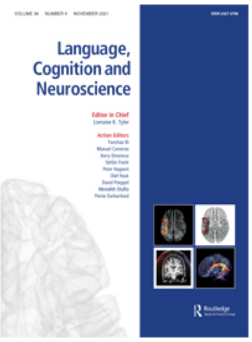

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/plcp21

The interaction of grammatically distinct agreement dependencies in predictive processing

Kate Stone, João Veríssimo, Daniel J. Schad, Elise Oltrogge, Shravan Vasishth & Sol Lago

To cite this article: Kate Stone, João Veríssimo, Daniel J. Schad, Elise Oltrogge, Shravan Vasishth & Sol Lago (2021) The interaction of grammatically distinct agreement dependencies in predictive processing, Language, Cognition and Neuroscience, 36:9, 1159-1179, DOI: 10.1080/23273798.2021.1921816

To link to this article: https://doi.org/10.1080/23273798.2021.1921816

LANGUAGE, COGNITION AND NEUROSCIENCE
2021, VOL. 36, NO. 9, 1159–1179
https://doi.org/10.1080/23273798.2021.1921816

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

The interaction of grammatically distinct agreement dependencies in predictive processing

Kate Stone [a], João Veríssimo [a,b], Daniel J. Schad [c], Elise Oltrogge[a], Shravan Vasishth [a] and Sol Lago [d]

[a]Department of Linguistics, University of Potsdam, Potsdam, Germany; [b]School of Arts and Humanities, University of Lisbon, Lisbon, Portugal; [c]Psychology Department, Health and Medical University Potsdam, Potsdam, Germany; [d]Institute for Romance Languages and Literatures, Goethe University Frankfurt, Frankfurt, Germany

ABSTRACT
Previous research has found that comprehenders sometimes predict information that is grammatically unlicensed by sentence constraints. An open question is why such grammatically unlicensed predictions occur. We examined the possibility that unlicensed predictions arise in situations of information conflict, for instance when comprehenders try to predict upcoming words while simultaneously building dependencies with previously encountered elements in memory. German possessive pronouns are a good testing ground for this hypothesis because they encode two grammatically distinct agreement dependencies: a retrospective one between the possessive and its previously mentioned referent, and a prospective one between the possessive and its following nominal head. In two visual world eye-tracking experiments, we estimated the onset of predictive effects in participants' fixations. The results showed that the retrospective dependency affected resolution of the prospective dependency by shifting the onset of predictive effects. We attribute this effect to an interaction between predictive and memory retrieval processes.

ARTICLE HISTORY
Received 18 December 2020
Accepted 19 April 2021

KEYWORDS
Sentence processing; visual world eye-tracking; prediction; gender agreement; German

Introduction

Grammatical relationships shape the morphological form of sentence constituents, providing speakers with powerful cues to track these relationships during comprehension. For example, the form of German articles and adjectives is governed by gender and number agreement with a nominal head, as in "*ein blauer Knopf*" (a$_{\text{.SG.MASC}}$ blue$_{\text{.SG.MASC}}$ button$_{\text{.MASC}}$). Comprehenders compute these relationships online, as demonstrated by the fact that agreement violations are quickly noticed within and across phrases (Barber & Carreiras, 2005; Foote, 2011; Hagoort & Brown, 1999; Keating, 2009; Molinaro et al., 2008; Nevins et al., 2007). Agreement relationships also appear to inform comprehenders' predictions about upcoming sentence constituents (Brouwer et al., 2017; Dahan et al., 2000; Dussias et al., 2013; Grüter et al., 2020; Hopp, 2013, 2012; Hopp & Lemmerth, 2018; Lew-Williams & Fernald, 2010). However, it is unclear whether agreement-based predictions are always fully grammatically constrained or whether they can be affected by other grammatically illicit types of information, as has been observed for other linguistic phenomena (Kamide & Kukona, 2018; Kukona et al., 2011, 2014; Rommers et al., 2013, 2015). The fallibility of agreement constraints is a useful issue to assess as it may yield insight into the mechanisms underlying predictive processing. With this goal in mind, we present two studies examining the predictive use of agreement and the circumstances in which comprehenders struggle to deploy agreement relations in a grammatically faithful manner.

