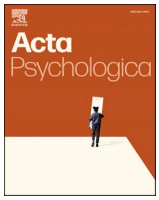
Acta Psychologica 246 (2024) 104241

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



What eye and hand movements tell us about expectations towards argument order: An eye- and mouse-tracking study in German

Judith Schlenter [a,*], Marit Westergaard [a,b]

[a] Department of Language and Culture, UiT The Arctic University of Norway, Tromsø 9037, Norway
[b] Department of Language and Literature, NTNU Norwegian University of Science and Technology, Trondheim 7491, Norway

ARTICLE INFO

Keywords:
Prediction
Reanalysis
Case marking
Word order
Transitive verb
Ditransitive verb

ABSTRACT

Previous research on real-time sentence processing in German has shown that listeners use the morphological marking of accusative case on a sentence-initial noun phrase to not only interpret the current argument as the object and patient, but also to predict a plausible agent. So far, less is known about the use of case marking to predict the semantic role of upcoming arguments after the subject/agent has been encountered. In the present study, we examined the use of case marking for argument interpretation in transitive as well as ditransitive structures. We aimed to control for multiple factors that could have influenced processing in previous studies, including the animacy of arguments, world knowledge, and the perceptibility of the case cue. Our results from eye- and mouse-tracking indicate that the exploitation of the first case cue that enables the interpretation of the unfolding sentence is influenced by (i) the strength of argument order expectation and (ii) the perceptual salience of the case cue.

PsycINFO code: 2720 Linguistics & Language & Speech.

1. Introduction

There is abundant evidence that sentence processing proceeds incrementally; that is, listeners do not wait until the end of the sentence to process its structure and meaning but do so along the way. Listeners can also predict what comes next, using multiple sources of information (for review, see e.g., Kamide, 2008; Kuperberg & Jaeger, 2016; Traxler, 2014). This raises the questions of whether and how sources of information are weighted, especially in cases of conflicting information. In the present paper, we focus on linear order and morphological case marking as two sources of information in German sentence processing. While previous research has shown that German native speakers exploit case marking on a sentence-initial argument to predict the upcoming one (e.g., Hopp, 2015; Kamide, Scheepers, et al., 2003), less is known about how argument interpretation is affected by listeners' prior expectations towards argument order. In the present study, we tested when case marking comes into play as a cue to argument interpretation in transitive and ditransitive sentence structures while controlling for factors such as animacy and perceptual salience of case marking that could have influenced real-time processing in previous studies. With this study, we aim to address potential confounds in prior materials and extend previous research on pre-verbal case marking to case marking on post-verbal arguments.

Before we present the experiments in this study, we briefly introduce the German case system and provide some background on the incremental interpretation of arguments, with a focus on sentence processing in German. Note that whenever we use 'order' here, we refer to surface order and not the underlying order. We use the commonly used terms agent, patient, recipient, and theme to refer to an argument's semantic role, although we are aware that there is considerable controversy about this in the field (Rissman & Majid, 2019).

1.1. The German case system

One important cue to argument interpretation (i.e., identifying 'who did what to whom') is morphological case marking on pronouns and/or noun phrases (NPs). Here, we focus on full NPs. German distinguishes between nominative (NOM), accusative (ACC), dative (DAT), and genitive (GEN) case. Case is fused with gender (masculine, feminine, neuter) and number (singular, plural). As shown for singular nouns in Table 1, NOM and ACC are unambiguously marked on the preceding article only for masculine nouns, whereas there is no overt marking for

* Corresponding author.
E-mail address: judith.schlenter@uit.no (J. Schlenter).

https://doi.org/10.1016/j.actpsy.2024.104241
Received 31 October 2023; Received in revised form 2 April 2024; Accepted 5 April 2024
Available online 12 April 2024

Table 1
Inflectional paradigm for singular nouns in German.

	Masculine		Feminine	Neuter
	weak	strong		
Nominative	der Prinz-ø	der Mann-ø	die Frau-ø	das Gespenst-ø
Accusative	den Prinz-en	den Mann-ø	die Frau-ø	das Gespenst-ø
Dative	dem Prinz-en	dem Mann-ø	der Frau-ø	dem Gespenst-ø
Genitive	des Prinz-en	des Mann-(e)s	der Frau-ø	des Gespenst-(e)s
	'the prince'	'the man'	'the woman'	'the ghost'

feminine and neuter nouns due to case syncretism. Masculine nouns are further subdivided into strong and weak nouns. The latter receive additional marking on the noun.

1.2. Incremental argument interpretation in transitive sentence structures

The incremental interpretation of arguments in transitive sentences with an agent and a patient argument has received considerable attention in psycho- and neurolinguistic research, where it has been repeatedly shown that language comprehenders have an initial focus on the agent. For instance, in a visual-world eye-tracking experiment of Tagalog, which marks the semantic role of a so-called pivot argument in sentence-final position on the sentence-initial verb, Sauppe (2016) showed that, upon hearing the verb, the participants first looked at the agent of the transitive event, irrespective of argument order. Both agents and patients are equally likely to appear as pivots, which typically follow the non-pivot argument. Only later, yet still prior to encountering the final argument, did listeners integrate all information to anticipate (or 'predict') the respective referent.

Much research has focused on semantic role ambiguities to examine comprehenders' interpretative biases for transitive events. In an event-related potential (ERP) experiment on the incremental interpretation of sentences with role-ambiguous initial NPs, Sauppe et al. (2023) tested speakers of Äiwoo. In the basic word order of this language, the patient precedes the agent. Nevertheless, when the first NP was human and disambiguation (via symmetrical voice marking) was towards a patient-initial order, an increased negativity was observed in comparison to the agent-initial order, pointing towards an agent-first bias. In a language like German, the agent (typically realized as subject) tends to precede all other arguments. Therefore, in a neutral context, object-initial sentences such as (1b) and (1c), although grammatical, are referred to as 'marked' (e.g., Bornkessel et al., 2002). In terms of frequency, sentences that start with an ACC object make up less than a quarter of structures in corpus analyses (Bader & Häussler, 2010; Bader & Portele, 2019). Altogether, this has implications for argument interpretation at the beginning of a German clause. Psycholinguistic studies that have included role ambiguities consistently report an agent-first preference (e.g., Hemforth, 1993; Hemforth & Konieczny, 2000; Weber et al., 2006), as e.g., in (1c) which has a feminine noun in initial position.

(1) a. *Der Drache* *besiegt schließlich* *den Prinzen*
[The dragon]$_{NOM}$ defeats finally [the prince]$_{ACC}$
'Finally, the dragon defeats the prince.'

b. *Den Drachen* *besiegt schließlich* *der Prinz*
[The dragon]$_{ACC}$ defeats finally [the prince]$_{NOM}$
'Finally, the prince defeats the dragon.'

c. *Die Hexe* *besiegt schließlich* *der Prinz*
[The witch]$_{NOM/ACC}$ defeats finally [the prince]$_{NOM}$
'Finally, the prince defeats the witch.'

The above findings indicate that the agent plays a special role in sentence processing. In the literature, the linear order of arguments and case marking (+ 1st position, + nominative), as well as animacy (+ animate), have been considered as prominence features (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2009, 2016) or cues (e.g., MacWhinney et al., 1984) that are used to identify the agent and that are weighted differently depending on the specific language. In German, overt case marking for masculine nouns, see examples (1a) and (1b), enables the immediate mapping to a semantic role: agent for NP1 in (1a) and patient for NP1 in (1b). Nevertheless, findings from ERP experiments in German show that object-initial orders in comparison to subject-initial orders induce processing costs, even when the object is unambiguously marked with ACC case (e.g., Bornkessel et al., 2002; Dröge et al., 2016; Matzke et al., 2002; Rösler et al., 1998; Schipke et al., 2012; Schlesewsky et al., 2003; but cf. Frisch et al., 2002, who did not find processing costs for unambiguous object-initial clauses). The ACC object-initial order has been found to elicit an enhanced negativity on the first NP in comparison to the NOM subject-initial order. As an explanation for this, Bornkessel et al. (2002) have argued that the parser expects a prototypical word order, but ACC signals a non-prototypical structure. In line with this, they interpret the neurophysiological response to an ACC-marked argument in clause-initial position as a reflection of a prediction mismatch.

Moreover, several visual-world eye-tracking experiments have shown that adult German native speakers use ACC to predict that the upcoming argument is the agent (Cristante, 2016; Henry et al., 2017; Hopp, 2015; Kamide, Scheepers, et al., 2003). For example, in Hopp (2015), participants were presented with a semi-realistic scene that included the image of a wolf, a deer, a hunter and, as distractor image, a mountain, while listening to a subject-verb-object (SVO) or object-verb-subject (OVS) sentence as illustrated in (2).

(2) a. *Der Wolf* *tötet* *gleich* *den Hirsch*
[The wolf]$_{NOM}$ kills soon [the deer]$_{ACC}$
'The wolf will soon kill the deer.'

b. *Den Wolf* *tötet* *gleich* *der Jäger*
[The wolf]$_{ACC}$ kills soon [the hunter]$_{NOM}$
'The hunter will soon kill the wolf.'

The results showed that German native speakers integrated both morphosyntactic and lexical-semantic information to predict either an agent or a patient, as reflected by a preference to look at the respective image of the target argument before its auditory onset. Another participant group of non-native speakers of German did not use ACC to predict an agent (*hunter*) upon encountering the transitive verb (*kill*) but anticipated the final argument to be a patient (*deer*), indicating that they used a word-order strategy to interpret the sentences.[1] This leads to a correct sentence interpretation in subject-initial sentences (2a), but to an incorrect interpretation in object-initial sentences (2b), at least temporarily. Hence, different groups of speakers may weigh cues to argument interpretation differently.[2] However, a closer look at the gaze pattern (Hopp, 2015, Fig. 3) reveals that, while the adverbial and first part of the second argument was heard, German native speakers fixated more on the competitor image in the object-initial condition (2b) than in the subject-initial condition (2a). This indicates that even in adult native processing, linear order as a cue may not be canceled out immediately upon encountering a case cue.

While the above-mentioned findings suggest that native speakers of

[1] The non-native speakers in this study were English native speakers. English has a rigid word order (SVO) and only marks case on pronouns but not on full NPs. Thus, one possible explanation for this outcome is the lack of familiarity with case.

[2] German-speaking children up to the age of six to seven years have also been found to differ from adult German native speakers in the processing of subject- and object-initial sentence (e.g., Cristante, 2016; Kröger et al., 2017; Schipke et al., 2012). However, Özge et al. (2022) have recently demonstrated that German children up to this age may use morphological case in object-initial sentences predictively.

German capitalize on case marking to revise an initial expectation towards a subject−/agent-first order and to predict upcoming information, there are other information sources that might help listeners identify semantic roles. In example (3), from Özge et al. (2022: 6), future tense is used to separate the case cue and the lexical-semantic cue provided by the verb. However, in their materials, the agent was always prototypically animate and the patient prototypically inanimate (3a) or less agent-like than another depicted referent (3b).

(3) a. *Der Hase* *wird* *im nächsten Moment* *den Kohl* *aufspüren*
[The hare]$_{NOM}$ will in the next moment [the cabbage]$_{ACC}$ find
'The hare will find the cabbage in the next moment.'

b. *Den Hasen* *wird* *im nächsten Moment* *der Fuchs* *aufspüren*
[The hare]$_{ACC}$ will in the next moment [the fox]$_{NOM}$ find
'The fox will find the hare in the next moment.'

We argue that animacy could have been an additional semantic cue in some studies testing transitive structures as could world knowledge, for instance, in example (3) but also in Hopp (2015): Consider the object-initial sentence *Den Wolf tötet gleich der Jäger* ('The hunter will soon kill the wolf') from example (2). A wolf is probably more likely to kill a deer than a hunter, so a deer is a good patient. In contrast, a wolf is comparatively less likely to be killed by a hunter and thus is a less prototypical patient. Moreover, we do not know how differences in the perceptibility of case marking cues might influence the processing of non-prototypical OVS word order. Previous studies only included strong masculine nouns (Cristante, 2016; Hopp, 2015) or mixed strong and weak masculine nouns, but did not report whether they tested for processing differences between strong and weak masculine nouns (Kamide, Scheepers, et al., 2003; Özge et al., 2022).

1.3. Incremental argument interpretation in ditransitive sentence structures

While much research has tested the incremental interpretation of arguments in transitive sentences, less research has focused on ditransitive structures and on expectations towards argument order after the agent has been encountered. In fact, there is little consensus as regards the prototypical order of arguments in German double-object constructions (e.g., Häussler & Bader, 2012; Røreng, 2011) like in example (4) from the present study.

(4) *Der Krankenpfleger bringt* … — 'The nurse brings …'
a. *dem Patienten* *morgens* *die Ärztin*
[the patient]$_{DAT}$ in the morning [the doctor]$_{ACC}$
'In the morning, the nurse brings the doctor to the patient.'
b. *den Patienten* *morgens* *der Ärztin*
[the patient]$_{ACC}$ in the morning [the doctor]$_{DAT}$
'In the morning, the nurse brings the patient to the doctor.'

The order recipient > theme is overall more frequent, as shown in corpus analyses, but not necessarily so when both objects are animate (Häussler & Bader, 2012).

Results from studies that have investigated the processing of ditransitive structures in German suggest that comprehenders have an expectation as regards the ordering of direct object (DO)/theme and indirect object (IO)/recipient with a preference for the argument immediately following the verb to be the recipient (Rösler et al., 1998; Scherger et al., 2022; Schlenter & Felser, 2021). Rösler and colleagues observed longer reaction times for the order DO-IO than for IO-DO. Moreover, the authors measured differences in ERP amplitudes between the two orderings: the elicited negativity at the first postverbal argument was larger in response to ACC-marked articles (DO/theme) than for DAT-marked articles (IO/recipient). Similarly, in a picture selection task conducted by Scherger et al. (2022) participants needed more time to select the correct picture for the DO-IO order than for the IO-DO order. In a visual-world eye-tracking experiment, Schlenter and Felser (2021) showed that when hearing the DAT argument in sentences such as (5a), where this argument is a possible recipient (*the crying baby*), native as well as non-native speakers of German with Russian as the first language anticipated a theme (*pacifier*) as the following argument. If the first postverbal argument carried ACC case (5b), listeners also initially considered the theme (*pacifier*) before starting to shift their gaze towards the plausible recipient (*mother*). Thus, listeners had initially interpreted the first postverbal argument as a recipient.

