TOPICS
TOPICS IN COGNITIVE SCIENCE

Check for updates

Topics in Cognitive Science 00 (2024) 1–13

ISSN: 1756-8765 online
DOI: 10.1111/tops.12724

This article is part of the topic "Parallelism in the Architecture of Language," Giosuè Baggio, Neil Cohn, and Eva Wittenberg (Topic Editors).

The Interaction of Linguistic and Visual Cues for the Processing of Case in Russian by Russian-German Bilinguals: An Eye Tracking Study

Serge Minor,[a] Natalia Mitrofanova,[a] Marit Westergaard[a,b]

[a]*Department of Language and Culture, UiT – The Arctic University of Norway*
[b]*Department of Language and Literature, NTNU-Norwegian University of Science and Technology*

Received 7 July 2023; received in revised form 29 January 2024; accepted 1 February 2024

Abstract

Modulation of visual attention in the Visual World Paradigm relies on parallel processing of linguistic and visual information. Previous studies have argued that the human linguistic capacity includes an aspect of anticipation of upcoming material. Such anticipation can be triggered by both lexical and grammatical/morphosyntactic cues. In this study, we investigated the relationship between comprehension and prediction by testing how subtle changes in visual representations can affect the processing of grammatical case cues in Russian by Russian-German bilingual children ($n = 49$, age $8-13$). The linguistic manipulation followed previous designs, contrasting SVO and OVS sentences, where the first NP (NP1) was marked with nominative or accusative case, respectively. Three types of visual displays were compared: (i) individual referents (potential agent/theme); (ii) pairs of referents (NP1 + potential agent/theme); and (iii) events (representing interactions between the referents). Participants were significantly more sensitive to the case manipulation when presented with events compared to the other two types of visual display. This suggests that they were able to quickly integrate the thematic role information signaled by grammatical case in the event representations. However, they were less likely to use the case information to anticipate upcoming arguments when the target pictures represented

---

Correspondence should be sent to Marit Westergaard, Institutt for språk og kultur, UiT Norges Arktiske Universitet, PO Box 6050 Langnes, N-9037 Tromsø, Norway. E-mail: marit.westergaard@uit.no

individual referents or pairs of noninteracting referents. We hypothesize that the process of argument anticipation is mediated by the activation of syntactic templates (SVO or OSV, depending on the case marking on NP1). The relatively weak anticipation effect observed may be attributed to the absence, or weak representation, of the noncanonical OVS template in the bilingual children's long-term memory.

Keywords: Grammatical case processing; Visual processing; Event representations; Visual World eye tracking; Bilingualism

1. Introduction, background, and research question

Modulation of visual attention in the Visual World Paradigm relies on parallel processing of linguistic and visual information. Previous studies have shown that processing the name of an object triggers eye movements toward that object, to objects with phonologically overlapping names (Allopenna, Magnuson, & Tanenhaus, 1998), as well as to semantically related objects (Dahan & Tanenhaus, 2005; Huettig & Altman, 2005; Myung, Blumstein, & Sedivy, 2006; Yee & Sedivy, 2006). These results have been taken as evidence that shifts in visual attention are triggered (or constituted) by increased activation of a particular depicted object on the screen due to an overlap between the phonological and conceptual features activated by that object and the features activated by the word(s) in the linguistic input (Altman & Kamide, 2007).

Importantly, many studies in the last two decades have argued that the human linguistic capacity also includes an aspect of anticipation of upcoming material. In the Visual World Paradigm, Altmann and Kamide (1999) showed that when participants were presented with a picture of a boy and several inanimate objects (a cake, a train, a ball, and a car) and the sentence *The boy will __ the cake*, their saccadic eye movements to the image of the cake started significantly earlier when the verb was *eat* (in which case there was only one possible direct object) than when the verb was *move* (when all items were possible objects). Importantly, the eye movements to the cake started before the onset of the word *cake* in the *eat* condition, but after in the *move* condition, leading the authors to argue that the semantics of the verb can be used to anticipate a subsequent grammatical object. Furthermore, Kamide, Scheepers, and Altmann (2003) show that anticipatory looks to an upcoming object may also be triggered by morphosyntactic cues. Thus, when presented with a display of a hare, a fox, and a cabbage as well as the German SVO and OVS sentences in (1)−(2), adult listeners will look at the fox earlier in the sentence where the initial element has accusative case (signaling that it is the object) than when the initial element has nominative case. Similar results were attested in Özge, Küntay, and Snedeker (2019) for Turkish monolingual children (age 4).

(1) **Der** Hase frißt gleich den Kohl.
the rabbit$_{\mathrm{NOM}}$ *eats shortly the cabbage*$_{\mathrm{ACC}}$
"The rabbit will shortly eat the cabbage."

