaneous (i. e., naturalistic) viewing and after typical formulation. Yet, after atypical formulation this effect is neutralized and the more moveable item is no longer more likely than the less moveable item to be accessed first. Like OrderOfAccess, FirstGazeDuration appears similar in naturalistic viewing and in viewing after a typical formulation, with the first-named easily moveable element receiving longer first looks, but typical formulation amplifies these differences. In viewing after atypical formulation, the difference in first looks to the more- vs. less-easily moveable element is, again, attenuated. In terms of OverallGazeDuration, an indicator of information integration efforts, after a typical description of the spatial relation between the objects in the scene, the more easily moveable element attracts the viewer's gaze for longer than the less easily moveable element. However, this difference becomes only marginally significant after an atypical formulation which mentions the less easily moveable element first.

### 4.1.2 Voice

In the voice alternation, the second in terms of strength of the predicted effect on static image viewing, we observe the differential effect of active vs. passive construal in the early measure, OrderOfAccess, only. Although the Event is significantly more likely to be accessed first in naturalistic viewing, this effect is less pronounced after an active voice description and no longer significant after a passive voice description; after a passive voice description, the Patient is equally likely to be accessed first. On the later measures, FirstGazeDuration and OverallGazeDuration, we observe no construction-specific effects, although there are effects of language that apply irrespective of active or passive voice. FirstGazeDuration to the Agent is longer than that to the Patient, which is, in turn, longer than that to the Event, and these differences are observed within each mode. They are thus enhanced by language but not further affected by the details of the linguistic construal. OverallGazeDuration is slightly longer when participants receive additional auditory information. This confirms the results of

our pre-analyses, where language-mediated conditions required additional cognitive effort.

This pattern of effects forms an interesting contrast: (a) while the Event may be accessed first, it simultaneously receives the shortest FirstGazeDuration; (b) while the IAs attract longer first and total gaze duration after language mediation, there are no construal-induced differences. This means that no elements in the scene are automatically made more salient by construal. The lack of differences in measures of information uptake between viewing after active or passive description are also interesting given the literature on the voice alternation, which has highlighted an increased difficulty in producing and comprehending passives.

### 4.1.3 Dative

For the subtlest differences in information expression, represented by the dative alternation, we observe a general effect of language only. Under naturalistic viewing conditions the Agent and Recipient have equal chances of being accessed first, while in both language-mediated modes, the Agent is accessed first, and the Recipient and Object have an equal chance of being accessed second. However, there is no significant difference in order of access between Recipient and Object for NP versus PP, even though that is where construal affects order of mention. The later measures, FirstGazeDuration and OverallGazeDuration, show the same general effect of language only, between gaze duration in naturalistic viewing and in the language-guided modes. Across all three modes, the Agent receives the longest first looks, followed by the Recipient and then the Object. But overall, the Object is looked at longest in total, followed by the Agent and the Recipient. These differences are enhanced in language-mediated viewing modes (and significantly so in the NP), but are not significantly altered by the construction chosen to describe the event.

## 4.2 Does language determine or modulate scene inspection?

In this paper we were interested in the question of what types of differences in linguistic encoding induce differences in visual inspection of a scene. Overall, a complex picture of effects emerges that supports the hypothesized cline from prepositional alternation, over voice alternation to dative alternation. While the prepositional alternation shows both early and late effects that mirror the order of mention, the dative alternation shows early effects of construal only,

and these do not reflect the change in order of mention. The voice alternation falls in between, with early effects of change in order of mention on order of access.