(5) *Der Vater überreicht* … — 'The father hands over …'
a. *dem schreienden Baby* *vorsichtig* *den Schnuller*
[the crying baby]$_{DAT}$ carefully [the pacifier]$_{ACC}$
'The father carefully hands the pacifier to the crying baby.'
b. *das schreiende Baby* *vorsichtig* *der Mutter*
[the crying baby]$_{ACC}$ carefully [the mother]$_{DAT}$
'The father carefully hands the crying baby to the mother.'

The competition between theme and recipient lingered in the non-native group, while the native group rapidly recovered from their initial misinterpretation. This finding is line with Hopp's (2015) study on transitive structures reported in Section 1.2, that native and non-native speakers of German may weigh the cues for argument structure differently.

While there is an indication that German comprehenders also expect a certain argument order after encountering a ditransitive verb, the methodology and materials used in previous studies do not allow firm conclusions in this regard: One reason why the IO-DO order could have been perceived as more prototypical than DO-IO in Schlenter and Felser (2021) is not only that the DAT-marked object preceded the ACC-marked object, but that the theme was also prototypically inanimate. Similarly, in Rösler et al. (1998), ACC-marked objects were always inanimate, so effects of case marking cannot be disentangled from effects of animacy. Thus, it is possible that participants in previous studies were guided by animacy, at least to some extent, and possibly also semantic associations between arguments (e.g., baby/pacifier).[3] Nevertheless, delayed reaction times for the order DO-IO as compared to IO-DO was also observed by Scherger et al. (2022), who used two animate objects (e.g., *Ich gebe dem Schwein das Schaf* – 'I give the pig the sheep'). Finally, while reaction times may indicate processing difficulties, they do not allow us to pin down when processing was disrupted and what caused it.

Next, we refer to the linearization hierarchies that have been proposed to constrain surface order in various languages, including German (e.g., Ellsiepen & Bader, 2018; Siewirska, 1993), to derive more concrete predictions as regards argument order in ditransitive structures. In four acceptability judgment experiments using magnitude estimation, Ellsiepen and Bader (2018) aimed to determine the weight of surface constraints in German sentences with several arguments. The final ranking derived from their experiments is shown in Fig. 1. The authors teased apart constraints by including, for example, passive sentences like in (6). In contrast to (6a), the syntactic constraint DAT > ACC and the semantic constraints animate > inanimate and recipient > theme are violated in (6b). In (6c), the constraint NOM > DAT is violated while the constraints animate > inanimate and recipient > theme are respected.

[3] The same applies to studies that have tested the exploitation of ACC and DAT case in ditransitive structures in Japanese. In both Kamide, Altmann, et al. (2003) and Mitsugi and MacWhinney (2016), the ACC-marked theme argument was prototypically inanimate.

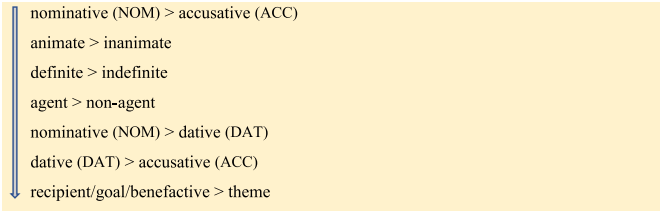


Fig. 1. Surface constraints identified by Ellsiepen and Bader (2018) in decreasing order with the highest weighted constraint at the top.

(6) *Der Internatsleiter sagte, ...* — 'The warden said …'

a.	*dass man*	*dem Erzieher*	*den Bericht*	*gebracht*	*hat*
	that one$_{NOM}$	[the educator]$_{DAT}$	[the report]$_{ACC}$	brought	has
b.	*dass man*	*den Bericht*	*dem Erzieher*	*gebracht*	*hat*
	that one$_{NOM}$	[the report]$_{ACC}$	[the educator]$_{DAT}$	brought	has
c.	*dass dem Erzieher*	*der Bericht*	*gebracht*	*wurde*	
	that [the educator]$_{DAT}$	[the report]$_{NOM}$	brought	was	
d.	*dass dem Erzieher*	*der Junge*	*gebracht*	*wurde*	
	that [the educator]$_{DAT}$	[the boy]$_{NOM}$	brought	was	

The ranking in Fig. 1 shows that German sentences received higher acceptability ratings if they conformed to NOM > ACC, even if an inanimate subject preceded an animate object. This supports the strong preference for the first argument encountered in a German sentence to be interpreted as the subject (typically also the agent). Crucially, the semantic constraint recipient > theme, which coincides with the syntactic constraint DAT > ACC, was the lowest ranked constraint. Hence, unlike for transitive structures with an agent and patient argument, we may find no or a reduced expectation towards the order of recipient and theme argument in ditransitive structures. While previous research has shown a preference for recipients to precede themes, this has been argued to be strongly associated with animacy (Häussler & Bader, 2012; Pappert et al., 2007). In the present study, we therefore control for animacy.

2. The present study

The present study investigates when native speakers of German exploit case marking on the sentence-initial NP in transitive sentences and the first object in ditransitive sentences to assign semantic roles to the arguments of the verb. Moreover, we want to know how argument interpretation is affected by expectations towards argument order. As shown previously, there seems to be a language-independent agent-first bias that should affect the processing of object/patient-initial transitive sentences and, likely, the predictive use of ACC on a sentence-initial NP. In contrast, less is known about argument order expectations in ditransitive sentences with two animate objects and the predictive use of ACC on a postverbal object. To norm our sentence stimuli and to get a first indication which object order is the preferred or prototypical one in our ditransitive sentences, we first conducted an acceptability judgment task (Experiment 1). Following stimulus norming, we conducted two experiments in the lab, in which we combined a picture selection task with the recording of eye movements (Experiment 2a) and then with hand movements (Experiment 2b). Prior to data collection, we preregistered the lab experiments on OSF (doi:10.17605/OSF.IO/Y8B23). In all experiments, participants provided informed consent. All procedures were in accordance with the Declaration of Helsinki. The processing of personal data (collected in the lab study) was assessed and approved by the Norwegian Center for Research Data (reference no. 548765). In line with local legislation, the experiments were exempt from an additional ethics vote as they included healthy adults, the methods were non-invasive, and the experiments did not pose a risk or physical/emotional burden to participants.

2.1. Animacy and world knowledge

Stimulus sentences and images were designed in a way that allows us to better control for the potentially confounding variables described above when testing for the use of case marking for argument interpretation during processing. In our stimulus sentences, we use fully reversible actions between animate referents and counterbalance the order of critical NPs to be able to rule out influence from lexical semantics and word knowledge. To determine when listeners decide on an argument interpretation, we present German native speakers with two visual scenes, one showing the target event and another showing the same event with role reversal, together with a subject/object-initial sentence (transitive structures) or a direct/indirect object first sentence (ditransitive structures). A similar forced-choice design has been used in a previous eye-tracking study on the use of case in Korean by Frenck-Mestre et al. (2019). Here, native but not non-native speakers displayed a preference for the target versus the competitor scene for both prototypical SOV and non-prototypical OSV sentences from the second NP onwards, which the authors took as an indication for the presence/absence of predictive processing. Similarly, Mitsugi (2017) employed a forced-choice design to investigate the use of case on NPs to predict active or passive voice in Japanese.

In Fig. 2, we show an experimental item for the transitive set in all four conditions. The critical manipulation is the order of arguments, that is, SVO (agent > patient, NOM > ACC) versus OVS (patient > agent, ACC > NOM). For counterbalancing reasons, two sentence versions are created for each argument order (for a similar design, see Knoeferle & Crocker, 2006). Thus, we can control for effects of NP order and potential differences in visual salience between scenes. As shown by Knoeferle et al. (2005) and Knoeferle and Crocker (2006, 2007), the event representation together with the case cue should facilitate immediate role assignment. In Knoeferle and colleagues' studies, German speakers used the visual information in event scenes with three referents, one patient and two possible agents, to anticipate not only stereotypical (e.g., wizard-jinxing-pilot), but also non-stereotypical agents (e.g., wizard-spying-on-pilot). Note, however, that in this work either the event or the target referent was depicted. In the current study, the same two referents perform the same action; thus visual information is no additional cue towards the upcoming referent, only towards the event (i.e., agent-action-patient).

In Fig. 3, we show an experimental item for the ditransitive set in all four conditions. Again, the critical manipulation is the order of arguments, here IO-DO (recipient > theme, DAT > ACC) versus DO-IO (theme > recipient, ACC > DAT) after the verb. Together with the order of NPs, we manipulate the perceptual salience of the case cue (*dem/den* vs. *der/die*), as further described below.

2.2. The perceptibility of the case cue

For masculine nouns, the NOM-marked definite article is *der*, the ACC-marked article is *den*. The two articles differ in the frequency at which they occur in written and spoken corpora (Strotseva-Feinschmidt et al., 2015); as illustrated in Table 1, *der* can also signal DAT and GEN for feminine nouns. In an ERP study, Strotseva-Feinschmidt et al. (2015) tested the auditory discrimination for both articles with German-speaking children and adults. They found that monolingual children were able to discriminate between *der* and *den* already at the age of three years. Yet, for both children and adults, recognition started earlier for the more frequent *der*: Deviant occurrence of *der* in comparison to standard occurrence of *der* elicited an early mismatch negativity (MMN) in children and adults, which was followed by a late negativity for the children but not the adults. Deviant occurrence of *den* elicited a late negativity in children and an MMN in adults that had a later onset than the MMN elicited after deviant *der*. Altogether, the results indicated that frequency facilitated the perceptibility and thus the processing of the article *der*.

				Target	Competitor
SVO, NP order A	*Der Drache* [The dragon]NOM	*besiegt schließlich* defeats finally	*den Prinzen* [the prince]ACC		
SVO, NP order B	*Der Prinz* [The prince]NOM		*den Drachen* [the dragon]ACC		
OVS, NP order A	*Den Drachen* [The dragon]ACC		*der Prinz* [the prince]NOM		
OVS, NP order B	*Den Prinzen* [The prince]ACC		*der Drache* [the dragon]NOM		

Fig. 2. Example of an item from the transitive structure set. The rightmost columns show the respective target and competitor scenes.

Der Krankenpfleger bringt … The (male) nurse brings …				Target	Competitor
IO-DO, NP order A	*dem* the-DAT	*Patienten morgens* patient in the morning	*die Ärztin* [the doctor]ACC		
DO-IO, NP order A	*den* the-ACC		*der Ärztin* [the doctor]DAT		
IO-DO, NP order B	*der* the-DAT	*Ärztin morgens* doctor in the morning	*den Patienten* [the patient]ACC		
DO-IO, NP order B	*die* the-ACC		*dem Patienten* [the patient]DAT		

Fig. 3. Example of an item from the ditransitive structure set. The rightmost columns show the respective target and competitor scenes.

In our study, we aim to account for differences in perceptibility between case marking cues in the following way: For half of our transitive sentences, we use a strong masculine noun for the first NP; for the other half the first NP includes a weak masculine noun. For weak masculine nouns, ACC is marked not only on the article, but additionally on the noun itself by the affix *-en* (see Table 1). Thus, by splitting our sentence stimuli into two halves, one with the double marking and the other with the simple marking of case, we also take into consideration the perceptual salience of the case cue in OVS sentences. One may reason that the processing of OVS sentences is easier when the less frequent article *den* is followed by a second case cue. A recent study investigating the comprehension of subject- and object-initial *wh*-questions in German showed a trend towards better identification of subject/agent and object/patient for nouns that have the double case marking (Binanzer et al., 2021). Note, however, this additional marking is often dropped in colloquial speech.

The perceptual difference between ACC *den* and DAT *dem* is also relatively small. This is different for feminine and neuter nouns. For feminine nouns, the ACC-marked definite article is *die* and the DAT-marked article *der*. In our investigation of ACC and DAT marking, we systematically vary the nouns' gender: For half of the ditransitive sentences, a masculine noun is part of the first postverbal argument (*dem/den Piraten* 'the$_{DAT/ACC}$ pirate'), and for the other half a feminine noun is part of the first postverbal argument (*der/die Prinzessin* 'the$_{DAT/ACC}$ princess'). Again, this manipulation allows us to detect potential differences in the perceptual salience of the case cue on the article and the effects on sentence processing. Note that in our materials DAT *der* and ACC *die* cannot be mistakenly taken for NOM, since at the time they are encountered, the subject/agent has already been mentioned (e.g., *Der Kapitän übergibt der/die Prinzessin …* 'The$_{NOM}$ captain surrenders the$_{DAT/ACC}$ princess …').

3. Experiment 1: acceptability judgment task

The aim of the acceptability judgment task (AJT) is to establish a baseline for the stimulus sentences when presented without a contextualizing scene. The AJT should show whether there are certain argument order preferences for our sets of stimuli and provide a first indication of acceptability differences between argument orders and sets. Moreover, the AJT will show whether reversing the order of NPs alters the acceptability of the sentences. The data and analysis code for this experiment is available on the Open Science Framework (OSF)

under https://osf.io/cg97f/.

3.1. Methods

3.1.1. Participants

Fifty native speakers of German completed the AJT. One person did not indicate whether s/he consented to data use, another person indicated that s/he did not acquire German before the age of six, so these two datasets were removed prior to analyses. The remaining 48 participants had an average age of 24 years, ranging from 18 to 49. Forty-four of them were females and four males. Nine participants indicated that they had learned German together with another language prior to the age of six.

3.1.2. Materials and design

For each experimental set, we created 24 items. All items include an animate subject and animate object(s). All NPs are definite, as definiteness is a factor that might influence the processing of argument order manipulations. For the transitive set, we use 12 obligatorily transitive verbs that — important for Experiment 2a and 2b — could be easily depicted and that clearly identified one event participant as agent and another as patient, such as *besiegen* ('defeat') or *ziehen* ('pull'). Each verb is used twice, each time with different event characters. For the ditransitive set, we use the same six verbs four times (*bringen* 'bring,' *holen* 'fetch,' *überlassen* 'leave to,' *übergeben* 'surrender,' *zeigen* 'show,' *empfehlen* 'recommend'), as there is a limited number of ditransitive verbs that sound natural in combination with two animate objects. As in the transitive set, every time a verb is repeated, it appears together with a new character combination. Event characters are either animals, human beings, or fictional characters. For the transitive events, all entities are referred to by a masculine noun, in order for the subject/agent or object/patient to be unambiguously marked by morphological case. For the ditransitive events, the two objects differ in gender with one being feminine and one being masculine. Gender is manipulated in such a way that in NP order A masculine nouns precede feminine nouns and vice versa for NP order B. The gender of the subject/agent varies.

We have a design with two categorical variables, argument order and NP order, with two levels each. The total of 48 items is distributed across four presentation lists with a Latin square design, so each participant encounters an item only once.