Several studies have shown that comprehenders use morphosyntactic relationships to predict upcoming nouns. Much of this evidence comes from the visual world paradigm, in which participants typically hear sentences while viewing objects on a screen. For example, Hopp and Lemmerth (2018) presented native and non-native German speakers with auditory sentences that contained gender-marked articles (or adjectives) and colour adjectives, together with displays that showed three colour-matching objects with the same or different gender. In the "different" trials, the gender marking of the article (1a) or adjective (1b) was a useful cue to differentiate the target object from the two other colour-matching objects. By contrast, in the "same"

CONTACT Kate Stone stone@uni-potsdam.de Department of Linguistics, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam.
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2021.1921816.

trials, gender was not a useful cue for target identification and the target object was only identified by the noun:

(1a) Wo ist **der/die/das** gelbe … ?

Where is the.$_{MASC/FEM/NEUT}$ yellow … .?

(1b) Wo ist ein **kleiner/kleines** gelber/s … ?

Where is a small.$_{MASC/NEUT}$ yellow.$_{MASC/NEUT}$ … .?

The results showed that native speakers were faster to look predictively at the target object in different than same gender trials, suggesting that they used agreement to constrain their noun expectations (see also Brouwer et al., 2017; Dahan et al., 2000; Dussias et al., 2013; Hopp, 2013; Lew-Williams & Fernald, 2010). Similar predictive effects have been reported for number agreement with German articles (Hopp, 2012) and with Mandarin Chinese classifiers, which provide cues about noun class membership (Grüter et al., 2020). In English, where articles and adjectives do not show number agreement, constructions such as "*There **is***" and "*There **are***" have been found to trigger preferential looks toward number-congruent objects (Kouider et al., 2006; Lukyanenko & Fisher, 2016; but see Riordan et al., 2015).

Relatedly, studies on case marking have shown that comprehenders of German, Turkish, English, Korean, Hindi, and Japanese can use case-marked noun phrases to anticipate constituents that are yet to be mentioned (Frenck-Mestre et al., 2019; Henry et al., 2017; Hopp, 2015; Husain et al., 2014; Kamide, Altmann, et al., 2003; Kamide, Scheepers, et al., 2003; Mitsugi & MacWhinney, 2016; Özge et al., 2016; Zhang & Knoeferle, 2012). Together, this evidence highlights a key role of morphosyntactic relationships in sentence comprehension: They allow speakers to proactively identify possible heads of noun phrases (gender and number agreement) and to assign thematic roles to these phrases (case marking), thus enabling incremental interpretation and enhancing comprehension speed and robustness to noise.

However, while previous evidence establishes that agreement-based relationships *inform* predictions, it does not address whether predictions are *restricted solely* to agreement-licit targets. On the one hand, agreement relations may strongly constrain prediction formation, such that comprehenders only predict morphosyntactically licit words (henceforth "grammatically constrained predictions"). Alternatively, agreement relations might play a more fallible role, such that they can be outweighed by other types of information. This latter possibility is consistent with accounts of prediction as a mechanism that creates representations at multiple levels (Kukona et al., 2011; Kuperberg & Jaeger, 2016) or that involves different mechanisms, each fine-tuned to different information sources (Huettig, 2015). Importantly, if agreement constraints are fallible, then it is important to understand the conditions under which this occurs. Below, we summarise evidence in support of each of these two possible roles of morphosyntactic relationships in predictive processing. We then describe our experiments, which tested for the existence of grammatically unlicensed predictions.