The cline in the strength of the effect that linguistic construal has on perception, and by extension conception, is also visible in the obviousness or subtlety of the linguistic mechanisms involved in a given alternation. The preposition alternation, in which two objects are described in spatial relation to each other, was introduced as baseline alternation. Previous findings reported that viewers look at the elements of a scene that are mentioned in the describing sentence (Altmann and Kamide 1999; Cooper 1974), and that speakers focus on those elements in a consistent order which is, afterwards, preserved in the sentence they use to describe that same scene (Griffin and Bock 2000; Myachykov et al. 2013). In that respect, the preposition alternation, with its straightforward description of the visual scene, was most likely to show effects of construal. After all, both the typical and atypical description denote the same spatial configuration. The difference consists in the presentation of the two objects in the scene, once as 'Figure', another time as 'Ground'.

In the voice alternation, the thematic roles of Agent and Patient are constant, yet variably assigned to grammatical functions. While the Agent occupies subject position and the Patient occupies object position in the active voice, their positions are switched in the passive voice and the Patient ends up in subject position, while the Agent may be omitted altogether. Hence, the choice of voice does not affect the event type, but it changes the focus on the participants in a given event. The active voice focuses on the active involvement of an Agent, while the passive voice defocuses the agentive participant in the event and refocuses on the passive involvement of the Patient. Yet, passives are only available for actives with an agentive agent, putting a semantic constraint on passive syntax (Ambridge et al. 2016; Pinker 1989).

In case of the dative alternation, much of the traditional literature has assumed that both constructions are semantically equivalent: dative shifting in English is considered a stylistic, discourse-pragmatic device (Givón 1984: 153). Goldberg (2005), on the other hand, promoted the idea that both constructions are semantically different, but statistical modelling of naturalistic data (Bresnan et al. 2007; Bresnan and Ford 2010) was needed to unravel the interplay of phonological, morphological and semantic variables that govern the probabilistic choice for one option over another. From this it can be concluded that, at the very least, any semantic differences are rather subtle. In the case of the dative alternation, construal appears to be more of an expressive device that is informative about discourse preferences but does not achieve refocusing of the elements in a scene. This interpretation is supported by our

pre-analyses (Section 3.1.) which showed that the Dative triggers the largest pupil sizes but requires only average gaze path length. In other words, the high cognitive processing costs that come with the Dative alternation are not imposed by the visual properties of the scene but by the linguistic properties of the description.

## 4.3 To what better end?

These findings are important for linguistic theorizing. Linguists are interested in the question of which aspects of our (linguistic) experience make it into (long-term) memory. Our experience is far too rich for us to remember every detail, and between experience and memory lies a process of attention and encoding, which makes memory not directly reflective of frequency (Pierrehumbert 2006) or active/passive decay of memory traces (cf., Hebb 1949; Underwood 1957; for recent findings and discussion see Mirković et al. 2019). The results we have obtained here shed light on the kinds of differences in linguistic encoding that affect attention for a specific aspect of a scene, thereby increasing the chances for those aspects of experience to leave traces in memory and to become gradually more entrenched. Language, in other words, reduces the uncertainty about which components of a scene are important and, eventually, which ones should be remembered.

Given the set-up of our study, we are looking at the very first stages of memory, but in order for something to make it into long term memory, it will need to have been retained at this first stage. As explained in Section 2, the average auditory stimulus duration was 2600 ms (range: 1877–3672 ms). The auditory stimulus was followed by a brief 250 ms fixation on a central point, before the image appeared and remained visible for 3500 ms. On average, 60 ms were needed from the time the image appeared before the eyes fixated for the first time. Although a stimulus can persist in auditory memory for up to 3 or 4 seconds, in most cases, only the end of the stimulus will have been available in echoic memory by the time the participants started to view the scene. Thus, at late viewing time ($\sim 6000$ ms) participants will have been relying on processed information available in working memory only. As reported in Section 3 and discussed above in Section 4.1, not all linguistic manipulations are equally successful in affecting attention, and we observed early and late effects of linguistic manipulation on visual information uptake. While the prepositional alternation shows both early and late effects that clearly mirror the linguistic manipulation, the voice alternation shows early effects of construal on order of access only and the dative alternation shows

some early effects that do not map straightforwardly onto the details of the linguistic manipulation.