3.1.3. Procedure

The acceptability judgment task was web-based with the use of SoSci Survey (Leiner, 2019). The URL of the experiment was shared on social media and distributed in classes at two German universities. A click on the URL led to a welcome screen with general information about the experiment and data use. Participants were then assigned to one of the four lists. This way drop-outs on the first page could not affect list assignment, which reduced the risk of highly unbalanced lists. Participants provided information about their gender, age, and language knowledge between the ages of zero and six (options: only German, German and another language, or another language than German). Then, two example sentences were provided that only differed in the position of an adverb together with the 5-point scale used throughout the questionnaire. Participants were told that they were to assign a higher score to the sentence version they perceived as 'better' and that in the following, only one sentence would be displayed on the screen. This had to be rated according to its acceptability on a scale from 1 (very bad) to 5 (very good). The sentences were presented in a fully randomized order together with eight additional filler sentences that differed from the experimental sentences in word order and/or animacy of the nouns. The filler sentences should further ensure that participants used the entire range of the rating scale. Participants could only continue to the next sentence after they had selected a score. Otherwise, a message appeared on the screen, reminding them to give a response. Altogether, participants rated 56 sentences.

3.2. Results

We analyzed the rating responses by means of cumulative link mixed models in R (2022, R version 4.2.1), using the package *ordinal* (Christensen, 2019). In a first step, we only included argument order as a fixed effect into the models, and in a second step, we tested whether the difference between argument order conditions depends on the level of the factor NP order (A, B). Note that NP order is manipulated to control for effects of lexical semantics and world knowledge, and, for the ditransitive set, the gender of the noun and the corresponding perceptual salience of the case cue.

A first model for the transitive set with argument order as fixed effect and the maximal random effect structure shows a difference between argument orders: when changing the argument order from SVO to OVS, the likelihood of selecting 5 vs. 1 to 4 on the rating scale, 4 vs. 1 to 3, and so forth decreases by 3.90 on the ordered logit scale ($SE = 0.35$, $z = -11.18$, $p \leq 0.001$). For the ditransitive set, the maximal model with argument order as fixed effect also reveals a significant difference between argument orders when IO-DO is taken as the reference level (Est. $= -1.36$, $SE = 0.26$, $z = -5.32$, $p \leq 0.001$). Additional models that include NP order indicate that this factor does not mediate the acceptability of the sentences. The maximal model including argument order (SVO as reference level), NP order (A as reference level) and their interaction for the transitive structure set shows an effect of argument order (Est. $= -3.75$, $SE = 0.4$, $z = -9.45$, $p \leq 0.001$), but no effect of NP order (Est. $= 0.33$, $SE = 0.27$, $z = 1.22$, $p = 0.222$) and no interaction between argument and NP order (Est. $= -0.44$, $SE = 0.32$, $z = -1.40$, $p = 0.160$). The maximal model including argument order (IO-DO as reference level), NP order (A as reference level), and their interaction for the ditransitive set, again shows an effect of argument order (Est. $= -1.76$, $SE = 0.32$, $z = -5.53$, $p \leq 0.001$), but no effect of NP order (Est. $= -0.07$, $SE = 0.32$, $z = -0.23$, $p = 0.82$) and no interaction between argument and NP order (Est. $= 0.6$, $SE = 0.49$, $z = 1.22$, $p = 0.22$). Thus, changing the order of NPs (e.g., 'the$_{DAT}$ (male) patient the$_{ACC}$ doctor' vs. 'the (female) doctor$_{DAT}$ the patient$_{ACC}$'), which includes a reversal of feminine and masculine NPs, does not change the rating significantly. The probabilities of selecting 1 to 5 on the rating scale as predicted by the two latter models are shown in Fig. 4.

3.3. Discussion

The results of the acceptability judgment task confirm our predictions for the transitive structure set that there is a strong preference for the order SVO (i.e., agent-first) and that OVS is largely dispreferred. Thus, the markedness of the latter structure is clearly reflected in these offline acceptability judgments. Additionally, the AJT results indicate that for the ditransitive structure set, there is a markedness difference between IO-DO and DO-IO orders even if both objects are animate. While the difference between argument orders is less pronounced than for the transitive structure set, this provides us with some indication that IO-DO is the preferred order. A possible explanation for this finding is that this order is also overall more frequent (Häussler & Bader, 2012). The results for both the transitive and ditransitive set show no effect of NP order on offline acceptability judgments. While we cannot rule out that there was one for single items, this provides us with some indication that the events are reversible (see OSF repository for the ratings per item). It should be pointed out that the sentences in the ditransitive set received lower ratings overall than (prototypical) sentences in the transitive set. We can only speculate that this is because double-object constructions with an animate theme are not as prototypical as double-object constructions with an inanimate theme, especially without a context. However, this does not affect the critical issue in our study, which is the manipulation of argument order. Moreover, in the eye-tracking and mouse-tracking experiments there will always be a contextualizing scene.

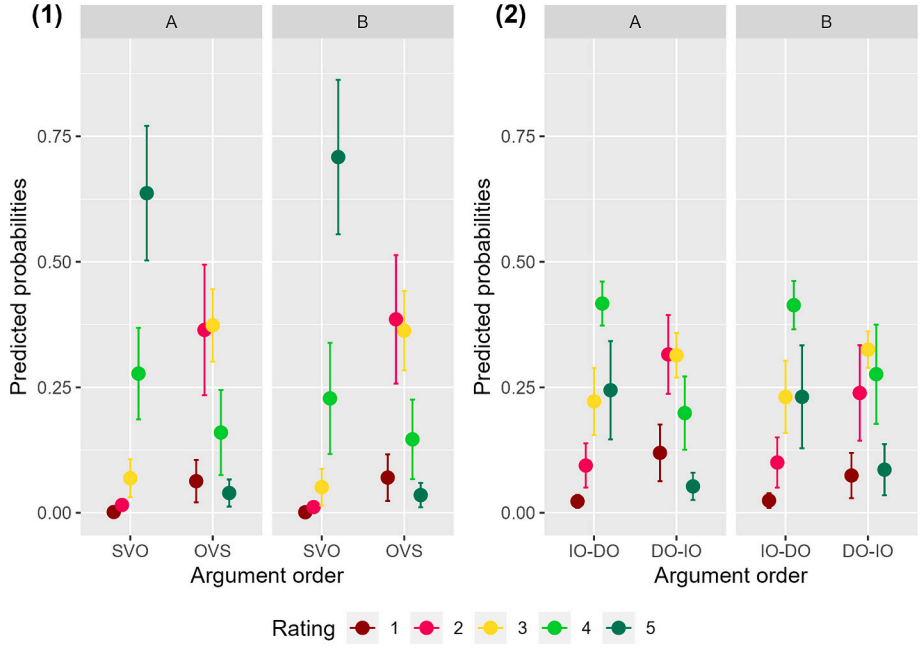


Fig. 4. Predicted probabilities of selecting 1 (very bad) to 5 (very good) on the rating scale per argument and NP order in (1) the transitive structure set and (2) the ditransitive structure set. Left panels show NP order A, right panels NP order B. Error bars indicate 95 % confidence intervals.

4. Experiment 2a: eye-tracking

In an experiment in the lab, we record German native speakers' eye movements while they listen to the sentences tested in Experiment 1 and select the respective target scene to answer the following research questions:

(1a) When do listeners utilize the morphological case marked on the first argument (ACC) in the absence of further cues to assign semantic roles (agent, patient) in transitive structures?
(2a) When do listeners utilize the morphological case marked on the first postverbal argument (ACC, DAT) in the absence of further cues to assign semantic roles (recipient, theme) in ditransitive structures?

Critically, we want to know whether listeners match the sentence to the target scene prior to the onset of the NP2 in the transitive set and prior to the second object in the ditransitive set in an experimental design in which we control for influences from lexical-semantics and world knowledge. Moreover, we explore how the perceptibility of the case cue may influence real-time processing. In addition to eye movements, we measure listeners' accuracy and reaction times.

If listeners use case marking predictively, looks to a target scene should diverge from looks to a competitor scene before the NP2 is encountered in transitive sentences and before the second object is encountered in ditransitive sentences. Importantly, prediction should be observed for all argument orders to demonstrate that listeners used the case cue and did not rely on linear order only: If listeners prefer the target scene when listening to a SVO sentence, then we do not know whether case marking or linear order was the primary cue. However, if listeners prefer the target scene when listening to an OVS sentence prior to the onset of the subject/agent, this indicates an effect of case. The data and analysis code for this experiment is available on OSF under https://osf.io/e9naj/.

4.1. Method

4.1.1. Participants

Thirty-nine native speakers of German participated in the lab study,[4] which was conducted at the University of Potsdam. The participants' age ranged from 18 to 55 years with an average age of 26 years. Twenty-eight of them were females and eleven males. All participants were born in a German-speaking country. Two participants had acquired an additional language from birth (Swiss German, Turkish). Thirty-four were right-handed, four left-handed and one participant was ambidextrous. Since all reported that they would use the right hand to hold a mouse, all participants performed the tasks using their right hand. All participants included into the analyses had either normal or corrected to normal vision and hearing.

4.1.2. Materials and design

Except for one item in the ditransitive set that had to be revised, the same stimulus sentences as in the previous acceptability judgment task are used in Experiment 2a but are now presented auditorily.[5] Prior to the lab experiments, we conducted a web-based picture selection task to assess the suitability of the visual stimuli designed for our study (see OSF repository). Based on the results of this additional norming experiment, seven out of 96 scenes were replaced.[6]

Our sentence material includes two experimental sets, a transitive and a ditransitive structure set, each including 24 experimental items in four conditions that are distributed across four lists. The two sets function as fillers for each other, no additional fillers are added. The sentences were recorded in a sound attenuated booth at a rate of 44.1 kHz.

[4] This exceeds the planned sample size of 36 in the pre-registration. We allowed a few more participants to sign up as a back-up for no-shows and potential problems with calibration in the eye-tracking experiment.

[5] Originally, one item included the noun *Angeklagte* ('defendant'). Unlike for all other NPs in this set, ACC is marked both on the article and the noun. Thus, this noun was replaced by *Betrügerin* ('swindler'). A picture selection task that was conducted to norm the pictures included the revised item.

[6] The new scenes were shown to two German native speakers, who considered them as a better match.

They were read by a female native speaker of German with an average speech rate of 4.32 syllables per second (min.: 3.35, max.: 5.05). This speech rate falls into the range that is considered normal in a literature review by Fernandez et al. (2020), who detected a noticeable decline in the predictive ability of young native speakers of English at a speech rate of 6 syllables per second. As regards prosody, the speaker aimed to read the OVS sentences with a neutral intonation, that is, without a contrastive pitch accent on the sentence-initial object. To control for the length and prosody of the critical segment for a prediction effect in Experiment 2a, we spliced the critical segment that is identical between sentence versions (transitive set: verb + adverb, ditransitive set: noun of first object + adverb) from one sentence to all other sentence versions in Praat (Boersma & Weenink, 2022). Table 2 shows the mean length and ranges in milliseconds (msec) for the critical segment for a prediction effect for the two sets of stimuli as well as the average speech rate per set.

To ease the visual recognition of the characters used in the sentences, these were selected in such a way that they could be easily depicted (e.g., a police officer in a uniform). A complete list of all items and scenes is available on OSF. In accordance with a general preference of German speakers to linearize events from left to right (e.g., Suitner et al., 2021), agents in transitive event scenes are almost always depicted to the left of other event characters (see Fig. 2).[7] The agent in ditransitive events is either positioned to the left of theme and recipient or between theme and recipient. During the experiments, the position of the target and competitor scene is counterbalanced so participants see an equal number of target scenes on the left and the right of the screen.

Like in Experiment 1, we have two categorical variables with two levels each: argument order and NP order. Note that for the transitive set, we include both NP orders, A and B, to control for potential effects of lexical semantics and world knowledge (e.g., 'the$_{NOM}$ dragon defeats the$_{ACC}$ prince' and 'the$_{NOM}$ prince defeats the$_{ACC}$ dragon'). Potential differences in perceptual salience (e.g., *den Drachen* 'the-ACC dragon-ACC vs. *den Clown* 'the-ACC clown') are tested separately in exploratory analyses. For the ditransitive set, NP order and perceptual salience of the case cue are linked, as one NP is feminine (*der* vs. *die*) and the other masculine (*dem* vs. *den*). This may affect real-time processing, although it did not seem to affect offline judgments (Experiment 1).[8]

4.1.3. Procedure

Prior to the experiments, we obtained participants' informed consent. The experimental session started with an ocular dominance test. The test served the purpose of determining the participant's dominant eye for the eye-movement recording. After completion of the eye-tracking experiment, the participants filled in a background questionnaire, while the experimenter set up the mouse-tracking experiment. The experimental session ended with a short debriefing and lasted altogether about 45 min. Participants either received course credit or a 10-Euro gift card for their participation.

For the eye-tracking, we used an EyeLink Portable Duo in the remote mode. Participants were seated in front of the display laptop (15.6" Full HD, 1920 × 1080 resolution, 144 Hz) to which the camera was attached. After an instruction screen that reminded participants not to move during the experiment, the camera setup started, including a 5-point calibration and validation. If calibration/validation was successful, the experimenter started the recording (500 Hz sampling rate). Fig. 5 shows an example trial. First, participants were given a preview of the two scenes for a duration of 2.5 s. Then a fixation cross appeared, which the participants had to fixate on for 500 consecutive msec. If they did not fixate on the cross, the experiment went back to the camera setup, and the aborted trial would be resumed later. Thus, the fixation cross not only served the purpose of attracting people's eye movements away from the two scenes prior to the audio presentation, but also functioned as a drift check. If participants accurately fixated on the cross, the fixation cross would disappear, and the audio was played. With the offset of the case cue in a trial, a mouse cursor appeared. Participants were instructed to select the scene that matched the sentence by clicking on it as soon as they felt certain about their choice. A trial ended when the participant clicked on one of the scenes with the mouse. Afterwards a start button would be shown, on which participants had to click to continue with the next trial. Trials appeared in a randomized order. Each recording started with four practice trials. The eye-tracking experiment was built in Experiment Builder (SR Research, Version 2.3.38).

Table 2
Mean length and ranges for the critical segment for a prediction effect.