Grammatically constrained predictions

The possibility that morphosyntactic information fully constrains prediction formation is consistent with findings in two other predictive dependencies: filler-gap dependencies and cataphora. These two dependencies are assumed to involve predictive processing in reading studies (for discussion, see Phillips et al., 2011). With regard to filler-gap dependencies, the proposal is that encountering a displaced filler phrase such as "*the book*" in "*We like the book that* … " triggers an active search for a gap representing the original syntactic position of the filler. This search is proactive, such that a reading disruption is observed when the earliest available gap site (underlined) leads to an implausible sentence: "*We like the book/*city* [*that the author wrote* __]". Crucially, the disruption does not occur within domains that are grammatically unlicensed to host a gap, such as inside an embedded relative clause: "*We like the book/*city* [*that the author* [*who wrote* __]] … " (Traxler & Pickering, 1996). These domains are called "islands" (Ross, 1967), and several studies have shown that island constraints restrict predictive dependencies (Bourdages, 1992; McElree & Griffith, 1998; Ness & Meltzer-Asscher, 2019; Omaki et al., 2015; Omaki & Schulz, 2011; Phillips, 2006; Pickering et al., 1994; Wagers & Phillips, 2009).

Similarly, encountering a cataphoric pronoun, which precedes its antecedent, is also assumed to trigger a predictive search for the antecedent. The evidence that this search is syntactically constrained comes from gender mismatch effects. In sentences like "*When **he** was at the party, the boy/*girl* … ", the noun that gender-mismatches the pronoun, "*girl*", elicits reading disruptions, confirming that the parser attempts to resolve coreference as early as possible (van Gompel & Liversedge, 2003). Crucially, the mismatch effect disappears when the antecedent is in a structural position in which coreference would violate binding constraints, e.g. in an embedded "while" clause: "*Because last semester **he** was taking classes full-time* [*while Russell/Kathryn was* *working*] … " (Aoshima et al., 2009; Drummer & Felser, 2018; Kazanina et al., 2007; Kazanina & Phillips, 2010). This suggests that only noun phrases in grammatically constrained positions are considered during the predictive antecedent search. Thus, research on both cataphora and filler-gap dependencies suggests that predictions are grammatically constrained. If these findings extend to agreement relationships, we would expect predictions to be restricted to morphosyntactically licensed words.

Predictions in the presence of information conflict

The literature summarised above suggests that predictions may be fully constrained by grammatical information. An alternative option is that morphosyntactic information is weighed against other types of information. Crucially, if morphosyntax has a weighted influence, agreement relations should inform predictions but may not fully constrain them, making them fallible; for example, when they conflict with other information sources, such as lexico-semantic or non-linguistic information.

Situations of information conflict can indeed elicit predictions that are locally coherent but inconsistent with sentence-level (i.e. combinatorial) relationships. One such case concerns the interaction between visual and sentence-level information (Knoeferle & Crocker, 2006, 2007; Kukona et al., 2011, 2014; Rommers et al., 2013, 2015). For example, Rommers et al. (2013) conducted a visual world experiment in Dutch that measured participants' eye movements while they were listening to preambles that were strongly predictive of a target word, such as "*moon*" in the sentence "*In 1969 Neil Armstrong was the first man to set foot on the moon*". Participants were presented with a display showing either the target object (the moon), an object with a similar shape (a tomato) or a shape-unrelated object (a box of rice). Before the target noun was heard, participants fixated both the target and shape competitor more often than the unrelated object, suggesting that they considered the shape competitor as a possible sentence completion despite its incongruency with the combinatorial meaning of the sentence. This effect was attributed to a priming mechanism, such that the preactivation of the target word spread to semantically related words—including those with similar shape attributes—causing comprehenders to consider semantically incongruent nouns.

A similar effect was observed by Kukona et al. (2014), who found that after hearing a preamble like "*The boy eats the white* … ", participants' fixations increased not only to images of edible objects (e.g. a white cake) but also to the image of a white car, which was not semantically congruent with the context. Crucially, the white car received more looks than a brown car, which again suggests that the processing of visual cues, which were consistent with the meaning of the adjective, influenced predictions even though these were ruled out by sentence-level constraints.

Further evidence comes from a study that addressed the interaction between lexical and sentence-level relationships. Kukona et al. (2011) presented English-speaking participants with sentences like "*Toby arrested the* … " and examined their fixations on images of verb-related patients and agents (e.g. "*crook*" and "*policeman*"). Surprisingly, after hearing the verb, listeners showed a similar proportion of anticipatory fixations to both the patient and the agent, even though the latter was sentence-incongruent because its role had already been filled (by "*Toby*"). Under the assumption that looks to the agent were due to its lexical association with the verb ("*arrest–policeman*"), this finding demonstrates that lexical associations can prime globally incongruent words, triggering predictions that violate sentence-level constraints.