(2) **Den** Hasen frißt gleich der Fuchs.
the rabbit$_{\text{ACC}}$ *eats shortly the fox*$_{\text{NOM}}$
"The fox will shortly eat the rabbit."

It is debated whether these types of prediction effects can be subsumed under the same mechanism that accounts for the activation of conceptually related items (see Altman & Kamide, 2007; Huettig, 2015; Magnuson, 2019 for discussion). On Altman and Kamide's (2007) account, case-marked NPs can activate conceptual representations of events that overlap with the *affordances* of the objects in the visual scene, leading to a shift in attention and consequent eye movements. For example, the accusative-marked NP *rabbit* may activate a representation of an event where a rabbit is acted upon by an active agent, while the visual representation of the fox activates its affordance to act as an agent in events. This match then triggers a shift of attention toward the fox (instead of the cabbage). On this account, prediction as observed in Kamide et al.'s (2003) experiment may be taken to operate in a similar priming-like fashion as the activation of conceptually related objects (e.g., a trumpet when hearing *piano*). Alternatively, the prediction of upcoming referents may require the involvement of an active forecasting system that preactivates linguistic input based on a combinatorial mechanism and is sensitive to linguistic constraints (Huettig, 2015). One possibility is that a sentence-initial accusative-marked NP activates the corresponding syntactic template [Object$_{Acc}$ V Subject$_{Nom}$], which in turn activates the expected semantic features of the event and the subject (e.g., its ability to act as an agent, given the propensity of the subject of a transitive clause to act as an agent). These features are then matched with the properties of the visually depicted referents, which are activated in parallel (Huettig et al., 2022).

This latter account opens up the possibility that certain speakers or populations may be able to link case marking with relevant thematic role features (Agent or Patient) but may nevertheless fail to use case to actively predict the upcoming referent in the sentence, for example, due to a weakened representation of noncanonical syntactic templates (OVS). Indeed, we found evidence for such a dissociation in a previous study which investigated case processing by heritage Russian bilingual 4- to 6-year-olds growing up in Germany, comparing them to age-matched German and Russian monolingual controls (Mitrofanova et al., submitted). We investigated the role of visual stimuli in the processing of case cues by contrasting two visual displays: (i) three individual referents (3-REF, following Kamide et al., 2003; Özge et al., 2016), and (ii) two events (2-PIC, e.g., the rabbit eating the cabbage vs. the fox eating the rabbit). Results showed that already before the onset of the second noun, the Russian monolingual children looked to the agent (e.g., the fox) significantly more when the initial element of the clause was accusative than when it was nominative, in both types of visual display. Even the youngest monolinguals (age 3) showed a significant effect of case in both visual contexts, indicating that grammatical case in Russian is acquired early by monolingual learners. The German monolingual preschoolers, in contrast, did not use case predictively in either display, indicating that the case cues in German may be weaker, less salient or later acquired than in Russian. Interestingly, the bilinguals patterned with the monolinguals in both languages (i.e., effect of case in Russian, no effect of case in German), with one crucial distinction: In Russian, the bilinguals used the case cue to identify the target picture only in the 2-PIC display (event representations), but not in the 3-REF display. Mitrofanova et al. hypothesized that the bilingual children were able to integrate the thematic role information conveyed by the case markers with the visual cues in the event representations but did not use the case cues to anticipate sentence continuations. However, since the number of pictures in the two visual conditions differed, these conditions could not be compared directly. Moreover, the 3-REF display included a picture of the referent of NP1, which may have distracted attention away from the potential Agent or Patient of the event, masking the effect of prediction.

The current study was designed to correct these limitations by allowing for a direct comparison of case processing depending on the type of visual display. The study addressed the following research question: How do visual and linguistic cues interact in language processing? More specifically, do heritage bilingual children use case cues more effectively when the visual display involves representations of events as opposed to individual referents?

2. Method

A group of 49 Russian-German bilingual children (age 8−13, mean 9.6) were recruited and tested at the premises of a bilingual school in Berlin. Approximately half of the children ($n = 25$) came from families with two Russian-speaking parents, and half came from mixed-language families ($n = 22$) or German-speaking families ($n = 2$). Most participants had lived in Germany their whole life ($n = 33$), while 16 participants were born outside of Germany and moved there later in life (between 1 and 7 years of age). The project was registered and approved by the Norwegian Social Science Data Service (*Sikt*). Written informed consent was obtained from the parents prior to testing. All children gave oral consent prior to participating in the study.

The experimental design combined a linguistic cue (nominative vs. accusative case on NP1) and a visual manipulation (three types of visual display). Each trial started with a preamble that involved a visual display depicting three new referents (Fig. 1) accompanied by a spoken sentence introducing them (3).

(3) Èto sobaka, moroženoje, malyš
This is dog ice cream baby
"This is a dog, an ice cream, a baby."