	Segment	Mean length	Min. length	Max. length	Speech rate (syllables per second)
Transitive set	verb + adverb	962	807	1104	4.12
Ditransitive set	noun + adverb	1117	844	1576	4.52

4.2. Results

Two participants were excluded from the eye-tracking analyses, one because the calibration was impossible with the glasses and one because the wearing of a hearing aid in combination with the headphones used for the eye-tracking experiment was problematic. Below, we start with the accuracy and response times before we report the eye-movement analyses. For both accuracy and response time data, we also computed a model that included NP order (to check whether NP order A and B were different from each other) and, if there was no effect of or interaction with NP order, collapse both orders and report the output of the model that includes argument order as a (treatment-coded) fixed effect. In models that test for effects and the interaction of both argument and NP order, the levels of each factor are coded as 0.5 and − 0.5. If the respective maximal model did not converge, we simplified the random-effects structure and report the results from the converging model with the lowest value for the Akaike information criterion (AIC). If varying slopes are included and if not indicated otherwise, the random-effects structure includes all possible correlations. Analyses were conducted in R (2022, version 4.2.2), using the *lmerTest* package (Kuznetsova et al., 2017).

4.2.1. Accuracy

In the pre-registration, we indicated as an exclusion criterion an error rate of >20 % in the SVO condition. On this basis, we had to exclude one participant from the analyses of the eye-tracking experiment, leaving us with data from 36 participants. Table 3 summarizes the accuracy scores for the transitive and ditransitive set in Experiment 2a.

We analyzed the correct (1) and incorrect (0) responses by means of generalized linear mixed-effects models (GLMM). A GLMM for the transitive set with argument order as fixed effect showed a significant effect (Est. = −1.556, $SE = 0.455$, $z = -3.420$, $p \leq 0.001$), indicating that participants were less accurate in selecting the target scene for OVS than for the reference level SVO.

For the ditransitive set, a GLMM that tested for effects of argument and NP order on accuracy as well as their interaction, including by-subject and by-item varying intercepts, revealed an effect of argument

[7] Exceptions are events like *A lifting B* and *A overtaking B*. Another reason for keeping the position of agent and patient constant in our study is that the position of the agent relative to other event characters may affect sentence processing (Schlenter & Penke, 2022) and thus varying its position may introduce a confounding variable.

[8] The results of the picture norming experiment showed an interaction effect on response times. The longest response times were observed for the order DO-IO with NP order A (ACC signaled by masculine *den*).

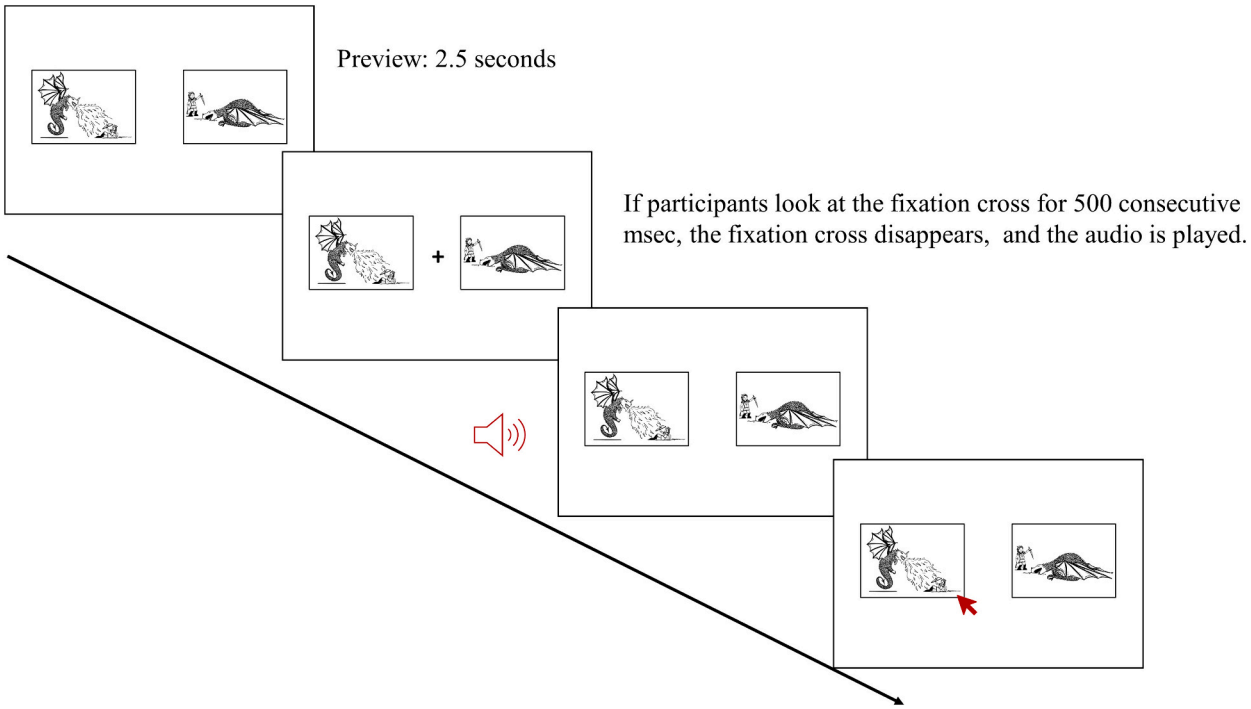


Fig. 5. Procedure in the eye-tracking experiment.

Table 3
Number and percentages of correct and incorrect responses in Experiment 2a. Numbers and percentages are further provided separately for the simple (article) and the double marking (article + noun) of ACC in the OVS condition of the transitive set and for masculine and feminine articles in the ditransitive set.

	Correct	Incorrect
Transitive set		
SVO	**425 (98 %)**	7 (2 %)
OVS	**405 (94 %)**	27 (6 %)
article	201 (93 %)	15 (7 %)
article + noun	204 (94 %)	12 (6 %)
Ditransitive set		
IO-DO	**382 (88 %)**	50 (12 %)
masc. (*dem*)	193 (89 %)	23 (11 %)
fem. (*der*)	189 (88 %)	27 (12 %)
DO-IO	**337 (78 %)**	95 (22 %)
masc. (*den*)	158 (73 %)	58 (27 %)
fem. (*die*)	179 (83 %)	37 (17 %)

order (Est. = 0.794, $SE = 0.196$, $z = 4.053$, $p \leq 0.001$), but not NP order (Est. = 0.214, $SE = 0.195$, $z = 1.096$, $p = 0.273$). Yet, there was a significant interaction between argument and NP order (Est. = −0.808, $SE = 0.391$, $z = -2.066$, $p = 0.039$). Pairwise comparisons that were conducted with the help of the package *emmeans* (Lenth, 2023, version 1.8.4–1) showed that for the order DO-IO with a masculine NP as first postverbal argument (i.e., ACC marked via *den*) the likelihood of selecting the target scene was lower than for all other conditions (*dem*, *der*, *die*); see Table A.1 in the appendix.

4.2.2. Response times

For the transitive set, response times (RT) were measured from NP1 offset. For the ditransitive set, RTs were measured from the offset of the article of the first object/postverbal argument. Here, a response means the selection of one of the two scenes. Below, we report the results from linear mixed-effects models (LMM) on log-transformed RTs. RTs were log-transformed to approach a normal distribution. For better interpretability, we also report the back-transformed values as predicted by the models as well as 95 % confidence intervals (in square brackets). Note that for the analyses of RTs, only correctly answered trials were considered, that is, trials in which participants selected the target scene.

Since the output of a LMM that included the effect of and interaction with NP order did not indicate an influence of this variable on RTs for the transitive set, we report the output of a LMM with argument order as the only fixed effect and SVO as the reference level. The output of the maximally specified model showed that participants needed more time to respond to the order OVS than to SVO (Est. = 0.102 $SE = 0.028$, $t = 3.598$, $p = 0.001$). The back-transformed values as predicted by the model are 2522 msec [2210, 2877] for SVO and 2792 msec [2417, 3225] for OVS. In an exploratory analysis, we used a subset of the data to test for the difference between strong and weak masculine nouns in the OVS condition. Unlike strong nouns, weak nouns mark ACC both on the article and noun. There was no indication that participants needed less time to respond after the double marking of case.

The output of a LMM that included argument order, NP order, and their interaction as fixed effects for the ditransitive set showed that RTs for the order DO-IO depended on the level of NP order (A vs. B). The model with by-subject and by-item varying intercepts as well as by-item varying slopes for argument order showed no effect of argument order (Est. = 0.004, $SE = 0.039$, $t = 0.107$, $p = 0.916$), a significant effect of NP order (Est. = −0.058, $SE = 0.024$, $t = -2.427$, $p = 0.015$), and a significant interaction between argument and NP order (Est. = 0.138, $SE = 0.048$, $t = 2.887$, $p = 0.004$). Pairwise comparisons conducted to follow-up on the interaction revealed a significant difference between a feminine NP and a masculine NP as first postverbal argument for the order DO-IO, with longer RTs for the masculine NP (i.e., *den*); see Table A.2 in the appendix. Averaged over the levels of NP order, predicted RTs were 3451 msec [2996, 3974] for the order IO-DO and 3436 msec [2979, 3963] for the order DO-IO.

4.2.3. Eye-movement data

The eye-tracking data were loaded into DataViewer (SR Research, Version 4.3.1) and down-sampled to 50 Hz. In a next pre-processing step, we loaded the data into R and excluded all rows with blinks and saccades. Then, we plotted the data to consider the fixation proportions

for the target and competitor scene as well as the background from sentence onset; see Fig. A.1 for the transitive and Figs. A.2 and A.3 for the ditransitive set. Background looks were included to show that there were no differences between experimental conditions. However, visual inspection of the time course graphs reveals the following for the ditransitive set: In trials in which the first postverbal argument is a masculine NP, there is an overall preference for the target scene that emerges already prior to the case cue (*dem/den*) that should enable participants to identify this scene. To better understand the initial target preference in these conditions, we also plotted the time course graphs for those trials in which participants selected the competitor scene; see Fig. A.4 in the appendix. As usual in eye-movement analyses, these trials were originally not included into the graphs or later analyses. However, they might be informative as regards the current baseline effect. In the graphs for incorrectly answered trials, the initial target preference in the prediction window is indeed absent, and we explain the pattern as follows: Participants seemed more likely to select the target scene when they were already looking at it while listening to the articles *dem* and, even more so, *den*. Note that we were more likely to obtain a baseline effect for the ditransitive set as participants had already heard the first part of the sentence (subject/agent verb article) after the presentation of the fixation cross, so it was more likely that their gaze was already on one of the two scenes when encountering the case cue.

In the statistical analyses below, we focus on target and competitor scene fixations starting at the offset of the case cue(s), that is, after NP1 offset in the transitive set and after the offset of the article of the first postverbal argument/object in the ditransitive set.

4.2.3.1. Cluster-based permutation analyses. Following a reviewer's suggestion, prior to the pre-registered divergence point analyses, we ran several cluster-based permutation analyses (CBPA) to identify clusters in which conditions significantly differ (for a review on the differences between eye-tracking analyses, see e.g., Ito & Knoeferle, 2023). Here we use the analysis code provided by Minor et al. (2022). First, we fitted a generalized linear mixed-effects model analyzing the likelihood of target looks in the transitive set depending on argument order condition (SVO vs. OVS). Participants and items are included as random intercepts. The model is run on each 20-msec time bin from NP1 offset up to 1160 msec later. This approximates the length of the verb + adverb segment with a mean length of 962 msec plus 200 msec to account for eye-movement latency. Adjacent time bins with a *p*-value below 0.05 are identified as clusters, and the *z*-values for all time bins within a cluster are summed to generate the sum statistics for that cluster (see Fig. 6 that shows the clusters that were detected for the comparison between SVO and OVS). In a next step, for each participant, the condition labels are randomly reassigned, and the model is run to identify significant clusters of time bins for the permuted data. The sum-statistic for each cluster is calculated, and the largest sum-statistic is stored. This step, including the permutation and calculation of the largest sum-statistic, is repeated 1000 times to generate a distribution of sum-statistics under the null hypothesis that the order of arguments does not have an effect. The sum-statistics for the original dataset and the null distribution is then used to obtain a *p*-value for the clusters.

The *p*-value for the cluster that was identified between 640 and 820 msec is 0.069, indicating a marginally significant effect of argument order. We ran the same analyses again, now focusing on the OVS trials and the comparison between the simple and double marking of ACC; see Fig. A.5 in the appendix. The CBPA detected several clusters within the first 320 msec after NP1 offset, in which the likelihood of target looks was higher for the double than for the simple marking of case, but none of these differences were statistically significant.

Next, we tested for differences in the ditransitive set. Here, the length of the analysis window is 1320 msec, spanning the noun + adverb segment with a mean length of 1117 msec plus 200 msec for eye-movement latency. First, we tested for potential differences between the IO-DO and DO-IO order for each of the NP order/gender conditions. No clusters were detected for NP order B; that is, following a feminine article (*der* versus *die*). For NP order A, as illustrated in Fig. 7, we obtain several clusters. Initially, we observe a higher likelihood of target looks for the DO-IO in comparison to the IO-DO order. Then, we observe a shift, and we obtain several clusters due to a lower likelihood of target looks after DAT *dem* relative to *den* towards the end of the analysis window (no significant differences). The pronounced baseline preference for *den* also becomes visible in the comparison between NP order/gender conditions within argument orders. In Fig. 8, we show the graph for the IO-DO order and, in Fig. 9, for the DO-IO order. Only one cluster with a single time bin is detected for the IO-DO order, and the difference is non-significant. For the order DO-IO, on the other hand, multiple clusters are detected. The earliest clusters are the result of a higher likelihood of target looks for the masculine article *den* relative to feminine *die* (no significant differences). Second, we obtained two late clusters, one between 840 and 900 msec and one from 920 until 1320 msec with the reverse pattern, the latter of which has a *p*-value of 0.04. The visible shift points to a difference in the development of a target preference for the ACC articles *den* and *die*.

4.2.3.2. Divergence point analyses. As stated in the pre-registration of this study, we originally set out to analyze our data by means of divergence point analyses (DPA). Our aim was to determine the point in time when there is a 'sustained' target-over-competitor preference, more specifically, whether this point falls into the time window for prediction, and to pin down the difference between argument orders. Following Stone et al. (2020), we first ran a one-sample *t*-test on target fixation proportions against chance (i.e., 50 %) over aggregated items. We set as a divergence point the first time point in a run of at least 10 consecutive time points with significant *t*-values (i.e., 200 msec). Next, new datasets were generated by means of non-parametric bootstrapping, where the data were resampled by replacement within the categories subject, timepoint (1 = 0 msec, 2 = 20 msec, etc.) and scene type (target/competitor); a new divergence point was estimated after each resampling. Our analyses deviate slightly from the pre-registration, for example, as regards the length of the analyzed windows, as described below, and number of iterations. To get more stable estimates, we increased the number of iterations from the planned 1500 to 2000.