Furthermore, the study by Kukona et al. (2011) suggested that increased time may help listeners resolve information conflict and prioritise sentence-level constraints. A second experiment, which used passive sentences like "*Toby was arrested by the* … " showed a larger difference in looks between the agent and patient images, consistent with a stronger influence of sentence-level constraints. However, this finding does not unambiguously show an effect of time, because the passive sentences in the study also contained additional syntactic cues that could influence participants' interpretation, e.g. the function words "*was*" and "*by*". However, a later study, which used active sentences similar to those of Kukona and colleagues but gave participants more time by using a slower speech rate, no longer found a fixation bias towards the sentence-illicit agent image, thus supporting the hypothesis that combinatorial relationships need time to fully constrain predictions (Gambi et al., 2016). Additional evidence of the role of time in prediction formation can be found in visual world (Huettig & Guerra, 2019) and electrophysiological studies (Chow et al., 2016, 2018; Dambacher et al., 2012; Ito et al., 2016; Momma et al., 2015; Wlotko & Federmeier, 2015).

Interestingly, the evidence above concerns mostly lexico-semantic and thematic constraints. Much less is known about conflicts involving syntactic constraints, although one recent visual world study by Kamide and Kukona (2018) supports the possibility that these may also be fallible. In this study, participants viewed scenes which contained several objects while hearing sentences with either a noun phrase (NP) or a verb phrase (VP) modifier: e.g. "*The girl who likes the man (from London/very much) will ride the carousel*" Among the objects on the screen, only one was congruent with the sentence-level meaning (carousel). However, in the NP condition, the motorbike was locally coherent within the embedded noun phrase "*the man from London will ride …*", but globally incongruent with the syntactic representation of the sentence. The results showed evidence of syntactically-unlicensed predictions: the motorbike elicited as many predictive looks as the carousel in the NP condition, but fewer looks in the VP condition.

Taken together, the evidence above suggests that the interaction between different types of information can give rise to predictions that are temporarily inconsistent with sentence-level semantic and syntactic relationships. Grammatically unlicensed predictions may thus be the result of different types of information being differentially weighted during prediction formation, or of errors that occur when the predictive process interfaces with processes supporting access to different information types (e.g. lexical retrieval, priming, visual encoding, etc.). It is also unclear whether cases of syntactically unlicensed predictions, such as those observed by Kamide and Kukona (2018), extend to morphosyntactic relationships like agreement, where an ungrammatical agreement configuration is neither locally nor globally licensed. Furthermore, while previous studies have proposed that sentence-level thematic relationships require time to override conflict and fully constrain predictions (Chow et al., 2018; Kukona et al., 2011), it is unknown whether time also affects the predictive use of agreement relationships, which are typically rapidly computed by native speakers during online processing (e.g. Brouwer et al., 2017; Dahan et al., 2000; Dussias et al., 2013; Hopp & Lemmerth, 2018; Lew-Williams & Fernald, 2010). Our study addresses these questions by examining the online application of agreement constraints to predictive processing.

The present study

To establish the use of agreement relationships in predictive processing we examined the comprehension of possessive pronouns. Possessive pronouns are useful in the study of agreement because they establish morphosyntactic and referential dependencies with other sentence constituents. Across languages, these relationships can be prospective and/or retrospective. For instance, the gender of an English possessive pronoun marks a retrospective or backward-looking relationship: "*his*" agrees in gender with a preceding masculine antecedent or "possessor", while "*her*" agrees with a preceding feminine possessor.

In contrast with English, German possessives like "*sein*" and "*ihr*" ('his' and 'her') establish gender agreement not only retrospectively, but also prospectively, because a German possessive needs to agree in gender with its syntactic head: the so-called "possessee" noun. In (2) the possessive stem ("*sein-*" vs. "*ihr-*") marks possessor agreement whereas the possessive suffix ("*-en*" vs. "*-e*") marks possessee agreement (note that German nouns have grammatical gender):

(2a) **Martin** nimmt **sein**en Knopf.