The preamble was followed by the test item. The participants viewed a visual display consisting of two images on the screen while listening to the test sentence. The linguistic cue (grammatical case) was manipulated by including two types of transitive sentences:

(i) SVO sentences, where NP1 appeared in the nominative case (4a);
(ii) OVS sentences, where NP1 appeared in the accusative case (4b).

The linguistic stimuli were the same as in Mitrofanova et al. (submitted).

(4a) Veselyj malyš sejčas lizn'ot moroženoje.
$\textit{cheerful}_{NOM}$ $\textit{baby}_{NOM}$ *now will lick* $\textit{ice cream}_{ACC}$
"The cheerful baby will now lick the ice cream."

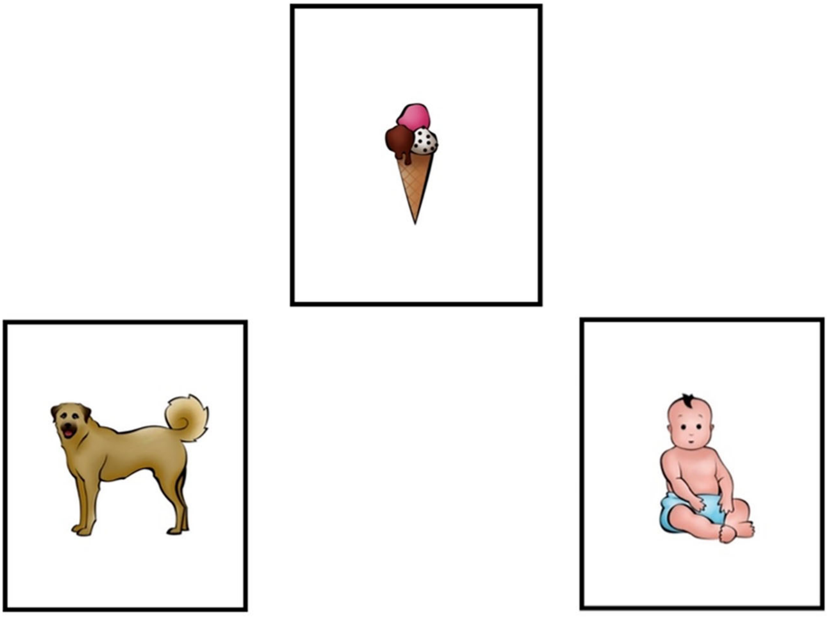

Fig. 1. Preamble display.

(4b) Veselogo malyša sejčas lizn'ot sobaka.
cheerful$_{\mathrm{ACC}}$ *baby*$_{\mathrm{ACC}}$ *now* *will lick* *dog*$_{\mathrm{NOM}}$
"The dog will now lick the cheerful baby."

The visual manipulation involved three types of visual display (Fig. 2):

(*a*) *Objects*—which contrasted individual referents, a plausible theme and agent of the action (e.g., ICE CREAM vs. DOG, Fig. 2a);
(*b*) *Pairs*—which contrasted pairs of pictures combining the theme/agent with the NP1 referent (ICE CREAM+BABY vs. DOG+BABY, Fig. 2b); and
(*c*) *Events*—which involved two pictures representing the NP1 referent interacting with the theme/agent (A BABY LICKING AN ICE CREAM vs. A DOG LICKING A BABY, Fig. 2c).

The 2 × 3 design—Linguistic manipulation (Nominative vs. Accusative case on NP1) × Visual display type (Objects vs. Pairs vs. Events)—rendered six experimental lists, each involving 12 experimental items and 12 fillers.[1] The participants were tested individually in a quiet room at the school premises with a Tobii Fusion 250 infrared eye-tracker mounted on a 17.3'' monitor. Eye gazes were sampled with 120 Hz frequency.

