Title
Visual cueing affects the processing of grammatical structure: A self-paced reading study on non-canonical word order in Bulgarian

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Author
Compensis, Paul

Publication Date
2023

Peer reviewed

Visual cueing affects the processing of grammatical structure: A self-paced reading study on non-canonical word order in Bulgarian

Paul Compensis (paul.compensis@uni-koeln.de)
Department of German Language and Literature I, University of Cologne, Germany

Abstract

Visual attention can influence how language is processed. For instance, previous research showed that visual cueing of a referent affects the choice of a particular grammatical structure. The present study extended this research to language comprehension and investigated whether visually cueing either the subject or the object of an event affects the interpretation of a textual event description in Bulgarian with either a subject-initial or object-initial word order. The presence of a visual cue matching the sentence-initial argument decreased self-paced reading reaction times and altered the accuracy in response to a comprehension question. Differences were found depending on the gender of the referent, highlighting the interaction of the visual cue and other linguistic information such as case marking. This study demonstrates that visually induced "context" can affect the linguistic salience of a referent in discourse and illustrates the interaction of domain-general cognitive mechanisms in language processing.

Keywords: attention; visual cueing; self-paced reading; non-canonical word order; object indexing; Bulgarian

Introduction

Visual attention and structural choice

Language is a complex cognitive system that interacts closely with other cognitive domains, such as attention. Interestingly, this link is not only present in the auditory domain but also when integrating visual attention cues from the environment: "Directing attention via referential *and perceptual* [emphasis added] priming causes people to construe a scene in a particular way and typically this is reflected in the linguistic structures people use" (Ibbotson, Lieven, & Tomasello, 2013, p. 457). When perceiving and describing the world with language, visual cues can determine the choice of an appropriate linguistic response, even to the level of structural choice (i.e., choice of a particular grammatical construction).

In this direction, visual cueing (or perceptual priming) studies showed that visual attention can directly affect language production (Gleitman, January, Nappa, & Trueswell, 2007; Myachykov, Pokhoday, & Tomlin, 2018; Tomlin, 1997). These authors used linguistic adaptations of Posner, Snyder, and Davidson (1980)'s visual cueing paradigm. In the classical Posner paradigm for testing visual attention, an object (a circle, for instance) appears either on the left or on the right side of the screen in an experimental setting. Usually, the position of the object is validly cued by a star or arrow before appearing whereas, in a smaller number of cases, the position indicated is not the actual position at which the object appears. Congruent cueing (i.e., high cue validity) of the position leads to faster and more accurate responses.

Similarly, in the *fish film* paradigm (Tomlin, 1997) participants see two fish approaching each other from two sides of the screen and finally one swallowing the other. Visual attention is cued by an arrow pointing to the eating fish (agent) or the eaten fish (patient). Participants have to describe the event afterwards. It was shown in various languages that cueing one or the other fish led to different sentence choices. When the patient fish was cued as the (visual) centre of attention, participants used more passives in English and more object-first sentences in Russian (Myachykov & Tomlin, 2008).

Two recent studies investigated the effect of visual cueing on sentence production in German (Esaulova, Penke, & Dolscheid, 2019, 2020). In both studies, participants were presented visually with depicted event scenes showing two referents in a transitive situation (e.g., *a witch pushing a hairdresser with a cart*). Participants were asked to describe the event with one-sentence descriptions (additionally, eye movement was measured). The pictures either showed two animate referents or one animate referent acting upon an inanimate referent. In half of the trials, the patient of the event was cued with a small circle appearing in the position of this referent, before the actual image was shown. Esaulova et al. (2019) reported that visual cueing caused more looks to the cued referent but did not affect the structural choice (i.e., did not elicit more passive sentences or non-canonical word order).

Esaulova et al. (2020) used a more explicit patient cueing by presenting an image of the patient referent before the event image was shown. Unlike the previous study, this type of object cueing elicited more passive sentences in this study (however, only for left-positioned ones). The authors concluded that German has a comparably strong agent preference and therefore visual cueing might be less effective in eliciting non-canonical structures. Nevertheless, these studies show that visual cueing might interact with the processing of linguistic structure, even in the case of more complex structures, such as non-canonical word orders or passives.

The present study

The studies mentioned so far were concerned with language production. However, the influence of visual cueing on the salience of a referent should also be observable in the in-

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

terpretation of referent-directed grammatical structure in language comprehension. To address this issue, the present study – focusing on Bulgarian – investigated whether visually cueing either the subject or the object of an event affects the interpretation of a textual event description with either a subject-initial or object-initial word order in a self-paced reading and comprehension task.