Martin takes his.$_{MASC}$ button.$_{MASC}$

(2b) **Sarah** nimmt **ihr**e Flasche**.**

Sarah takes her.$_{FEM}$ bottle.$_{FEM}$

These bidirectional prospective and retrospective relationships mean that upon encountering a German possessive, comprehenders may implement at least two cognitive processes: identify a suitable antecedent (based on the possessive stem) and predict an upcoming noun (based on the suffix). Crucially, only the gender of the possessive stem—possessor agreement—is grammatically licit for identifying the antecedent, and only its suffix—possessee agreement—is grammatically licit for predicting an upcoming noun.

However, possessor and possessee gender features may differ, such as in **Martin** nimmt **sein**e.$_{FEM}$ Flasche.$_{FEM}$ ('*Martin takes his.$_{FEM}$ bottle.$_{FEM}$*'). Thus, possessive pronouns can give rise to situations of conflict between the agreement features of the prospective and retrospective dependencies. If both dependencies are resolved independently and in a grammatically faithful manner, then the gender of the possessor should not interfere with the computation of possessee agreement. However, if the prospective computation is not fully grammatically constrained, then the gender of the possessor may affect its resolution, suggesting an interaction between the two grammatically separate dependencies. In the General Discussion, we attribute the resolution of the retrospective and prospective dependencies to two different cognitive mechanisms, memory retrieval and prediction, and we discuss how our results enable inferences about the interaction of these processes in a cue-based retrieval framework (Lewis & Vasishth, 2005).

To examine the possibility that the grammatically illicit possessor gender feature interferes with agreement-based predictions, we conducted two visual world experiments with German native speakers. Participants were instructed to click on objects on a screen, and we manipulated whether the possessee and possessor in the instruction (referring to the object and its owner respectively) matched or mismatched in gender. In the *match condition*, the possessor and possessee matched in gender, while in the *mismatch condition* they did not (note that both versions are grammatical in German):

(3) Match condition (masculine target noun)

Klicke auf **sein**en blauen Knopf!

Click on his.$_{MASC}$ blue.$_{MASC}$ button.$_{MASC}$

Mismatch condition (masculine target noun)

Klicke auf **ihr**en blauen Knopf!

Click on her.$_{MASC}$ blue.$_{MASC}$ button.$_{MASC}$

With regard to the retrospective relationship, we expected that upon hearing the possessive, German speakers would use the gender of its stem to identify a gender-matching referent. This assumption was based on previous work showing that gender cues are rapidly used to resolve coreference relationships involving pronouns and reflexives (Arnold et al., 2000; Badecker & Straub, 2002; Carreiras et al., 1996; Garnham et al., 1995; Jäger et al., 2015; Kennison, 2003; Laurinavichyute et al., 2017; Osterhout & Mobley, 1995; Runner et al., 2006; Sturt, 2003; van Berkum et al., 2004).

With regard to the prospective relationship, we expected that participants would use the possessive suffix to predict the target word, consistent with previous findings on gender-marked articles and adjectives. If so, they should predictively look at the object corresponding to the possessee noun prior to its mention in the auditory instruction. Our research question was whether predictions would be affected by the processing of the retrospective dependency. If predictions are fully morphosyntactically constrained, then the gender of the possessive stem should not affect the predictive process, because possessor gender is syntactically irrelevant for possessee agreement. In this case, object predictions should not differ between the match and mismatch conditions.

Alternatively, if predictions are not fully morphosyntactically constrained, then the gender of the possessive stem might affect the predictive process. Specifically, the need to distinguish between the two different gender features relevant for possessor and possessee agreement may create a conflict in the mismatch condition. Comprehenders may struggle to dissociate the two gender features, resulting in delayed or absent target predictions. In contrast, in the match condition the two gender features align and thus conflict should not arise when predicting the upcoming possessee. We evaluated these predictions in two experiments, the first with complex 4-object visual displays (Experiment 1) and the second with simpler 2-object displays (Experiment 2).