The strength of the effect that linguistic description has on scene perception determines the claims language can lay to affecting visual information uptake and hence conceptualization of a static scene (Huettig et al. 2011; Lupyan 2012; Lupyan and Lewis 2017). In the case of Preposition, the effects of language on perception were clear and strong. It seems convincing to assume that the two reinforce each other mutually, and work together during information uptake and processing. Language guides attention, thereby influencing how a scene is perceived, and shaping the experience that is committed to memory. In the case of the Dative, the effects of language on perception were weak and indistinct. It is, thus, much less plausible to postulate some (or any) level of mutual reinforcement. Here, language modulates attention by highlighting aspects of a scene, while not fundamentally changing how the scene is perceived and committed to memory. The linguistic preferences, although entrenched in memory (Bresnan and Ford 2010), remain restricted to the level of the code, as some sort of 'larpurlartistic' device that, essentially, does not alter the uncertainty of the message (by either increasing or decreasing it substantially) but, perhaps, serves to build plasticity of the communicative system – the language.

# 5 Conclusions

We started from the observation that language provides a variety of ways to express events. On a Cognitive Linguistic approach to meaning, the choice of describing a scene in one way rather than another is not fully dictated by the scene itself but reflects the speaker's conceptualization of the scene. The theoretical concept of construal is used to account for alternating ways of describing and thereby "construing" a situation. To test the reach of construal experimentally, we examined whether alternative constructions could evoke different conceptualizations of a situation in the hearer across the larger population of language users. Relying on the concept of ception – which conjoins conception and perception – as operationalization of the idea, we employed a Visual World Paradigm to establish which aspects of linguistic scene description modulate visual scene perception, thereby affecting event conception.

We obtained support for a modulatory role of construal. We found that construal affects the spontaneous visual perception of a scene by affecting the order in which the components of a scene are accessed or by modulating the distribution of attention over the components, making them more or less

interesting and salient. Which effect was found depended on the type of linguistic manipulation, with stronger linguistic manipulations triggering stronger perceptual effects. Our findings thus also support our hypothesis that there would be a cline in the strength of the effect that language has on perception. The strength of the effect that linguistic description has on scene perception determines the claims language can lay to affecting visual information uptake and hence conceptualization of a static scene.

**Conflict of Interest:** John Newman was appointed as Editor-In-Chief for the review purposes of this article. The submission was blinded in ScholarOne and the standard double-blind peer-reviewed procedure was adhered to.

# Datasets and code

Datasets and code underlying the experiment can be found at the following links:
R-scripts: https://github.com/ooominds/Construal_in_language
Data files: https://doi.org/10.25500/edata.bham.00000385

# Appendix 1: Descriptive statistics for the three main dependent variables

| Variable name | Preposition | | | | |
|---|---|---|---|---|---|
| | **M** | **Mdn** | **sd** | **IQR** | **Min/Max** |
| OrdOfAccess* | 1 | 1 | – | – | 1/2 |
| FirstGazeDur | 2.27 | 2.26 | 0.23 | 0.34 | 1.6/2.8 |
| TotalGazeDur | 27.09 | 27.46 | 8.11 | 11.54 | 6.1/47.4 |
| | Voice | | | | |
| | **M** | **Mdn** | **sd** | **IQR** | **Min/Max** |
| OrdOfAccess* | 1 | 1 | – | – | 1/3 |
| FirstGazeDur | 2.40 | 2.41 | 0.26 | 0.37 | 1.5/3.1 |
| TotalGazeDur | 28.40 | 28.53 | 7.99 | 11.36 | 7.9/49.8 |
| | Dative | | | | |
| | **M** | **Mdn** | **sd** | **IQR** | **Min/Max** |
| OrdOfAccess* | 1 | 1 | – | – | 1/3 |
| FirstGazeDur | 1.64 | 1.64 | 0.10 | 0.13 | 1.3/1.9 |
| TotalGazeDur | 6.80 | 6.80 | 1.29 | 1.73 | 2.6/10.6 |

*For accessing order of interest areas (OrdOfAccess), a rank-type of variable, we provide the Mode (i. e., the most frequent value) instead of the Mean, the Median, and the range of values, only.