Following Esaulova et al. (2020), referents of an event were cued by presenting them for a short amount of time prior to the event itself and a subsequent textual description of the event. Presenting one of the two referents beforehand should lead to a cueing effect making this referent more salient in this cross-modal mini-discourse – just like textual context might impact the salience of a referent due to repeated mention.

Bulgarian is an ideal testing candidate for the investigation of the interaction of visual cueing and linguistic cues due to some specific features of Bulgarian grammar. The most relevant linguistic dimension investigated in this study is word order. Half of the sentences were simple canonical SVO sentences with the agent inhabiting the subject position. This was contrasted with event descriptions that put the patient in the sentence-initial position. Non-canonical OVS in Bulgarian is commonly used for marking highly salient referents and is often accompanied by an object index (”clitic doubling”, ”differential object indexing”) in Bulgarian, with the index highlighting the deviation from the canonical word order or other preferred patterns (Guentchéva, 2008; Compensis, 2022). Therefore, visually cued (i.e., more salient) object referents should match this structure more closely. In addition, (definite) masculine referents unlike feminine or indefinite referents are additionally flagged with case on the noun allowing for the testing of referents that are marked with an additional role-assigning cue.

Assuming that visual attention and grammatical structure interact in this setting, I expected that converging cues (e.g., visual object cueing and object-initial order) should improve accuracy and reaction time to the comprehension question and during self-paced reading. Divergent cueing (e.g., subject cueing and object-initial order) on the other hand should negatively impact accuracy and slow down processing as measured by the self-paced reading reaction times per word. This study is a first attempt to investigate the interaction of visual salience manipulated by visual cueing and non-canonical word order during language comprehension in Bulgarian.

Methods

Participants

In total, 38 native speakers of Bulgarian were recruited for this experiment on *Prolific* (`www.prolific.co`). Unfortunately, twelve participants had to be excluded from the analysis because they failed on attention checks in form of easy-to-answer grammatical filler sentences with unambiguous event descriptions. The remaining 26 participants (15 females, 57.69 %) had a mean age of 28.96 (*SD* = 6.12).

Materials and design

For this study, I used the visual material adapted from Esaulova et al. (2019, 2020). Two example stimuli are given in figure 1 and two event descriptions are presented in example 1. All stimuli material, scripts and raw data of this experiment are publicly available at `https://osf.io/c9wjb/`. Twelve pictures showed pairs of feminine referents and the other 12 showed masculine referents. All pictures were additionally rotated as mirror images, leading to 48 visual target stimuli. Additional 24 filler pictures combined a human referent and an inanimate entity, with the person operating on the inanimate object. Each participant saw 72 events. For the visual cues, the respective referents were extracted and presented as stand-alone.

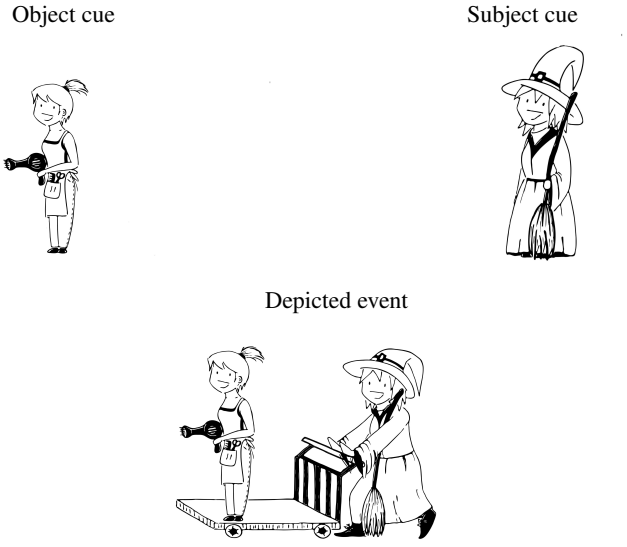

Figure 1: Illustration of visual stimuli

(1) Illustration of target sentence stimuli

a. Canonical SVO condition
Vešticata e butala
witch-ART.SG.F be.PRS.3SG push-PTCP.F
friz'orkata
hairdresser-F-ART.SG.F
'The witch pushed the (female) hairdresser.'

b. Non-canonical OVS condition
Friz'orkata **ja** e
hairdresser-F-ART.SG.F 3.SG.F.ACC be.PRS.3SG
butala vešticata
push-PTCP.F witch-ART.SG.F
'The witch pushed the (female) hairdresser.'
(theoretically also: 'The (female) hairdresser pushed the witch.')