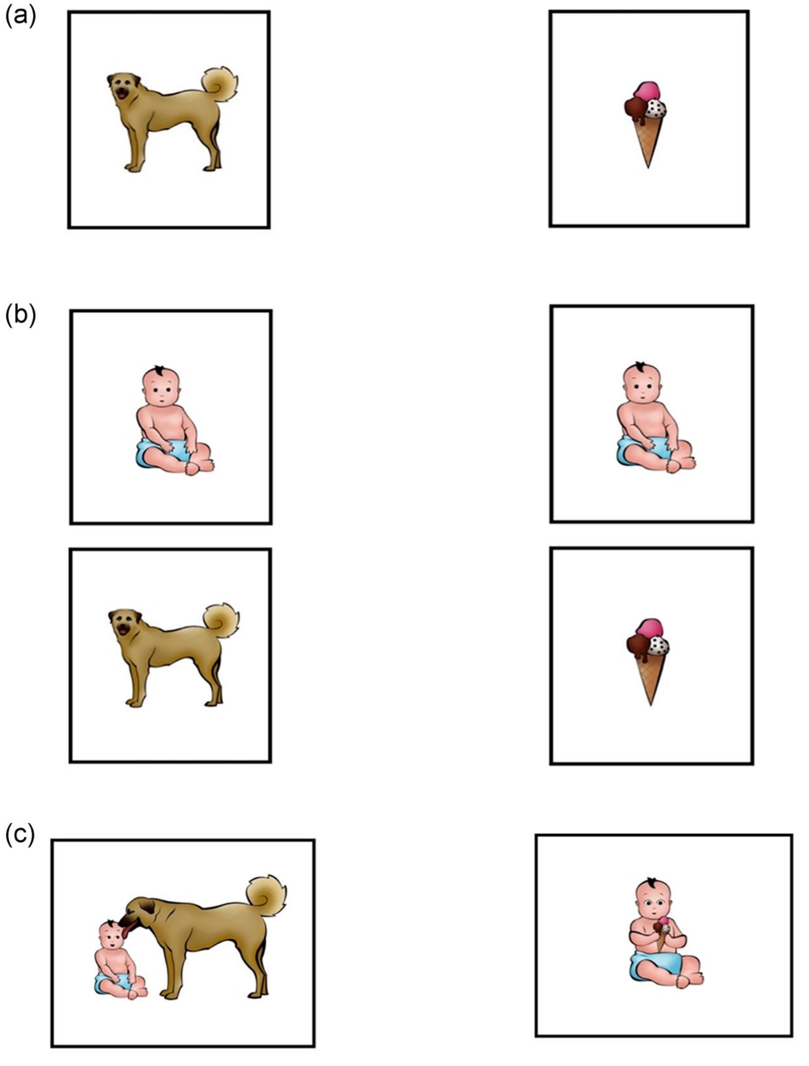

Fig. 2. Three types of visual display involved in the experiment: (a) Objects; (b) Pairs; and (c) Events.

3. Results

To analyze the participants' eye movements, the screen was divided into two equal-sized areas-of-interest (AOIs), corresponding to the left and right sides of the screen. We defined the *prediction window* as the time window which began after the presentation of the first Case cue, that is, at Adjective offset, and ended at the onset of the second NP (NP2). We added 200 ms to the onset and offset of the prediction window to account for the time needed to plan and execute a saccade. More looks to the Agent AOI within the prediction window in the Accusative condition compared to the Nominative would indicate that the participants were

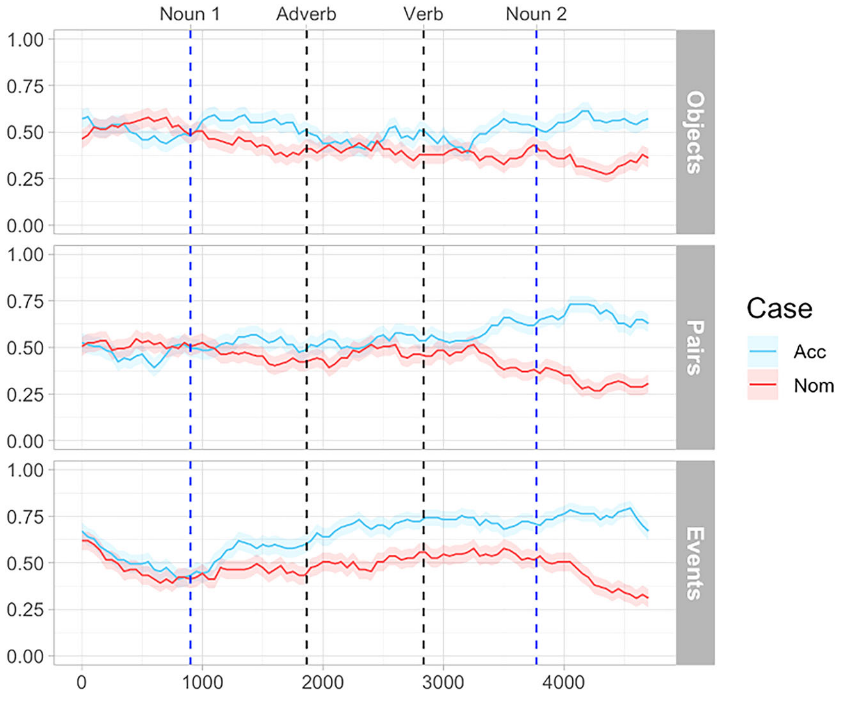

Fig. 3. Proportions of looks to the Agent AOI in the three Visual Display conditions by Case. Proportions were calculated in one hundred 50-ms time-bins starting from the onset of NP1. Vertical lines mark average word onsets (+200 ms). Blue vertical lines indicate the onset and offset of the prediction window.

able to process the grammatical Case cue to assign the correct thematic role to the referent of NP1 (Agent or Patient) and use this information to predict the referent of NP2.