For the transitive set, visual inspection of fixation proportions indicated that target fixations in the OVS condition reach a peak only around 1500 msec after NP1 offset, and we therefore chose this time window for the DPA. Fig. 10 shows the graphs for the SVO and OVS orders, to which we added the bootstrapped divergence points and, as a measure of temporal uncertainty, the 95 % percentile confidence intervals (CI; based on 1952 bootstrap replicates). The estimated mean divergence point for the SVO condition is 493 msec [440, 620] after NP1 offset, indicating that listeners showed a preference for the target scene well in advance of NP2 onset (~962 msec after NP1 offset). For the OVS condition, the bootstrap mean lies at 1054 msec [740, 1320]. Given the assumption that it takes around 200 msec for the eye to initiate a movement, a sustained target preference indicating an NP1-as-patient interpretation (and, correspondingly, NP2-as-agent interpretation) was thus obtained at a time point at which information from the NP2 was still unlikely to have had an influence. The difference between conditions is 561 msec [112, 820]. Notably, there is a wide CI. Moreover, as shown in Figs. A.6 and A.7 in the appendix, the empirical mean for the OVS order lies outside the range of the 95 % CI, which is also the case when we plot the bootstrap distribution of the difference in divergence points between SVO and OVS.

Descriptively, there were two clusters of divergence points in the 1500 msec time window after NP1 offset for OVS: an early cluster around 300 msec and a later cluster around 1000 msec, that is, around the average onset of NP2. A CBPA for the OVS trials detected a cluster between 240 and 320 msec in which the likelihood of target looks was

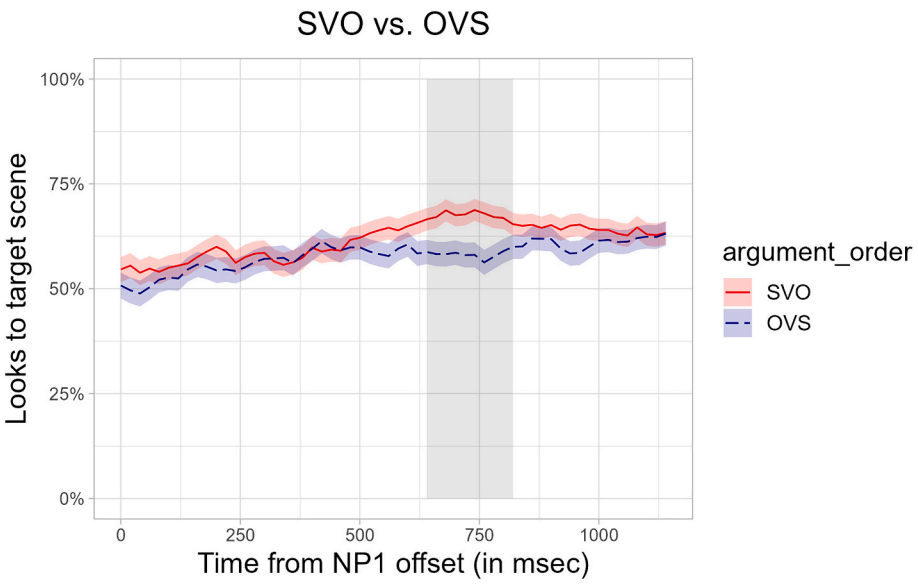


Fig. 6. Fixations to the target in 20-msec time bins starting from NP1 offset for the transitive set by argument order: SVO (solid, red) vs. OVS (dashed, blue). Shaded areas represent clusters identified via a cluster-based permutation analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

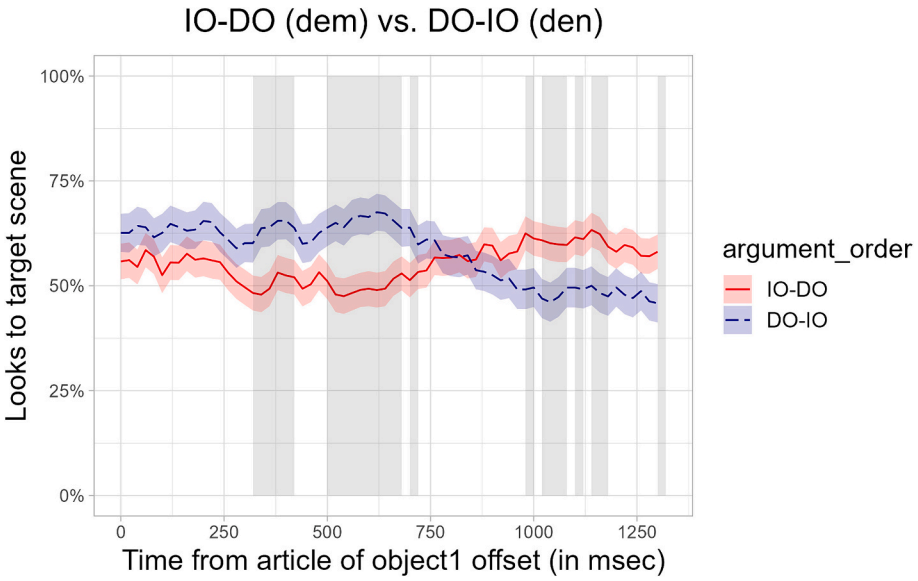


Fig. 7. Fixations to the target in 20-msec time bins starting from article of object1 offset for NP order A (masculine NP precedes feminine NP) by argument order condition: IO-DO (solid, red) vs. DO-IO (dashed, blue). Shaded areas represent clusters identified via a cluster-based permutation analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

higher for the double than the simple marking of ACC, but the difference was not statistically significant. Thus, while the marking of ACC case may seem to be a likely explanation, we must attribute these early onsets to unexplained variance or noise in our data.

Given the baseline effect in the ditransitive set for the sentence conditions in which a masculine NP preceded a feminine NP after the verb (NP order A), it was difficult to pursue the planned DPA. In the following, we focus on the subset of the ditransitive data in which a feminine NP preceded a masculine NP after the verb (NP order B). We computed a DPA for the time window including the noun of the first object/first postverbal argument and adverb plus 200 msec for eye-movement latency (1320 msec after article offset). In Fig. 11, we show the graphs for the IO-DO order (article *der*) and DO-IO order (article *die*) to which we added the bootstrapped divergence points and, as a measure of temporal uncertainty, the 95 % CIs (based on 1397 bootstrap replicates)[9]; for the bootstrap distributions of divergence points per argument order and the difference in divergence points, see Figs. A.8 and A.9 in the appendix. For both argument orders, the bootstrap means lies before the average onset of the second object, that is, the final argument: for IO-DO at 868 msec [720, 1080] and for DO-IO at 816 msec [700, 1080]. Thus, for the order DO-IO, the onset of a sustained target-over-competitor preference was slightly earlier than for the order IO-DO. However, as the CI for the difference between divergence points

[9] A likely explanation for the replication failure — given that we have 2000 iterations, ideally, we should have as many replicates — is the sparsity of the data (we can only use a subset of the ditransitive set) and because the fixation curves are close together (i.e., there are no divergence points) in the beginning.

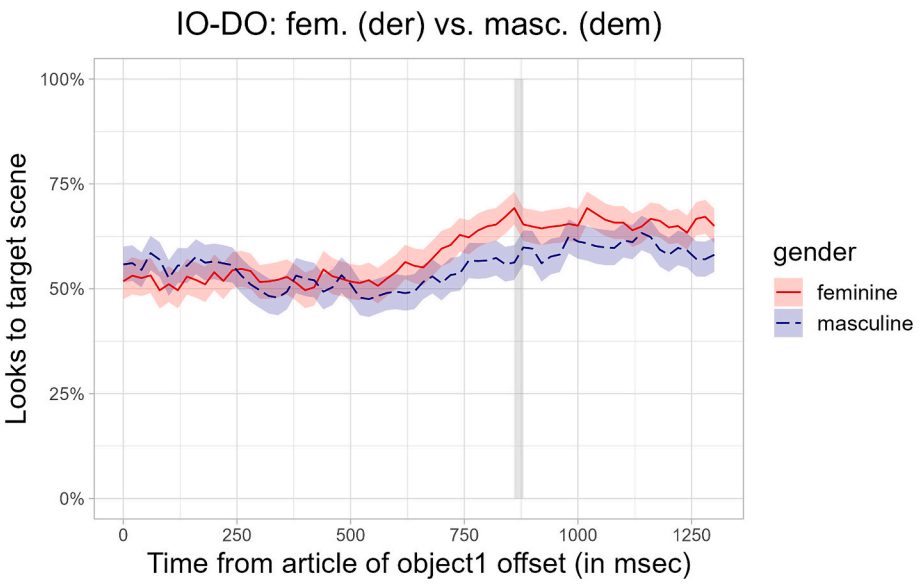


Fig. 8. Fixations to the target in 20-msec time bins starting from article of object1 offset for the IO-DO by NP order/gender condition: feminine (solid, red) vs. masculine (dashed, blue). Shaded areas represent clusters identified via a cluster-based permutation analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

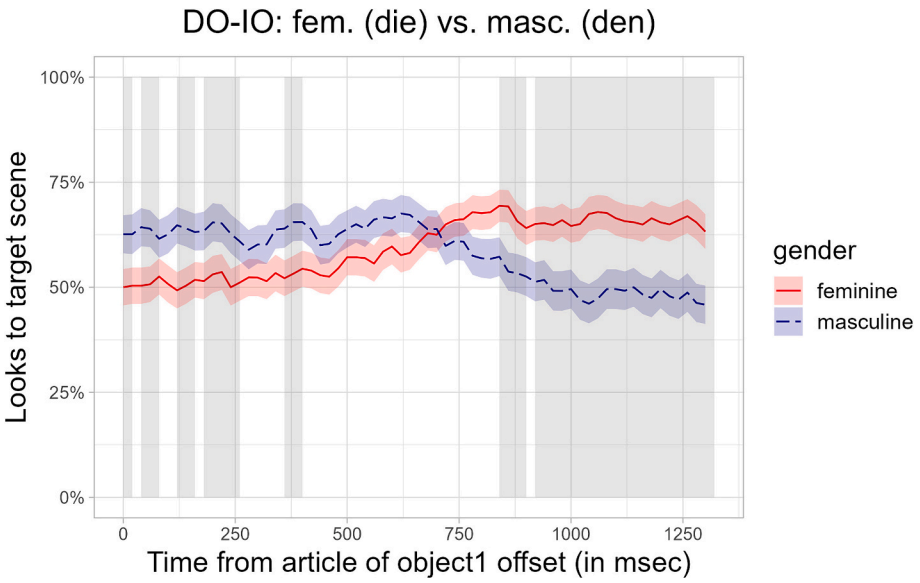


Fig. 9. Fixations to the target in 20-msec time bins starting from article of object1 offset for the DO-IO by NP order/gender condition: feminine (solid, red) vs. masculine (dashed, blue). Shaded areas represent clusters identified via a cluster-based permutation analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contains zero [−300, 240], we must assume that it is not reliably different, which aligns with the results of the CBPA that detected no clusters when testing for the difference between argument order conditions for this NP order.

4.3. Discussion

In Experiment 2a, we addressed when case marking is exploited to identify semantic roles. Research question (1a) addressed the timing of the case cue in transitive sentences. To answer (1a), ACC case on the NP1 was exploited prior to the onset of the NP2 to interpret the unfolding sentence and identify the target scene. Overall, our eye-movement data replicate previous research indicating that German listeners use the marking of ACC on a sentence-initial NP to predict the semantic role of the upcoming argument (e.g., Hopp, 2015; Özge et al., 2022), here in a design in which we controlled for influences from semantics, world knowledge, and visual salience by presenting fully reversible actions and counterbalancing the order of NPs. We found a marginally significant difference between argument orders in a CBPA as regards the likelihood of looking at the target scene. A DPA showed a sustained target-over-competitor preference within a critical time window for prediction that, however, had a later and more variable onset for OVS than for prototypically ordered SVO sentences. While looks to the target scene quickly diverged from looks to the competitor scene after NP1 offset for SVO when the agent of the event could be identified by both order (+1st position) and case (+ NOM), there was more variance for OVS when the agent was yet to come. In addition, participants were more accurate and faster in selecting the target scene for the order SVO than for the order

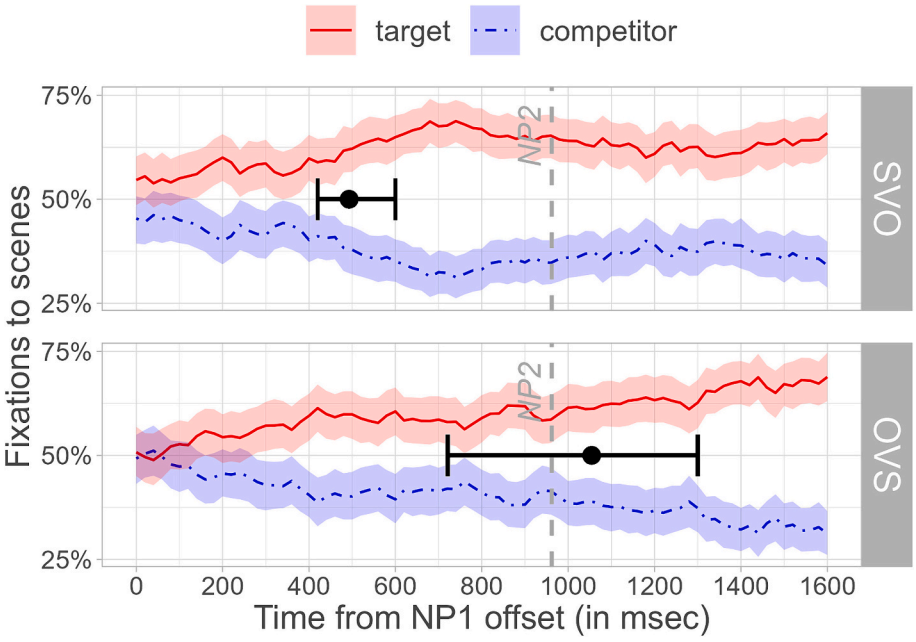


Fig. 10. Fixations to the target (solid, red) and competitor scenes (dot dashed, blue) per argument order in the transitive set to which we added the bootstrapped divergence points and 95 % CIs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

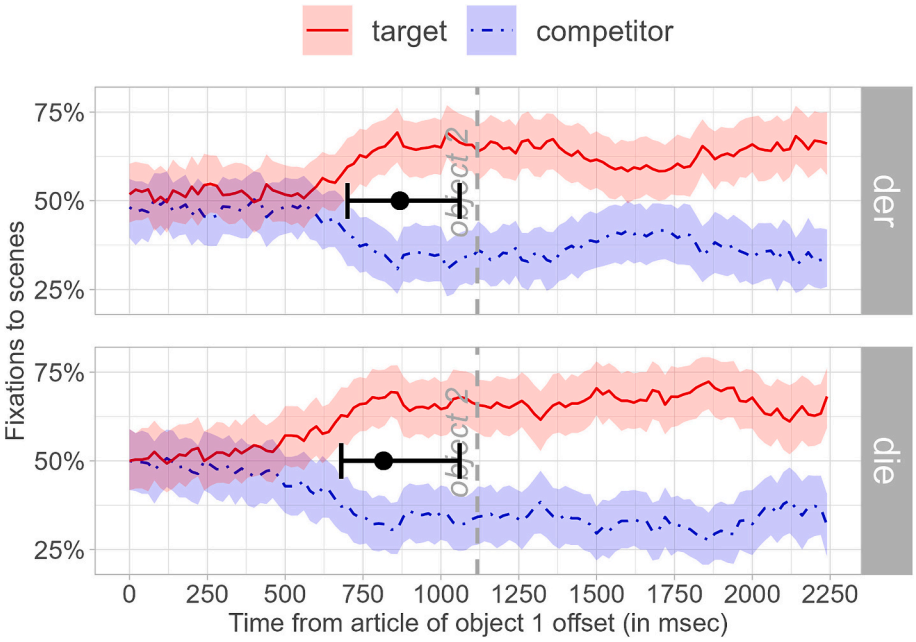


Fig. 11. Fixation proportions for the target (solid, red) and competitor scenes (dot dashed, blue) for the subset with a feminine article (*der/die*), IO-DO at the top and DO-IO at the bottom, to which we added the bootstrapped divergence points and 95 % CIs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

OVS, in line with previous findings that object-initial orders are more difficult to process than subject-initial orders. In Experiment 2b, we further address the possibility that an NP1-as-agent interpretation is not immediately rejected but maintained. An exploratory examination of the simple and double marking of ACC provided no strong indication that listeners benefited from a second case cue. Only descriptively did we see a higher likelihood of target looks for the double versus the simple marking shortly after NP1 offset (Fig. A.5).