Table 1: Means and standard deviations of accuracy, the reaction time to choice and the self-paced reading times

condition		Accuracy		Reaction time in ms											
		in %		NP1		OI		AUX		V		NP2		to choice	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
feminine referents															
subject cue	SVO	96.79	18	571	292			403	172	442	291	600	496	1652	605
	OVS	92.95	26	675	376	458	176	410	176	544	359	701	482	1703	683
object cue	SVO	94.87	22	610	353			434	261	485	350	611	453	1614	628
	OVS	90.38	30	586	298	457	249	400	252	475	274	649	465	1564	654
masculine referents															
subject cue	SVO	93.59	25	608	360			439	209	441	262	640	478	1585	587
	OVS	87.82	33	635	419	507	402	408	211	516	321	767	576	1548	650
object cue	SVO	94.87	22	639	371			461	251	461	309	673	539	1600	643
	OVS	89.10	31	680	426	476	289	411	181	497	308	696	516	1634	654

Based on the visual events, textual event descriptions were developed for the purpose of this study. Each sentence contained a very simple structure, consisting of a first NP, an object index (in the OVS condition, marked in bold in the example), an auxiliary, a participle verb and a second NP. After the event picture, the respective sentence was presented (either with SVO or OVS order). OVS sentences always contained object indexes. The event descriptions in the fillers were matched by unambiguous filler sentences that also served as attention checks. The comprehension question always asked for the event itself (e.g.,"Did the witch push the hairdresser?") and was presented in SVO order.

Procedure

The presentation script for this web-based experiment was developed using the *jsPsych* package (de Leeuw, 2015) and the experiment was hosted on *Pavlovia* (`www.pavlovia.org`).

When participants started the experiment, the screen automatically changed to full-screen mode. Consecutively, the instructions were presented and participants were asked to start four exercise trials. After the exercise, participants could take a short break and then start the actual experiment where they saw 72 visual cues, event pictures and 72 sentences with a subsequent comprehension question. The presentation of the stimuli was individually randomized by *jsPsych*.

Each trial started with a fixation cross (size: 50 points, duration: 500 ms) on a white screen, followed by a blank screen for 500 ms. Then, the visual cue (either cueing the subject or the object) was presented for 700 ms, followed by a short blank screen for another 500 ms. Then, the event picture was presented for 3000 ms, after which a button appeared, prompting the participants to push it when they were ready to read the respective target sentence. The sentence was presented in a self-paced reading format, i.e., participants had to press the space key to read the next word. At the end of the target sentence, three question marks appeared for 700 ms, indicating that the comprehension question turned up. The question tested for the comprehension of the full event and had to be answered by pressing "e" for yes or "i" for no.

Statistical analyses

Pre-processing and analyses were conducted in *R* (R Core Team, 2019). Trials with a reaction time of more than 3000 ms or below 100 ms were excluded from the reaction time data, leading to the exclusion of 297 trials of the 1248 trials in total for reaction time to choice, eight out of 1248 for reaction time to the first NP, three for the verb and 44 for reaction time to the second NP.

Means and standard deviations were calculated for each dependent variable (accuracy, reaction time to choice and reaction time per word) per subject and condition. The mean reaction times per word from the self-paced reading task were plotted in a point plot (with connected lines to highlight differences between the conditions).

For each dependent variable, a linear mixed effects model with the fixed factors CUE, ORDER and GENDER was calculated using the *lmer()* function from the *R* packages *"lme4"* (Bates, Mächler, Bolker, & Walker, 2015). In addition, the model included random effects for subject and item as well as random slopes for ORDER to account for different sensitivities to non-canonical word orders per subject. Additional group-level linear mixed effects models for the interaction of CUE and ORDER were calculated for each gender separately.

Results

Means and standard deviations

The means and standard deviations for each dependent variable are summarized in table 1. Reaction times per word are additionally plotted in figure 2.