Seven trials with above 50% track loss were removed prior to analysis (1.2% of the data). Fig. 3 illustrates the proportions of looks to the AOI involving the Agent depending on NP1 Case in the three Visual Display conditions. Fig. 4 summarizes the mean proportions of looks to the Agent image within the prediction window. These plots suggest that the effect of Case (i.e., the difference between the proportions of looks in the Accusative and Nominative conditions) was larger in trials where the images depicted events compared to those where the images depicted individual objects or vertically aligned pairs of objects.

To compare the effect of Case in the different Visual Display conditions, we fit a Bayesian generalized mixed regression model, predicting the proportion of looks to the Agent AOI with Case, Visual Display type, and the interaction between Case and Visual Display as fixed effects (see Appendix for details). Fig. 5 illustrates differences between the marginal posterior distributions of Accusative and Nominative conditions for the three Visual Display types. The medians, as well as the 85% and 95% highest density intervals (HDIs) for the Case effect sizes, are given in Table 1.

With 95% probability, the proportion of looks to the Agent picture in the Events condition increased by 10−26% of the maximal range when NP1 was marked with the Accusative

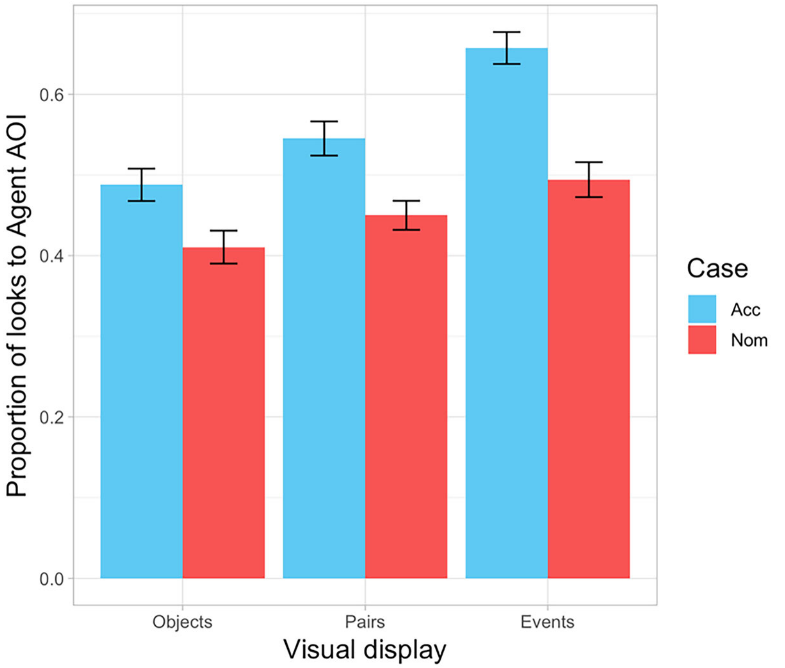

Fig. 4. Average proportions of looks to the Agent AOI in the prediction window.

case compared to the Nominative ($\hat{\Theta} = -0.75,\ SD = 0.16$ [in beta odds]). There is weaker evidence for an effect of Case in the Objects and Pairs Visual Display conditions. With 85% probability, the size of the Case effect was 2−13% of the maximal range in the Objects condition, and 1−14% of the maximal range in the Pairs condition.

To compare the effects of Case between the Visual Display conditions, we calculated the pairwise differences between the distributions in Fig. 5 and computed the HDIs for these differences (see Fig. 6 and Table 2).

These results suggest that with 95% probability, the effect of Case in the Events condition was 2−19% (of the maximal range) stronger than in the Objects condition, and 2−20% (of the maximal range) stronger than in the Pairs condition. There is no evidence that the effect of Case was different in the Objects and Pairs conditions (the median of the difference is close to 0).

4. Discussion

Our study investigated the parallel processing of grammatical and visual information by bilingual Russian-German children (age 8−13). We employed Visual World eye-tracking to test how the processing of grammatical case cues in Russian is affected by the type of visual display presented to the participants. The linguistic manipulation followed previous designs