# References

Allopenna, Paul D., James S. Magnuson & Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38. 419–439.

Altmann, Gerry T. M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73. 247–264.

Ambridge, Ben, Amy Bidgood, Julian Pine, Caroline Rowland & Daneil Freudenthal. 2016. Is passive syntax semantically constrained? Evidence from adult grammaticality judgment and comprehension studies. *Cognitive Science* 40(6). 1435–1459.

Baayen, R. Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28.

Bacon, William F. & Howard E. Egeth. 1994. Overriding stimulus-driven attentional capture. *Perception & Psychophysics* 55(5). 485–496.

Baker, R. W. R. & J. A. Nissim. 1963. Expressions for combining standard errors of two groups and for sequential standard error. *Nature* 198(4884). 1020.

Boston, Marisa Ferrara, John Hale, Reinhold Kliegl, Umesh Patil & Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence Corpus.

Box, George E. P. & David R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)* 26(2). 211–243.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 186–213.

Buswell, Guy Thomas. 1935. *How people look at pictures: A study of the psychology of perception in art*. Chicago: The University of Chicago Press.

Cooper, Roger M. 1974. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6(1). 84–107.

Coventry, Kenny R., Dermot Lynott, Angelo Cangelosi, Lynn Monrouxe, Dan Joyce & Daniel C. Richardson. 2010. Spatial language, visual attention, and perceptual simulation. *Brain & Language* 112(3). 202–213.

Croft, William & Alan D. Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dahan, Delphine, James S. Magnuson & Michael K. Tanenhaus. 2001. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology* 42. 317–367.

Desimone, Robert & John Duncan. 1995. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18. 193–222.

Dunn, Olive J. 1961. Multiple comparisons among means. *Journal of the American Statistical Association* 56(293). 52–64.

Egeth, Howard E. & Steven Yantis. 1997. Visual attention: Control, representation, and time course. *Annual Review of Psychology* 48. 269–297.

Givón, Talmy. 1984. *Syntax: A functional-typological introduction*. Amsterdam: John Benjamins.

Goldberg, Adele E. 2005. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

Griffin, Zenzi M. & Kathryn Bock. 2000. What the eyes say about speaking. *Psychological Science* 11(4). 274–279.

Hebb, Donald Olding. 1949. *The organisation of behavior: A neuropsychological theory*. New York: Wiley.

Huettig, Falk, Joost Rommers & Antje S. Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137(2). 151–171.

Hwang, Heeju & Elsi Kaiser. 2009. The effects of lexical vs. perceptual primes on sentence production in Korean: An online investigation of event apprehension and sentence formulation. Paper presented at the 22nd CUNY conference on sentence processing, Davis, CA.

Kuperman, Victor & Julie A. Van Dyke. 2011. Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language* 65(1). 42–73.

Langacker, Ronald. 1987. *Foundations of cognitive grammar. Volume 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, Ronald. 2007. Cognitive grammar. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The oxford handbook of cognitive linguistics*, 421–462. Oxford: Oxford University Press.

Langton, Stephen R. H., Anna S. Law, Mike Burton & Stefan R. Schweinberger. 2008. Attention capture by faces. *Cognition* 107(1). 330–342.

Lindsay, Shane, Christoph Scheepers & Yuki Kamide. 2013. To dash or dawdle: Verb-associated speed of motion influences eye movements during spoken sentence comprehension. *Plos One* 8(6). e67187.

Lupyan, Gary. 2012. Linguistically modulated perception and cognition: The label feedback hypothesis. *Frontiers in Psychology* 3(54). 1–13.