Accuracy of choice is generally better for sentences with SVO order (irrespective of gender or cue), ranging between 93.59 to 96.79 % than for sentences with OVS (range: 87.82

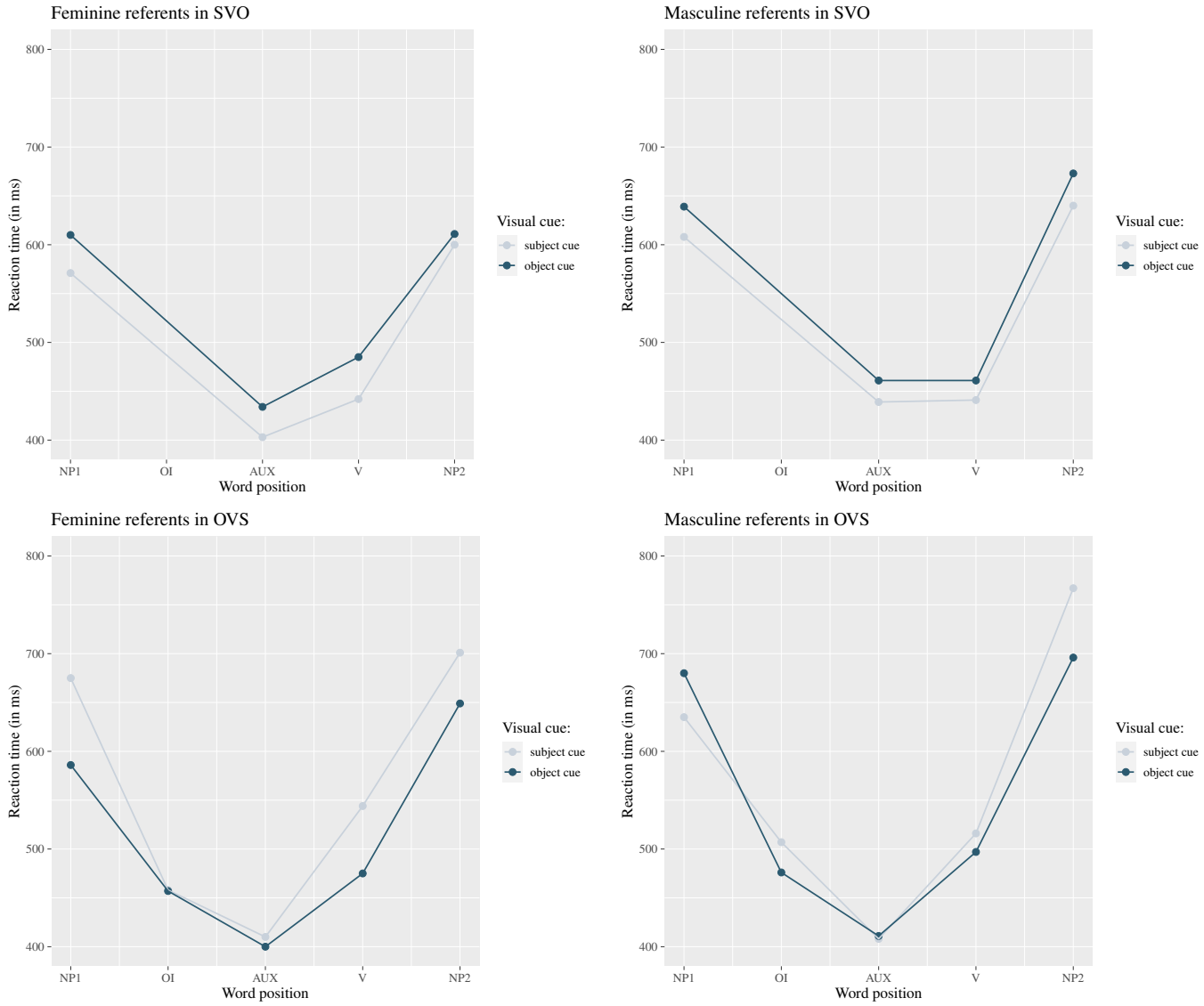

Figure 2: Mean self-paced reading (button press) reaction times per word (in ms)

to 92.92 %). With feminine referents, the presence of an object cue decreases the accuracy, both for SVO (from 96.79 % to 94.87 %) and OVS (from 92.95 % to 90.38 %). The opposite is true for masculine referents. Here, the presence of an object cue improves accuracy for OVS (from 87.82 % to 89.10 %) as well as for SVO (from 93.59 % to 94.87 %).

With respect to reaction time to choice for feminine referents, reaction time was the fastest for OVS after object cues (M = 1564, SD = 654), followed by SVO after an object cue (M = 1614, SD = 628) and slower after a subject cue (with OVS after a subject cue being the slowest condition in terms of reaction time to choice). For masculine referents, the choice was quicker with subject cues (SVO: M = 1585, SD = 587; OVS: M = 1584, SD = 650) than for object cues (SVO: M = 1600, SD = 643; OVS: M = 1634, SD = 654). Interestingly, however, after a subject cue, OVS was quicker and after an object cue, response to SVO was quicker.