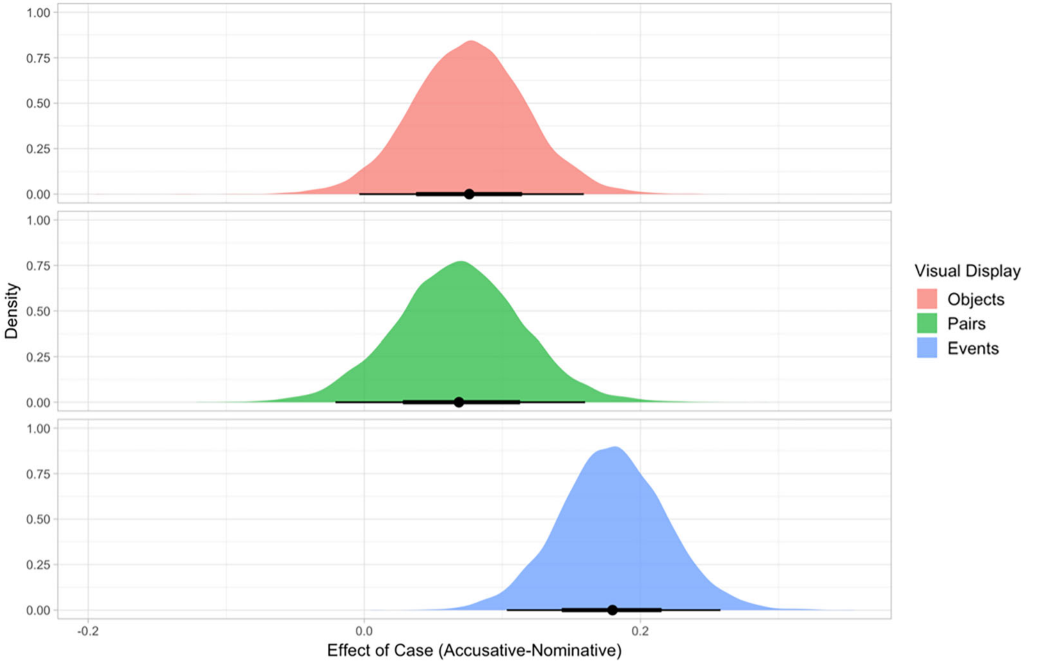

Fig. 5. Differences between the marginal posterior distributions of Accusative and Nominative case conditions for the three Visual Display types. The 95% HDI is represented by the thin line, the 66% HDI is represented by the thick line, and the dot represents the median value.

Table 1
The medians and 85% and 95% HDIs for Case effect size by Visual Display type

Visual display type	Median	85% HDI	95% HDI
Objects	0.08	[0.02, 0.13]	[0, 0.16]
Pairs	0.07	[0.01, 0.14]	[−0.02, 0.16]
Events	0.18	[0.13, 0.24]	[0.1, 0.26]

(Kamide et al., 2003; Özge et al., 2016, 2019; Mitrofanova et al. submitted) contrasting SVO sentences—with NP1 marked with nominative case and NP2 marked with accusative, and OVS sentences—where the case marking was reversed. Unlike previous studies, we compared three types of visual display representing either individual referents (referents for NP2 = potential agent/theme of the action), pairs of referents (the referent of NP1 + potential agent/theme), or events (representing interactions between the referent of NP1 and the potential agent/theme). We found clear evidence of sensitivity to the case cues when the visual display involved representations of events and weaker evidence of such sensitivity when the visual display represented individual referents or pairs of referents. Most importantly, we found that the effect of the case manipulation was significantly stronger when the

Table 2
The medians and 85% and 95% HDIs for the differences in Case effect size between Visual Display types

Difference in Case effects	Median	85% HDI	95% HDI
Events−Objects	0.1	[0.04, 0.17]	[0.02, 0.19]
Events−Pairs	0.11	[0.05, 0.18]	[0.02, 0.20]
Objects−Pairs	0.007	[−0.06, 0.08]	[−0.09, 0.10]