In research question (2a), we asked when listeners use case marking on the article of the first object following a ditransitive verb to identify the semantic roles of the arguments. As a quick recap, the first postverbal argument in the ditransitive sentences was either marked as DAT or as ACC, so the first object either corresponded to the IO/recipient or the DO/theme in the event (e.g., 'bringing'). If DAT and ACC were marked by *der* and *die* respectively (feminine gender), participants preferred the target over the competitor scene prior to the onset of the second object, indicating that they had assigned a semantic role to the first object and, in a predictive fashion, already to the not-yet-encountered second object. As regards differences between DAT and ACC for argument interpretation, we found none between *der* and *die* in both the CBPA and the DPA. However, if DAT and ACC were marked by *dem* and *den* (masculine gender), a different picture emerged.

Unfortunately, statistical analyses for the latter conditions — in particular for *den* — were impaired by a baseline effect that we attributed to participants being more likely to select the target scene when already fixating on it at the point in time they encountered the first case cue. At least descriptively, a target preference after DAT *dem* developed similarly as for feminine *der* (see Fig. 7), whereas it developed much later for ACC *den*. A deviant pattern for *den* also showed up when we analyzed accuracy and response times. The most incorrect responses were elicited after *den*. Moreover, listeners were slower in selecting the target scene when ACC on the first object was marked by *den* than when it was marked by *die*.

The results for the ditransitive set largely align with previous findings by Schlenter and Felser (2021) for native and non-native speakers of German, showing a predictive use of morphological case after a ditransitive verb in a design with separate images in a visual display. In their study, the first postverbal argument was always a feminine or neuter NP, so case would be marked additionally on an included adjective. The authors reasoned that the perceptual salience of the case cue, together with the less marked argument order manipulation (in comparison to OVS in other studies) influenced the results. The current findings lend support to the assumption that perceptual salience influences predictive sentence processing; we will return to the question about markedness in the general discussion.

5. Experiment 2b: mouse-tracking

Complementary to the eye-tracking experiment, we record German native speakers' hand movements during scene selection. In mouse-tracking, we get one movement trajectory per trial. Mouse-tracking allows us to measure whether and how strongly participants are attracted by the target scene of the prototypical order. This is done by comparing the degree of curvature between argument order conditions. The mouse-tracking experiment thus gives us an indication about the amount of conflict between the two response options, that is, between target and competitor scenes (e.g., Freeman, 2018; Kieslich et al., 2020). By means of mouse-tracking, we aim to answer the following research questions:

(1b) Do listeners maintain an agent-first interpretation in non-prototypically ordered transitive structures (OVS) despite unambiguous case marking?
(2b) Do listeners have an expectation towards a certain argument order after a ditransitive verb? If yes, do listeners maintain an initial interpretation despite unambiguous case marking?

We added the mouse-tracking experiment to better understand the processes in the eye-tracking experiment, that is, whether a delayed onset of divergence for one argument order relative to another may be the result of an ongoing competition between argument interpretations.

If listeners expect a certain argument linearization, we expect them to start moving the mouse towards the scene that shows the event as described by a prototypically ordered sentence. For a transitive event, we expect participants to start moving the mouse towards the scene that corresponds to an agent interpretation of the first argument. When case marking rules out that interpretation, we expect them to move the mouse towards the scene corresponding to a patient interpretation. This switching should surface in differences in the mouse trajectory between the two orderings, so mouse trajectories for OVS should be more curved than mouse trajectories for SVO. This way, we are also able to tell if there is a prototypical order for two animate objects after a ditransitive verb, that is, whether listeners expect a recipient to follow the verb, as in previous research, or a theme. If there is a prototypical order, this should be the order that is immediately selected. If not, listeners should show no differences in mouse trajectories between argument order conditions. As in Experiment 2a, we explore how the perceptual salience of the case cue influences the outcome variables. The data and analysis code for this experiment is available on OSF under https://osf.io/e9naj/.

5.1. Method

5.1.1. Participants

The same 39 participants as in Experiment 2a were also tested in Experiment 2b.

5.1.2. Materials and design

The same sentence and scenes as in Experiment 2a were used in Experiment 2b. In the mouse-tracking experiment, participants were assigned to a different list than the one they saw previously in the eye-tracking experiment.

5.1.3. Procedure

The mouse-tracking experiment was conducted after the eye-tracking experiment in one experimental session. Fig. 12 shows an example trial. To start a new trial, participants clicked on a start button at the bottom center of the 1920 × 1080 pixels screen, resulting in the two scenes of an item appearing in the left and right upper corner. As soon as participants moved the mouse upwards and crossed an invisible horizontal boundary above the start position, the audio file started playing. The trial ended as soon as participants had selected one of the scenes by clicking on it. Participants were instructed to move the mouse as fast as possible and click when feeling certain about their choice. Four practice trials at the beginning of the experiment familiarized the participants with the task. The experiment was built in OpenSesame (Mathôt et al., 2012, version 3.2.8, using the legacy backend) together with the mousetrap-OS plugin (Kieslich & Henninger, 2017, version 2.0.0). Cursor coordinates were recorded at a sampling rate of 100 Hz.[10]

5.2. Results

5.2.1. Accuracy

In Experiment 2b, all 39 participants had >80 % correct responses in the SVO condition. In Table 4, we summarize the accuracy scores for the transitive and ditransitive set.

A GLMM for the transitive set with argument order as fixed effect and by-subject and by-item varying intercepts as well as by-subject varying argument order slopes, showed an estimate that had a negative direction, but the difference between levels did not reach significance (Est. = −0.858, $SE = 0.771$, $z = -1.112$, $p = 0.266$).

A GLMM that included argument and NP order as well as their interaction for the ditransitive set showed a significant effect of argument order (Est. = 0.973, $SE = 0.339$, $z = 2.874$, $p = 0.004$), but not NP order (Est. = 0.169, $SE = 0.223$, $z = 0.758$, $p = 0.449$), and an interaction that was marginally significant (Est. = −0.826, $SE = 0.445$, $z = -1.858$, $p = 0.063$). The model included by-subject and by-item varying intercepts and by-subject varying argument order slopes. Pairwise comparisons revealed that the accuracy for the order DO-IO with a masculine NP as first postverbal argument (i.e., ACC marked via *den*) was lower than for IO-DO with a masculine NP (i.e., *dem*) as well as lower than for IO-DO with a feminine NP (i.e., *der*), yet not significantly lower than for DO-IO with a feminine NP (i.e., ACC marked via *die*); see Table A.3 in the appendix.

5.2.2. Response times

After inspection of the RT data, we excluded one trial from the

[10] Both the dynamic starting procedure and the response collection via click have been associated with more curved trajectories in comparison to static starting procedures and responses indicated by mouse movements to a response area (Kieslich et al., 2020). We did not manipulate the cursor speed to be slower, as we reasoned that such an additional manipulation could negatively affect how participants move the 'slower-than-usual' mouse. Instead, we used the default setting, that is, in Windows 10 medium speed, but switched off the mouse acceleration.

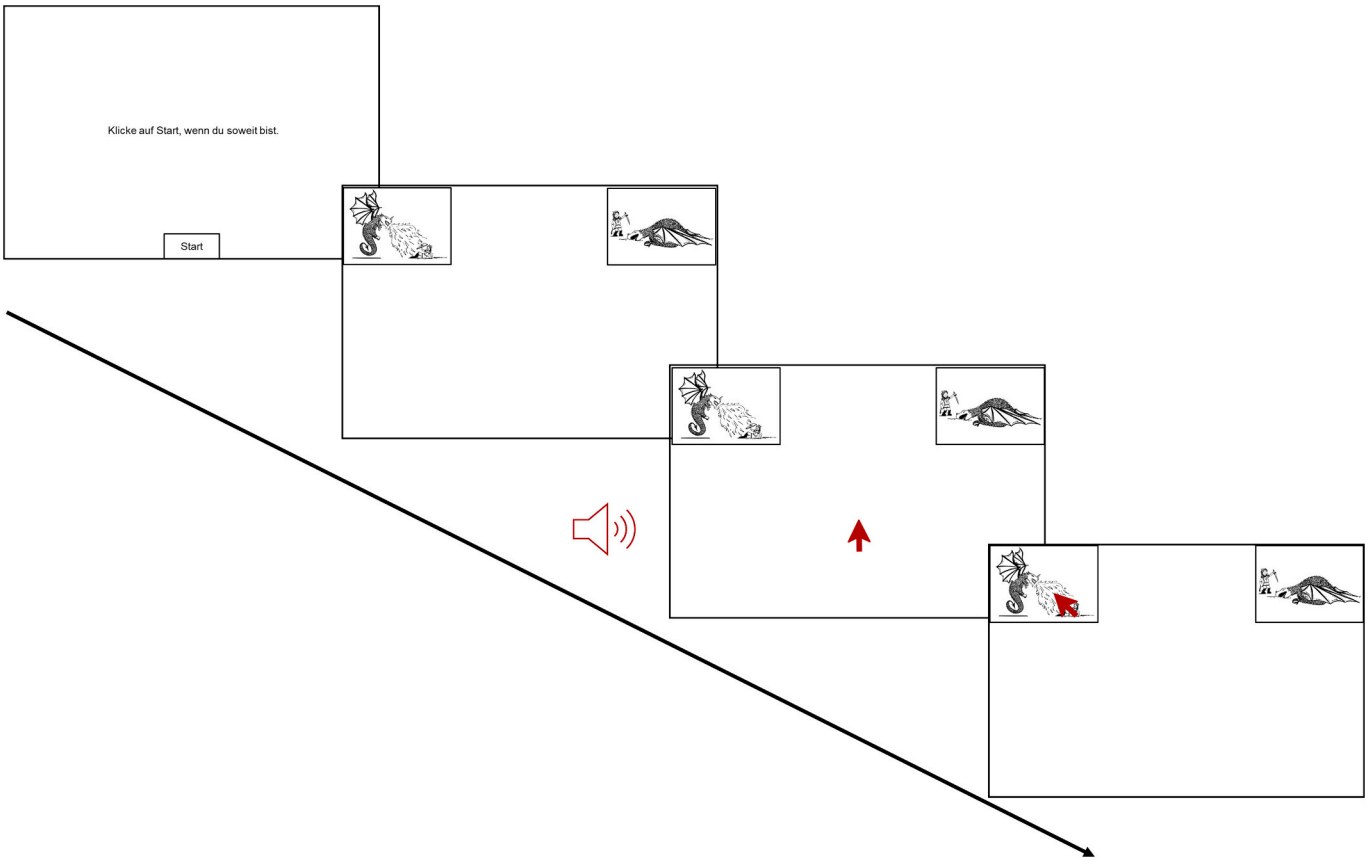


Fig. 12. Procedure in the mouse-tracking experiment.

Table 4
Number and percentages of correct and incorrect responses in Experiment 2b. Numbers and percentages are further provided separately for the simple (article) and the double marking (article + noun) of ACC in the OVS condition of the transitive set and for masculine and feminine articles in the ditransitive set.

	Correct	Incorrect
Transitive set		
SVO	457 (98 %)	11 (2 %)
OVS	440 (94 %)	28 (6 %)
article	219 (94 %)	15 (6 %)
article + noun	221 (94 %)	13 (6 %)
Ditransitive set		
IO-DO	**426 (91 %)**	42 (9 %)
masc. (*dem*)	215 (92 %)	19 (8 %)
fem. (*der*)	211 (90 %)	23 (10 %)
DO-IO	**392 (84 %)**	76 (16 %)
masc. (*den*)	188 (80 %)	46 (20 %)
fem. (*die*)	204 (87 %)	30 (13 %)

transitive set in the MT in which a participant responded in <200 msec after NP1 offset. For the transitive set, a LMM that included by-subject and by-item varying intercepts as well as by-subject varying slopes for argument order (SVO = reference level) revealed a significant difference between SVO and OVS (Est. = 0.095, $SE = 0.027$, $t = 3.473$, $p = 0.001$). The model predicted that participants need 2258 msec [1995, 2555] to respond to SVO and 2482 msec [2194, 2808] to respond to OVS. Like in Experiment 2a, an exploratory analysis with the OVS subset did not show a difference between the simple and double marking of ACC.

The output of a LMM on RTs for the ditransitive set that included by-subject and by-item varying intercepts as well as by-item varying slopes for argument order showed a marginally significant effect of argument order (Est. = −0.068, $SE = 0.036$, $t = -1.891$, $p = 0.071$), an effect of NP order (Est. = −0.12, $SE = 0.023$, $t = -5.257$, $p < 0.001$) and an interaction between both factors (Est. = 0.180, $SE = 0.046$, $t = 3.950$, $p < 0.001$). As indicated by pairwise comparisons (see Table A.4) and as visible in the right panel of Fig. 13 that shows the plotted model outputs, participants needed longer to respond to the order DO-IO with a masculine NP as first postverbal argument relative to all other conditions. Averaged over the levels of NP order, participants predicted RTs were 2779 msec [2452, 3150] for the order IO-DO and 2974 msec [2604, 3395] for DO-IO.