Lupyan, Gary & Molly Lewis. 2017. From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*. 34(10). 1319–1337.

Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324.

Mirković, Jelena, Lydia Vinals & M. Gareth Gaskell. 2019. The role of complementary learning systems in learning and consolidation in a quasi-regular domain. *Cortex* 116. 228–249.

Myachykov, Andriy, Christoph Scheepers, Simon Garrod, Dominic Thompson & O. Fedorova. 2013. Syntactic flexibility and competition in sentence production: The case of English and Russian. *The Quarterly Journal of Experimental Psychology* 66(8). 1601–1619.

Myachykov, Andriy, Dominic Thompson, Christoph Scheepers & Simon Garrod. 2011. Visual attention and structural choice in sentence production across languages. *Language and Linguistics Compass* 5(2). 95–107.

Myachykov, Andriy & Russell Tomlin. 2008. Perceptual priming and syntactic choice in Russian sentence production. *Journal of Cognitive Science* 9(1). 31–48.

Parkhurst, Derrick, Klinton Law & Ernst Niebur. 2002. Modeling the role of salience in the allocation of overt visual attention. *Vision Research* 42(1). 107–123.

Piaget, Jean. 2013. *The construction of reality in the child*. London: Routledge.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 34(4). 516–530.

Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, Keith. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62(8). 1457–1506.

Roland, Douglas W., Frederic D. Dick & Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3). 348–379.

Rubin, Edgar. 1921. *Visuell wahrgenommene Figuren: Studien in psychologischer Analyse*. Kobenhaven: Gyldendal.

Shank, Matthew D. & James T. Walker. 1989. Figure-ground organization in real and subjective contours: A new ambiguous figure, some novel measures of ambiguity, and apparent distance across regions of figure and ground. *Perception & Psychophysics* 46(2). 127–138.

Talmy, Leonard. 1978. Figure and ground in complex sentences. In Joseph Greenberg (ed.), *Universals of human language. Volume 4. Syntax.*, 625–649. Stanford: Stanford University Press.

Talmy, Leonard. 1988. The relation of grammar to cognition. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics* (Current Issues in Linguistic Theory). Amsterdam: John Benjamins.

Talmy, Leonard. 2000. *Toward a cognitive semantics. Volume 1. Concept structuring systems*. Cambridge, MA: MIT Press.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–1634.

Thompson, Dominic. 2012. *Getting at the passive: Functions of passive-types in English*. Glasgow: University of Glasgow.

Tomlin, Russell. 1995. Focal attention, voice, and word order. In P. Downing & M. Noonan (eds.), *Word order in discourse*, 517–552. Amsterdam: John Benjamins.

Tomlin, Russell & Andriy Myachykov. 2015. Attention and salience. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics* (Handbooks of Linguistics and Communication Science), 31–52. Berlin: De Gruyter Mouton.

Underwood, Benton J. 1957. Interference and forgetting. *Psychological Review* 64(1). 49.

van Rij, J., M. Wieling, R. Harald Baayen & H. van Rijn. 2016. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.

Verhagen, Arie. 2007. Construal and Perspectivization. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 48–81. Oxford: Oxford University Press.

Von Ehrenfels, C. 1890. Ueber "Gestaltqualitaeten". *Vierteljahrsschrift fuer wissenschaftliche Philosophie* 14. 249–292.

Wever, Ernest Glen. 1927. Figure and ground in the visual perception of form. *The American Journal of Psychology* 38(2). 194–226.

Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton: CRC Press.

Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(1). 3–36.

Wood, Simon N., Natalya Pya & Benjamin Säfken. 2016. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* 111(516). 1548–1563.

Yarbus, Alfred L. 1967. *Eye movements and vision*. New York: Plenum Press.

Yeo, In-Kwon & Richard A. Johnson. 2000. A new family of power transformations to improve normality or symmetry. *Biometrika* 87(4). 954–959.