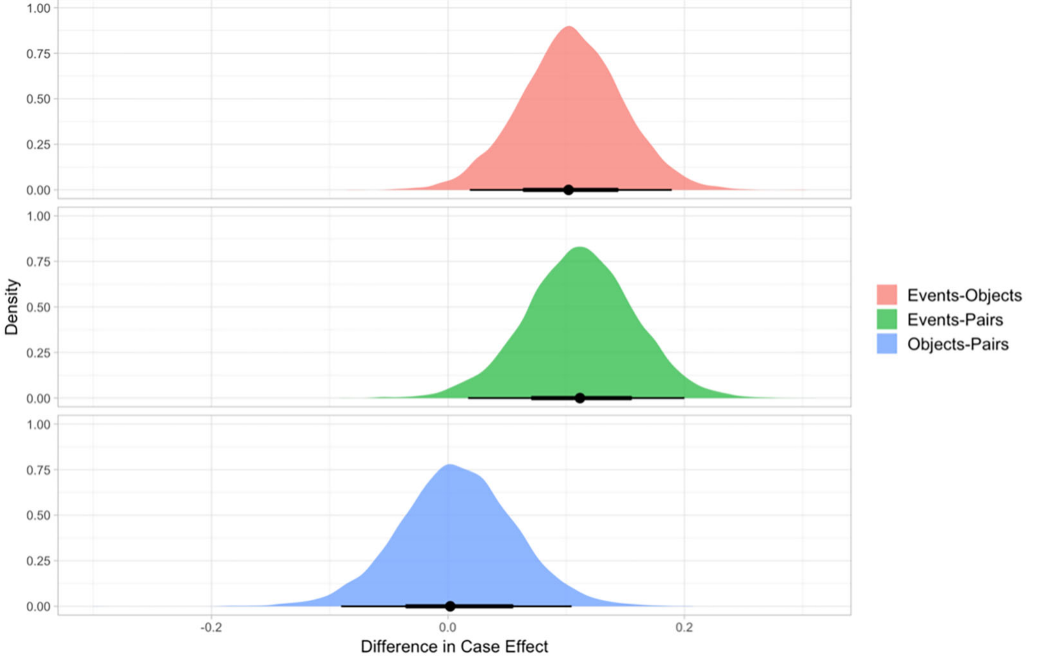

Fig. 6. Differences between the Case effect sizes in the Visual Display conditions. The 95% HDI is represented by the thin line, the 66% HDI is represented by the thick line, and the dot represents the median value.

visual display involved event representations compared to individual referents and pairs of referents (and no evidence that the effect of case differed between the latter two conditions).

These results are in line with Mitrofanova et al. (submitted), who found that younger Russian-German bilingual children (age 4−6) showed sensitivity to case cues in Russian when presented with two event representations but not when viewing representations of three individual referents. Our current study allowed for a direct comparison between the visual conditions and thus strengthened the conclusion that sensitivity to the case manipulation is increased when the visual display involves event representations compared to representations of individual referents.

In order to further isolate the effect of event representations on case processing, we included a third visual condition which contrasted pairs of referents. The Events and Pairs conditions both included representations of two referents in each image but differed in whether the two referents were depicted as interacting with each other or not. We found that the participants' sensitivity to the case cues in the Pairs condition was similar to that in the Objects condition, and significantly weaker than in the Events condition. This suggests that it is the visual representation of events, rather than a mere combination of referents, that facilitated the use of grammatical case to identify the target picture.

Grammatical case conveys information on the role that the referent plays in the event described by the verb. The results of this study suggest that bilingual children are able to quickly integrate this information with the visual cues in the event representations, and this facilitates the identification of the target picture in the Events condition (see also Knoeferle, Crocker, Scheepers, & Pickering, 2005; Knoeferle & Crocker, 2007; Zhang & Knoeferle, 2012). On the other hand, in order to identify the target picture in the Objects and Pairs conditions, the participants needed to actively anticipate the upcoming argument of the sentence. Our results show that this proves more difficult for bilingual children. Following Huettig et al. (2022; see also Jackendoff, this volume), we hypothesize that the process of argument anticipation is mediated by the activation of a particular syntactic structure, or template, stored in long-term memory, where NP1 functions as the initial element (the *incipit*) − [NP_{Nom} V NP_{Acc}] (= SVO order) or [NP_{Acc} V NP_{Nom}] (= OVS order), depending on the case marking on the initial NP. Successful anticipation obtains when activation of the syntactic template leads to the activation of the likely semantic features of the second argument (part of the *remainder* of the template), which are in turn matched with the properties of the visually depicted referents. The relatively weak anticipation effect that we observed may be attributed to the absence, or weak representation, of the noncanonical OVS template in the bilingual children's long-term memory. Indeed, the OVS structure is relatively rare in Russian—11% of all three-member sentences in written language (Bivon, 1971; Bailyn, 1995) and 7% in spoken Russian (Sirotinina, 1965; Slioussar, 2011), and rarer still in the participants' societal language German (less than 1%, Hoberg, 1981; Bader & Haeussler, 2010). It is thus plausible that the bilingual children's reduced input in Russian, in combination with potential influence from German, resulted in a weakened representation of the OVS template.

More generally, the results of the current study support the view of prediction as an important, but not indispensable aspect of language processing and comprehension (Huettig, 2015; Huettig & Mani, 2016). Methodologically, our results warn against interpreting failure to use case to predict upcoming referents (e.g., by children, bilinguals, etc.) as evidence for a total lack of sensitivity to case marking. Instead, such failure may result from deficits specific to the process of prediction, for example, weak representation of noncanonical syntactic templates.

Notes

1 All the linguistic and visual stimuli for the test sentences are available at https://osf.io/3nz5g/?view_only=b9f6e78ff6314bccba72a837e8a2e629.

2 Model formula: PropAgent ∼ 1 + Case*VisualDisplay + (1 + Case*VisualDisplay | Participant) + (1 + Case*VisualDisplay | Item)

References

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558.