For the single reaction times per word, a more consistent picture emerged for both genders (as visible in figure 2). For SVO sentences, the presence of an object cue slowed down reaction time for each word (to different degrees). This difference is slightly more pronounced for feminine referents. In contrast, in OVS the object cue speeds up reaction time for almost every word (except for the NP1 position with masculine referents that additionally carry case flagging).

Linear mixed-effects models

The linear mixed-effects models revealed a main effect of ORDER on accuracy ($\chi^2(1) = 4.86$, $p = 0.028$*) and on the reaction times for the auxiliary ($\chi^2(1) = 5.65$, $p = .017$*), the verb ($\chi^2(1) = 14.16$, $p < .001$***) and the second nominal phrase ($\chi^2(1) = 6.38$, $p = .012$*). There is a significant effect

for the interaction of CUE and ORDER at the verbal position ($\chi^2(1) = 4.08$, $p = .043$). At the position of the first NP, the interaction of all three independent dimensions was significant ($\chi^2(1) = 4$, $p = .046$*). At the position of the object index (OI), there was no significant main effect or interaction effect. Note that this is not necessarily indicative of no interaction of the visual cue and the index since interaction effects on self-paced reading may occur slightly shifted in time (i.e., the button presses are potentially slowed down or speeded up on the following words).

Group-level comparisons

Additional group-level comparisons for the interaction of CUE and ORDER – calculated for each gender separately – revealed no significant interaction effects for masculine referents (only main effects for order). In contrast, for feminine referents, there was a highly significant interaction of CUE and ORDER in the self-paced reading reaction times at the position of NP1 ($\chi^2(1) = 7.48$, $p = .006$**) and at the verb ($\chi^2(1) = 4.23$, $p = .040$*).

Apparently, for SVO, the type of cue did not play a huge role, but object cues slightly increased (i.e., slowed down) button press reaction time at both positions. For sentences with OVS, however, the object cue speeded up reaction time for NP1 (before the object index was examined) and for the verb (after the object index was presented). The interaction effect can be nicely illustrated at the verb position. Here, order of the sentence affected the reaction time for the verb ($\chi^2(1) = 3.98$, $p = .046$*), with OVS increasing reaction time by 85 ms ($\pm$30 standard errors). However, the significant interaction of an object cue and OVS ($\chi^2(1) = 4.23$, $p = .040$*) decreased reaction time by 83 ms ($\pm$40 standard errors).

Discussion

This experiment investigated whether the visual cueing of an agent/subject or patient/object of a visually presented event affects the interpretation of textual event descriptions that either matched or mismatched the salience induced by visual attention by using a corresponding word order and other supportive linguistic cues (case flagging of masculine referents, indexing of object arguments) or a divergent alignment of these linguistic cues in a sample of Bulgarian native speakers.

The presence of a visual object cue reduced the accuracy (as response to the comprehension question) for sentences with feminine referents in contrast to visual subject cues (irrespective of the word order in the sentence). The opposite effect was found with regard to masculine referents, where cueing the (future) object of the event slightly improved accuracy for SVO as well as OVS sentences. For reaction time to answering the question, no strong differences occurred.

For the self-paced reading reaction times, a more consistent picture emerged for feminine and masculine referents. SVO was processed faster after subject cues (reflected in the button presses per word), whereas OVS was processed faster after object cues. Divergent cue and order combinations (e.g., OVS after subject cues) slowed down the SPR reaction times. The positive effect of visually induced "context" on the interpretation of either a canonical or non-canonical word order is in line with the well-known finding that supportive (textual) context improves the processing also of (dis-preferred) object-initial orders (Kaiser & Trueswell, 2004).

These findings indicate a contribution of the visual cue, the presence of a case marker on masculine nouns and the object index on the outcome and time course of interpretation. All three can be understood as cues facilitating role assignment in terms of argument interpretation, in line with the idea that humans always try to determine as quickly and unambiguously as possible *who did what to whom* (Bornkessel-Schlesewsky & Schlesewsky, 2009). In that sense, the visual cue, word order, and the two morphosyntactic cues (case flag and object index) jointly facilitate the assignment of roles and illustrate how linguistic and non-linguistic cues might affect the interpretation of referents.