5.2.3. *Mouse-movement data*

The mouse movement data were pre-processed with the help of the *mousetrap R* package (Kieslich et al., 2019; Wulff et al., 2021). First, we aligned the starting position, so all trajectories start at (0,0). Note that this corresponds to a position approximately 350 pixels above the initial start position as we implemented a dynamic starting procedure, and the audio presentation was triggered by an upward movement. Second, we remapped all mouse trajectories to the left. Finally, we normalized the data, so that trajectories are represented by the same number of spatially equidistant data points or movement segments (see Wulff et al., 2021: 6) because we planned to analyze the shape of trajectories (irrespective of movement speed). Since we were interested in – ideally – continuous movements towards the scene, we excluded one participant who repeatedly released the mouse throughout the experiment as well as one participant who produced circles with the mouse and thus contributed many anomalous trajectories (10 trials for the transitive and 9 for the ditransitive set). We must emphasize that, for the remaining 37 participants, movements often were not continuous: Most participants moved the mouse upwards to listen to a sentence and paused before moving the mouse further towards a scene.

In Fig. 14, we plot the length-normalized trajectories per

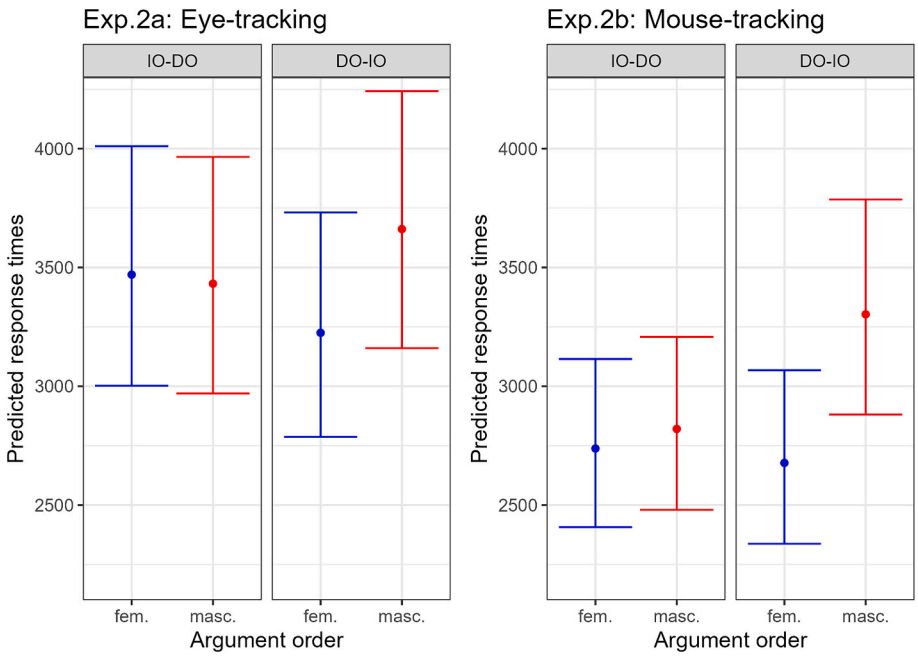


Fig. 13. Predicted response times and 95 % confidence intervals for the two argument orders in the ditransitive set depending on the gender of the first object noun in both Experiment 2a and 2b.

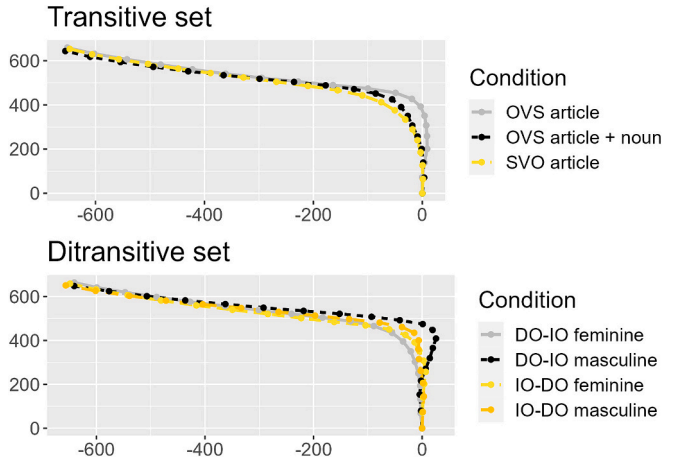


Fig. 14. Length-normalized and aggregated trajectories per argument order condition, also considering the marking of case. Trajectories start with the auditory onset of the sentence and last until response selection.

Table 5
Number and percentages of trajectory types per argument order and marking of case.

Argument order	Marking	Straight	Curved	cCoM	dCoM	dCoM2
Transitive set						
SVO	article	129 (30 %)	226 (52 %)	63 (14 %)	12 (3 %)	3 (1 %)
OVS	article	56 (27 %)	97 (47 %)	29 (14 %)	22 (11 %)	2 (1 %)
OVS	article + noun	65 (31 %)	96 (46 %)	28 (13 %)	17 (8 %)	4 (2 %)
Ditransitive set						
IO-DO	masc. (*dem*)	55 (27 %)	89 (44 %)	26 (13 %)	26 (13 %)	7 (3 %)
IO-DO	fem. (*der*)	56 (28 %)	91 (45 %)	35 (18 %)	17 (8 %)	2 (1 %)
DO-IO	masc. (*den*)	47 (26 %)	73 (41 %)	28 (15 %)	25 (14 %)	7 (4 %)
DO-IO	fem. (*die*)	46 (24 %)	102 (53 %)	30 (16 %)	12 (6 %)	2 (1 %)

experimental condition for the transitive (top) and ditransitive set (bottom). The OVS order is split into OVS trials with the double marking and the simple marking of case. For the ditransitive set, we plot the trajectories per argument order and gender of the first object noun, as — considering the previous findings — collapsing them might distort the results.

Next, we focus on the shape of trajectories. In Table 5, we report the number and percentages of trajectories per experimental condition that map onto a certain trajectory type. We use the set of prototypes that is available in the *mousetrap R* package as exemplified in Fig. 15. Descriptively, we observe more discrete change of mind (dCoM) trajectories for the order OVS than for the order SVO, especially when ACC is only marked on the article. Slight differences are also visible between the levels in the ditransitive set.

To statistically test for differences between levels, we use the ordering of prototypes from the least to maximal competition between response options (straight < curved < cCoM < dCoM < dCoM2) within an ordinal regression model. A model with random intercepts for subjects and items that takes the order OVS as the reference level shows that there is significantly less competition for the order SVO (-0.29, $SE = 0.14$, $z = -2.10$, $p = 0.04$). When taking as a reference level the order OVS with the simple marking of case (i.e., article only), the model output shows a marginally significant effect with less response competition for SVO (Est. $= -0.33$, $SE = 0.18$, $z = -1.87$, $p = 0.06$), but no difference for the same order with the double marking of case (Est. $= -0.07$, $SE = 0.21$, $z = -0.35$, $p = 0.72$). However, note that this analysis is based on a limited number of observations per level for the OVS order.

A model that takes the order DO-IO with a feminine NP as first postverbal argument as the reference level shows no significant differences between this level and the IO-DO orders. The difference between the order DO-IO with a feminine NP as first postverbal argument and the order DO-IO with a masculine NP as first postverbal argument is marginally significant (Est. $= 0.37$, $SE = 0.20$, $z = 1.84$, $p = 0.07$).

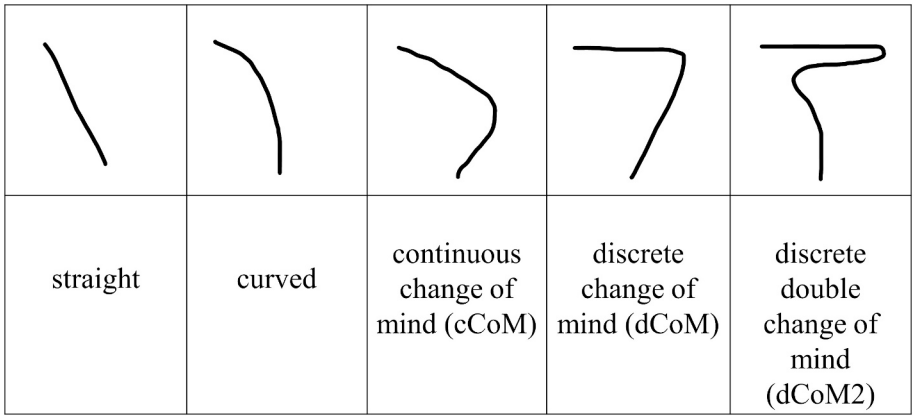


Fig. 15. Set of prototype trajectories provided by the *mousetrap R* package.

5.3. Discussion

Like in Experiment 2a, participants needed more time to select the target scene for the order OVS than for the order SVO. Again, participants scored numerically lower for OVS than for SVO, but the difference did not reach significance. The mouse-movement analyses showed a difference in the amount of conflict between SVO and OVS. We predicted that German speakers have a strong expectation that the first argument in a sentence is the subject/agent. Thus, if this expectation was violated, listeners would have to re-analyze an initial argument interpretation that was based on linear order. Our research question (1b) addressed the competition between an NP1-as-agent and NP1-as-patient interpretation and asked whether listeners maintain an agent-first interpretation in non-prototypically ordered transitive structures despite unambiguous case marking. Indeed, we found that participants were more likely to consider the competitor scene when the first argument was the object/patient, indicating that they did not immediately reject an NP1-as-agent interpretation but (temporarily) maintained an agent-first interpretation. Descriptively, we observed more discrete change of mind trajectories for the order OVS than for the order SVO. It might have been this competition that led to the variance in the eye-tracking experiment as regards the onset of a sustained target-over-competitor preference. The indication for an increased competition after the simple marking of ACC relative to the double marking is limited to a visual trend (Fig. 14).

In research question (2b), we addressed whether listeners have an expectation towards a certain argument order for ditransitive events with two animate object and, if yes, whether an initial interpretation is maintained. Our results from Experiment 1 indicated that the order IO-DO was preferred over DO-IO, and thus might be expected during processing. Our mouse-tracking findings provide no clear evidence that listeners maintain an initial object1-as-recipient interpretation in ditransitive structures when processing sentences that follow the order DO-IO. The mouse movement results point to an initial consideration of the competitor scene for the DO-IO order with a masculine noun as first postverbal argument (i.e., *den*), but not for the same order with a feminine noun (i.e., *die*). Like in Experiment 2a, we thus found that the processing of ditransitives was affected by both argument order and the perceptual salience of the case cue. Like in Experiment 2a, the order DO-IO with a masculine noun as first postverbal argument also 'stuck out' in terms of accuracy and response times.

6. General discussion

In a series of experiments, we manipulated the order of arguments in sentences with a transitive verb (subject/agent, object/patient) or ditransitive verb (DO/theme, IO/recipient) and examined how German speakers rate and process these sentences. We found that, in an acceptability judgment task (Experiment 1), German speakers preferred SVO over OVS and, though less strongly, the order IO-DO over DO-IO after a ditransitive verb. Next, we presented another group of German speakers with the same sentences together with two visual scenes (a target and a competitor scene) and recorded their eye movements (Experiment 2a) and then their mouse movements (Experiment 2b) while they selected the scene that was described by the sentence.

Our mouse-tracking findings align with previous research that shows a strong expectation towards argument order in transitive sentence structures (i.e., subject/agent-first) and adds evidence from a method that enabled us to capture the decision process. We found more response competition for OVS relative to SVO. In the eye-tracking experiment, we found that a sustained target-over-competitor preference emerged later for OVS than for SVO, which is in line with a lingering competition. In addition, in both the eye-tracking and the mouse-tracking experiment, participants' response times were longer for OVS than for SVO, reflecting processing difficulties. In Experiment 2a, participants were also more likely to select the target scene for the prototypical SVO order than for the OVS order. Nevertheless, even in the absence of semantic cues, ACC marking on NP1 was exploited prior to NP2 to identify the target scene, indicating that, overall, it is a reliable cue to argument interpretation.

We set out to explore whether the perceptual salience of the case cue influences the processing of transitive sentence structures. To this end, half of our OVS sentences marked ACC on the article only and half on both the article and noun. In both experiments, any differences between the simple and the double marking of ACC only showed up as trends (see also Binanzer et al., 2021). Hence, we have only limited evidence that the double marking of case facilitated comprehension. Future research might want to insert an adjective between article and noun (e.g., *den riesigen Drachen* – 'the giant dragon') and/or a prosodic cue to test whether an increasing number of cues reduces the (higher) uncertainty in the processing of OVS sentences. Henry et al. (2017), for example, found that multiple redundant cues can facilitate German listeners' processing of transitive sentences. The results from their growth curve analyses indicated that, if case marking was paired with a pitch accent on the NP1, listeners were faster and more accurate in fixating the target than for case marking alone, at least initially.

Previous research on ditransitive sentence structures suggested that expectations towards the order of recipient and theme argument are influenced by the animacy of arguments (e.g., Ellsiepen & Bader, 2018; Häussler & Bader, 2012), a factor we controlled for in this study. For our sentence stimuli with two animate objects, the order IO/recipient before DO/theme was overall preferred in an acceptability judgment task. The results from both the eye- and mouse-tracking experiment indicated that, if there was an expectation towards IO-DO, it was easily overridden by the case cue at the article of the first postverbal argument if perceptually salient (ACC marked by *die*). For the article *die*, case is not

only perceptually more salient than for *der*, *dem*, *den*, but can also be assigned from the second phoneme onwards. Yet, when we (exploratively) looked at the eye gaze pattern in trials in which the DO-IO order had a masculine noun as first object (ACC marked by *den*) and in which participants selected the competitor scene, the preference for the competitor scene (i.e., an object1-as-recipient interpretation) visibly developed in the critical time window for prediction (Fig. A.4). Lower accuracy rates and longer response times for the DO-IO order with a masculine noun as first object indicated an increased processing difficulty for this experimental condition. Moreover, participants tended to exhibit more response competition for the DO-IO order with a masculine noun as first object (*den … der*) compared to the same order with a feminine noun (*die … dem*). We observed no differences between the DO-IO and IO-DO order when the first object was feminine (*der … den* vs. *die … dem*), neither in the mouse- or eye-movement data nor in the accuracy or response time data. We thus speculate that IO-DO represents the default and that listeners rely on a default strategy — at least temporarily — if the case cue is low in perceptual salience. Otherwise, any expectation towards IO-DO may be easily overridden by a case cue. In other words, if in doubt, listeners may opt for the overall more frequent recipient > theme order, which in our acceptability judgment task was also the order that was rated as more acceptable. Future research might want to test the processing of ditransitive sentences with two animate objects using EEG to examine whether there is a processing difficulty associated with the DO-IO order that we were unable to detect with our methods.