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

Altmann, G., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004.

Bader, M., & Häussler, J. (2010). Word order in German: A corpus study. *Lingua*, *120*(3), 717–762. https://doi.org/10.1016/j.lingua.2009.05.007.

Bailyn, J. F. (1995). *A configurational approach to Russian "free" word order*. [Ph.D. dissertation], Cornell University.

Betancourt, M. (2018). Calibrating model-based inferences and decisions. arXiv: 1803.08393 [stat.ME]. https://arxiv.org/abs/1803.08393

Bivon, R. (1971). *Element order*. Cambridge: Cambridge University Press.

Bürkner, P.-C. (2020). *brms*: Bayesian regression models using 'Stan'. Vienna: R Foundation for Statistical Computing. https://cran.r-project.org/package=brms

Carignan, C., Coretta, S., Frahm, J., Harrington, J., Hoole, P., Joseph, A., Kunay, E., & Voit, D. (2021). Planting the seed for sound change: Evidence from real-time MRI of velum kinematics in German. *Language*, *97*(2), 333–364.

Dahan, D., & Tanenhaus, M. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453–459. https://doi.org/10.3758/BF03193787.

Falk H., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050.

Hoberg, U. (1981). *Die Wortstellung in der Geschriebenen Deutschen Gegenwartssprache*. Hueber.

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

Huettig, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. https://doi.org/10.1016/j.cognition.2004.10.003.

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*, 19–31. https://doi.org/10.1080/23273798.2015.1072223.

Jackendoff, R. (2023). The parallel architecture in language and elsewhere. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12698.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55.

Knoeferle, P., & Crocker, M. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*, 519–543.

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

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Amsterdam: Academic Press.

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Scicene*, *3*, 113–139. https://doi.org/10.1007/s41809-019-00035-3.

Mitrofanova, N., Minor, S., Westergaard, M., Sauerman, A., Gagarina, N., Ozge, D., Sekerina, I. (submitted). Processing of Grammatical Case in Russian and German by Monolingual and Bilingual Preschoolers: Evidence from Eye-tracking.

Myung, J.-Y., Blumstein, S., & Sedivy, J. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*, *98*(3), 223–243. https://doi.org/10.1016/j.cognition.2004.11.010.

Özge, D., Kornfilt, J., Münster, K., Knoeferle, P., Küntay, A., & Snedeker, J. (2016). Predictive use of case markers in German children. In *Proceedings of the 40th Annual Boston University Conference on Language Development* (pp. 291–303). Somerville, MA: Cascadilla Press.

Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition*, *183*, 152–180.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org

Sirotinina, O. B. (1965). *Porjadok slov v russkom jazyke*. Saratov State University.

Slioussar, N. (2011). Russian and the EPP requirement in the tense domain. *Lingua*, *121*, 2048–2068.

Stan Development Team. (2017). Stan: A C++ library for probability and sampling. Version 2.14.0. http://mc-stan.org/

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14. https://doi.org/10.1037/0278-7393.32.1.1.

Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Appendix: Statistical analysis

To compare the effect of Case in the different Visual Display conditions, we fit a Bayesian generalized mixed regression model (BRM), using the *brms* package (Bükner, 2020; Stan Development Team, 2017) in *R* (R Core Team, 2020). We modeled the proportion of looks to the Agent AOI with Case, Visual Display type, and the interaction between Case and Visual Display as fixed effects (Accusative case and Event visual display were coded as the baseline). Since the dependent variable was a proportion (on a scale from 0 to 1), we fit a BRM with a Beta distribution. The model included a maximal random effects structure with random intercepts for Participants and Items, as well as random slopes for Case, Visual Display, and the interaction between Case and Visual Display.[2] Weakly informative priors were used for the intercept and all the predictors (following Carignan et al., 2021). The model was run with four Markov chain Monte Carlo chains with 6000 iterations (3000 warm-up samples). The model converged with $\hat{R} = 1$, and an inspection of trace-plots revealed no divergences in the MCMC chains. A sensitivity analysis based on z-scores and shrinkage (Betancourt, 2018) was performed to confirm that the priors did not strongly influence the results. An examination of the posterior predictive check plot indicated that the model captured the distribution of the data well.

We analyzed differences between conditions by calculating the median and 85% and 95% highest density intervals (HDI) of the difference between the marginal posterior distributions of the Case conditions: Acc − Nom (Kruschke, 2014). These differences correspond to Case effect sizes, and the HDI intervals establish the range within which the true size of the Case effect falls with 85% and 95% probabilities. To compare the effects of Case between the Visual Display conditions, we calculated the pairwise differences between the posterior distributions of the Case effect (Events − Objects, Events − Pairs, Objects − Pairs), and computed the HDIs for these differences.