Differences between feminine and masculine referents can possibly be explained by the presence of an additional cue in the case of masculine referents. In this study design, this case marker apparently distracted and slowed-down the interpretation of event descriptions with masculine referents. This could either be due to a cognitive overload when three cues appear in such a controlled design aiming at a quick response (especially when the cues diverge) or due to the decreasing use of this case marker in contemporary Bulgarian.

This study has some limitations. As was pointed out by one reviewer, it is important to note that comprehension questions always had a SVO order, thereby leading to a potentially disadvantageous mismatch between the question and target sentences with OVS order. Due to the inclusion of three factors, the number of trials per condition was comparably low. Also, I only used 12 different event descriptions as targets (and their mirror counterpart). The total number of trials was comparably low as well. This was unavoidable, given that web-based experiments typically have a higher drop-out rate to ensure motivation to participate (nevertheless, the high exclusion rate suggests that this task was still comparably complex and strenuous for many participants). Clearly, the general pattern identified by this study should be substantiated by lab-based or massive online experiments. Also, in this study, no pre-set breaks or a split of the trials into several blocks were included. Breaks were only possible between seeing the pictures and reading the text. In theory, this could confound potential cueing effects of the visual cue on the text when participants did take a break at this position. On the other hand, the relatively high accuracy for each condition still shows that participants diligently worked on the task.

Nevertheless, this experiment identified interactions of visual cueing on the interpretation of different sentence structures in Bulgarian. This research complements production studies that showed that visual object cueing increases the likelihood of the speaker selecting a non-canonical structure

and provides evidence that visual cueing can also influence the processing of such sentences in language comprehension. In addition, some indication was found for an interaction of the visual cue and other linguistic cues, such as order, case and object indexing. These interactions are indicative of these cues' drawing on domain-general cognitive mechanisms associated with attention.

Acknowledgments

I would like to thank Petra B. Schumacher for supervising and supporting this research. I would also like to thank Martina Penke, Yulia Esaulova and Judith Schlenter for providing me with the visual stimuli and Petyo and Denitsa Hristova for correction and proof-reading of the sentence material. This study was part of my dissertation project that received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 713600 as part of the *a.r.t.e.s. EUmanities programme*. In addition, this study received financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 - SFB 1252 "Prominence in Language" in the project C07 "Forward and backward functions of discourse anaphora" and the *Cologne Center of Language Sciences (CCLS)*.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Linguistics and Language Compass*, *3*(1), 19–58. doi: 10.1111/j.1749-818X.2008.00099.x

Compensis, P. (2022). *Differential object indexing in Bulgarian: The role of discourse prominence and predictability* (Dissertation). University of Cologne, Cologne.

de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

Esaulova, Y., Penke, M., & Dolscheid, S. (2019). Describing events: Changes in eye movements and language production due to visual and conceptual properties of scenes. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.00835

Esaulova, Y., Penke, M., & Dolscheid, S. (2020). Referent cueing, position, and animacy as accessibility factors in visually situated sentence production. *Frontiers in Psychology*, *11*. doi: 10.3389/fpsyg.2020.02111

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. doi: 10.1016/j.jml.2007.01.007

Guentchéva, Z. (2008). Object clitic doubling constructions and topicality in Bulgarian. In D. Kallulli & L. Tasmowski (Eds.), *Clitic doubling in the Balkan languages* (pp. 203–226). Amsterdam: John Benjamins.

Ibbotson, P., Lieven, E. V. M., & Tomasello, M. (2013). The attention-grammar interface: Eye-gaze cues structural choice in children and adults. *Cognitive Linguistics*, *24*(3), 457–481. doi: 10.1515/cog-2013-0020

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147. doi: 10.1016/j.cognition.2004.01.002

Myachykov, A., Pokhoday, M., & Tomlin, R. (2018). Attention and structural choice in sentence production. In S.-A. Rueschemeyer, M. G. Gaskell, A. E. Hernández, E. M. Fernández, & N. Aznar-besé (Eds.), *The Oxford Handbook of Psycholinguistics* (Vol. 1). Oxford University Press. doi: 10.1093/oxfordhb/9780198786825.013.22

Myachykov, A., & Tomlin, R. (2008). Perceptual priming and structural choice in Russian sentence production. *Journal of Cognitive Science*, *9*(1), 31–48. doi: 10.17791/jcs.2008.9.1.31

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–174.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from `https://www.R-project.org/`

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In E. Pederson & J. Nuyts (Eds.), *Language and conceptualization* (pp. 162–189). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139086677.007