Interestingly, we did find that listeners preferentially fixated on the target scene towards the NP2 in object-initial sentences, while there appeared to be no stable target preference after *den* following a ditransitive verb. If markedness negatively affected processing, then it should affect ACC/non-agent > NOM/agent more than ACC/theme > DAT/recipient. However, we realize that the processing of the two structures is difficult to compare as, for the ditransitive structures, listeners have to assign semantic roles to three arguments (and keep them in memory), while there are only two in transitive structures. Moreover, the visual scenes for the ditransitive set are more complex for the given reason. In our view, this does not render the findings less valuable, as it shows the complex interplay of different factors (e.g., perceptibility, number of arguments, position in the sentence) during real-time processing.

In the context of the present study, we referred to perceptual salience in terms of articles ending with a consonant that is acoustically hard to discriminate from another (e.g., *dem* vs. *den*) or in terms of case being marked once (article only) or twice (article and noun). Of course, more fine-grained analyses, considering, for example, co-articulation between the article and following noun are possible. Our findings revealed that the perceptibility of a case marking cue can be critical in real-time listening comprehension. Note that besides being perceptually more dissimilar, the articles *der* and *die* are also more frequent in the input than the articles *dem* and *den*. While this could be another reason for a better performance for the order DO-IO following *die* in contrast to *den*, the slight differences between the simple and double marking of case in the OVS sentences speak in favor of a role of perceptual salience for the predictive use of morphological case marking; see also Binanzer et al. (2021). Better performance in a similar picture selection task for experimental conditions in which the salience of the case morpheme was enhanced or in which the case morpheme was two-syllabic has also been shown for heritage speakers of Korean (Kim et al., 2018).

The finding that the perceptual salience of the case cue influences real-time processing, as shown most clearly for the ditransitive set of the current study, aligns with the assumption that the strength of prediction hinges on its utility (e.g., Kaan & Grüter, 2021): If the case cue is low in perceptual salience, it may receive less weight as there is a higher risk of prediction error, which then requires reanalysis. The findings are also consistent with noisy-channel processing (e.g., Traxler, 2014): When encountering a non-prototypical order, the listener may interpret the case cue as noise and takes an article such as *den* for another article. This might be especially the case for two objects following a verb when the described and depicted event is reversible and there is no additional semantic cue (e.g., Gibson et al., 2013).

This study is not without limitations. To address a reviewer's concern, in the present design with event scenes, strictly speaking we cannot tease apart the verification of a current semantic role assignment and semantic role prediction. Here, we took a sustained target-over-competitor preference prior to the critical argument as indication of a completed identification of all semantic roles, in line with previous research using a similar forced-choice design (e.g., Frenck-Mestre et al., 2019; Mitsugi, 2017). However, we acknowledge that an alternative interpretation is possible. Thus, some readers may prefer to interpret our eye-tracking findings as incremental but not as predictive processing (i.e., 'maximally' incremental processing). Moreover, in the present study, the eye-tracking experiment always preceded the mouse-tracking experiment. Ideally, the sequence in which the experiments were conducted would have been counterbalanced with some time (e.g., 1 to 2 weeks) between the experiments. Here, both experiments were administered in one experimental session as the data were collected during a research stay outside of the authors' home institution. We decided to start with the eye-tracking experiment because we reasoned that the familiarity with the pictures etc. from another experiment potentially affects the time course of prediction, while the effects may be less severe for the mouse-tracking experiment. That being said, we cannot rule out that our findings from the mouse-tracking experiment were influenced by prior exposure to the same pictures and similar sentences. In fact, we do see that overall response times were faster in the second experiment (see Fig. 13). However, we still obtained very similar effects in both experiments as regards accuracy and response times. Moreover, we observed more response competition for OVS than for SVO sentences. This may indicate that participants in our study did not adapt to this non-prototypical structure due to prior exposure. However, our study was not designed to test for adaptation effects and as already pointed out, lacks a comparison group that completed the mouse-tracking experiment prior to the eye-tracking experiment. Another limitation is the number of trials in the analyses of subsets. Due to the baseline effect, the divergence point analyses only include the subset of ditransitive structures in which the first object is a feminine NP.

7. Conclusion

The current paper was concerned with the interplay of linear order and morphological case as two cues to argument interpretation in transitive and ditransitive sentences in German. We found that, if there was an expectation towards an indirect object/recipient after a ditransitive verb, it was immediately overridden by the case cue at the article of the first postverbal argument when perceptually salient. For transitive sentences that started with the object/patient, listeners were found to be more attracted towards the competitor scene than for prototypically ordered SVO sentences. Listeners preferentially fixated the target scene in object-initial sentences most clearly towards the onset of the subject/agent, while the onset of prediction for SVO was well in advance of the onset of the object/patient. Based on our findings from eye-tracking and mouse-tracking, we conclude that adult native speakers of German expect a sentence to start with the subject/agent, but they have less strong expectations as regards the order of objects when encountering a ditransitive verb, at least when both objects are animate. As a result, depending on the perceptibility of the case cue, researchers may be more likely to find a predictive use of morphological case in one structure versus the other in the absence of additional semantic cues.

CRediT authorship contribution statement

Judith Schlenter: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Marit**

Westergaard: Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data and analysis code from the present study is available on the Open Science Framework under https://osf.io/e9naj/

Acknowledgments

We thank Nadine Kolb and Cecilia Puebla Antunes for their help with the creation of the auditory and visual materials for this study, as well as the University of Potsdam, especially Claudia Felser, for the support with data collection. Many thanks to all colleagues who helped us, e.g., by forwarding the links to the norming experiments and by providing feedback.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101024414.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2024.104241.

References

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

Bader, M., & Portele, Y. (2019). Giveness and the licensing of object-first order in German: The effect of referential form. In A. Gattnar, R. Hörnig, M. Störzer, & S. Featherston (Eds.), *Proceedings of linguistic evidence 2018: Experimental data drives linguistic theory* (pp. 208–228). University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/87132.

Binanzer, A., Cristante, V., & Bittner, A. (2021). Verarbeitung von deutschen W-Fragen mit starken und schwachen Maskulina durch ein- und mehrsprachige Kinder und Erwachsene – Evidenz für morphosyntaktische Schemata? In A. Binanzer, J. Gamper, & V. Wecker (Eds.), *Prototypen – Schemata – Konstruktionen* (pp. 375–408). De Gruyter. https://doi.org/10.1515/9783110710595-012.

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (version 6.2.09) [computer software]. http://www.praat.org/.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: Evidence from the online processing of flexible word order. *Cognition, 85*, B21–B30. https://doi.org/10.1016/S0010-0277(02)00076-8

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

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016). The argument dependency model. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 357–367). Elsevier/AP Academic Press.

Christensen, R. H. B. (2019). Ordinal — Regression models for ordinal data (version 2019.12-10). https://CRAN.R-project.org/package=ordinal.

Cristante, V. (2016). The processing of non-canonical sentences in children with German as a first or second language and German adults: Evidence from an eye-tracking study [doctoral dissertation, Westfälische Wilhelms-Universität Münster]. Publication server of the University Münster. https://nbn-resolving.de/urn:nbn:de:hbz:6-11269715348.

Dröge, A., Fleischer, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2016). Neural mechanisms of sentence comprehension based on predictive processes and decision certainty: Electrophysiological evidence from non-canonical linearizations in a flexible word order language. *Brain Research, 1633*, 149–166. https://doi.org/10.1016/j.brainres.2015.12.045

Ellsiepen, E., & Bader, M. (2018). Constraints on argument linearization in German. *Glossa: A Journal of General Linguistics, 3(1):6*, 1–36. https://doi.org/10.5334/gjgl.258

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology, 73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science, 27*(5), 315–323. https://doi.org/10.1177/0963721417746793

Frenck-Mestre, C., Kim, S. K., Choo, H., Ghio, A., Herschensohn, J., & Koh, S. (2019). Look and listen! The online processing of Korean case by native and non-native speakers. *Language, Cognition and Neuroscience, 34*(3), 385–404. https://doi.org/10.1080/23273798.2018.1549332

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition, 85*(3), B83–B92. https://doi.org/10.1016/S0010-0277(02)00126-9

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of cross-linguistic word-order variation. *Psychological Science, 24*(7), 1079–1088. https://doi.org/10.1177/0956797612463705

Häussler, J., & Bader, M. (2012). Grammar- versus frequency-driven syntactic ambiguity resolution: The case of double-object constructions. In M. Lamers, & P. de Swart (Eds.), *Studies in theoretical Psycholinguistics. Case, word order and prominence* (pp. 273–301). Netherlands: Springer. https://doi.org/10.1007/978-94-007-1463-2_12.

Hemforth, B. (1993). *Kognitives parsing: Repräsentation und Verarbeitung sprachlichen Wissens [cognitive parsing: Representation and processing of linguistic knowledge]*. Infix.

Hemforth, B., & Konieczny, L. (2000). *German sentence processing*. Kluwer Academic Publishers.

Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience, 32*(10), 1229–1249. https://doi.org/10.1080/23273798.2017.1327080

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods, 55*, 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan, & T. Grüter (Eds.), *Bilingual processing and acquisition: Prediction in second language processing and learning* (pp. 2–24). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.01kaa.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2*(4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 111–130). Routledge. https://doi.org/10.31234/osf.io/zuvqa.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods, 52*(1), 317–341. https://doi.org/10.3758/s13428-019-01228-y

Kim, K., O'Grady, W., & Schwartz, B. D. (2018). Case in heritage Korean. *Linguistic Approaches to Bilingualism, 8*(2), 252–282. https://doi.org/10.1075/lab.16001.kim

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Kröger, J. M., Münster, K., & Knoeferle, P. (2017). The influence of prosody and case marking on thematic role assignment in ambiguous action scenes: Adults versus children. In R. Granger, U. Hahn, & R. Sutton (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2463–2468). CogSci 2017.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leiner, D. J. (2019). SoSci Survey (Version 3.1.065). https://www.soscisurvey.de.

Lenth, R. (2023). Emmeans: Estimated marginal means, aka least-squares means (Version 1.8.5). https://CRAN.R-project.org/package=emmeans.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The costs of freedom: An ERP-study of non-canonical sentences. *Clinical Neurophysiology, 113*, 844–852. https://doi.org/10.1016/S1388-2457(02)00059-7

Minor, S., Mitrofanova, N., & Ramchand, G. (2022). Fine-grained time course of verb aspect. *PLoS One, 17*(2), Article e0264132. https://doi.org/10.1371/journal.pone.0264132

Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied PsychoLinguistics, 38*(4), 953–983. https://doi.org/10.1017/S0142716416000515

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition, 19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

Özge, D., Kornfilt, J., Maquate, K., Küntay, A. C., & Snedeker, J. (2022). German-speaking children use sentence-initial case marking for predictive language processing at age four. *Cognition, 221*, Article 104988. https://doi.org/10.1016/j.cognition.2021.104988

Pappert, S., Schließer, J., Janssen, D. P., & Pechmann, T. (2007). Corpus- and psycholinguistic investigations of linguistic constraints on German object order. In A. Späth (Ed.), *Interfaces and Interface conditions* (pp. 299–328). De Gruyter. https://doi.org/10.1515/9783110926002.299.

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review, 26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Røreng, A. (2011). Die deutsche Doppelobjektkonstruktion: Eine korpusbasierte Untersuchung zur relativen Abfolge nominaler Akkusativ- und Dativobjekte im geschriebenen Deutsch [Doctoral dissertation, Universitetet i Tromsø]. Munin open research archive. https://hdl.handle.net/10037/3377.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language, 38*, 150–176.

Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology, 7*, 95. https://doi.org/10.3389/fpsyg.2016.00095

Sauppe, S., Næss, Å., Roversi, G., Meyer, M., Bornkessel-Schlesewsky, I., & Bickel, B. (2023). An agent-first preference in a patient-first language during sentence comprehension. *Cognitive Science, 47*(9), Article e13340. https://doi.org/10.1111/cogs.13340

Scherger, A.-L., Kizilirmak, J. M., & Folta-Schoofs, K. (2022). Ditransitive structures in child language acquisition: An investigation of production and comprehension in children aged five to seven. *Journal of Child Language, 1–18*. https://doi.org/10.1017/S0305000922000174

Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science, 15*(6), 762–774. https://doi.org/10.1111/j.1467-7687.2012.01167.x

Schlenter, J., & Felser, C. (2021). Second language prediction ability across different linguistic domains: Evidence from German. In E. Kaan, & T. Grüter (Eds.), *Bilingual processing and acquisition: Prediction in second language processing and learning* (pp. 48–68). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.03sch.

Schlenter, J., & Penke, M. (2022). (How) Visual properties affect the perception and description of transitive events. *Glossa Psycholinguistics, 1*(1), 1–29. https://doi.org/10.5070/G601193

Schlesewsky, M., Bornkessel, I., & Frisch, S. (2003). The neurophysiological basis of word order variations in German. *Brain and Language, 86*(1), 116–128. https://doi.org/10.1016/S0093-934X(02)00540-0

Siewirska, A. (1993). On the interplay of factors in the determination of word order. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: An international handbook of contemporary research* (pp. 826–846). De Gruyter.

Stone, K., Lago, S., & Schad, D. J. (2020). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition, 24*, 833–841. https://doi.org/10.1017/S1366728920000607

Strotseva-Feinschmidt, A., Cunitz, K., Friederici, A. D., & Gunter, T. C. (2015). Auditory discrimination between function words in children and adults: A mismatch negativity study. *Frontiers in Psychology, 6*, 1930. https://doi.org/10.3389/fpsyg.2015.01930

Suitner, C., Maass, A., Navarrete, E., Formanowicz, M., Bratanova, B., Cervone, C., … Carrier, A. (2021). Spatial agency bias and word order flexibility: A comparison of 14 European languages. *Applied PsychoLinguistics, 42*(3), 657–671. https://doi.org/10.1017/S0142716420000831

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences, 18*(11), 605–611. https://doi.org/10.1016/j.tics.2014.08.001

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*(2), B63–B72. https://doi.org/10.1016/j.cognition.2005.07.001

Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2021). Movement tracking of cognitive processes: A tutorial using mousetrap. *PsyArXiv*. https://doi.org/10.31234/osf.io/v685r