Experiment 1

In Experiment 1 participants heard auditory instructions like (3) while looking at displays with four objects. The auditory instruction contained a possessive pronoun, a colour adjective, and a noun indicating the target object to be clicked on (Figure 1). The target object matched the gender and colour cues in the instruction. The other objects matched only the colour cue (*colour competitor*), only the gender cue (*gender competitor*) or neither (*distractor*). The possessive indicated the gender of the target object, while the adjective indicated both its gender and colour. While our research question concerned the processing of the possessive, the adjective gave listeners more time to process the gender cues in the possessive. Given the differently coloured objects on screen, the target object only became fully predictable at the adjective and thus the adjective was the critical window for analysis. The critical comparison was between the target and colour competitor (henceforth, *competitor*): As both objects matched the colour of the adjective, but only the target had the appropriate gender, any target-over-competitor advantage should reflect the predictive use of gender information.

In the match condition, the target possessee noun in the instruction matched in gender with the possessor, whereas in the mismatch condition, they did not. Thus, if conflict between gender features elicits syntactically unlicensed predictions, we expected target predictions during the adjective time window to be less likely or more delayed in the mismatch vs. match condition. To assess this, we evaluated both the likelihood of the predictive effect in the match vs. mismatch condition (time-window analysis) as well as whether the onset of the predictive effect was delayed in the mismatch condition (onset analysis).

The use of onset analyses is a methodological contribution of this study and allowed us to examine not only *whether* participants were more likely to look at one object than the other, but also *when* this preference emerged, enabling inferences about the timecourse of predictions. As described above, some types of grammatical information may require more time to constrain predictions (Chow et al., 2018; Kukona et al., 2011), but

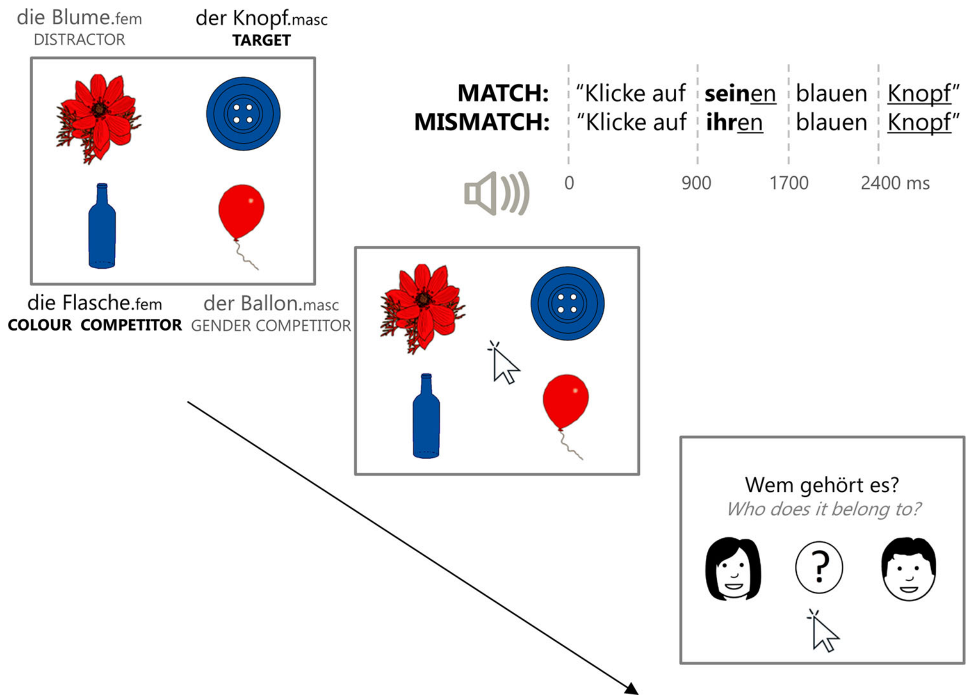

Figure 1. Sample experimental trial in Experiment 1. Participants saw a 1000 ms preview of the four objects before the onset of the auditory instruction. Only the target object matched both the gender and colour cues in the instruction ('the button.$_{MASC}$'). The colour competitor matched in colour but not in gender ('the bottle.$_{FEM}$'), the gender competitor matched in gender but not in colour ('the balloon.$_{MASC}$'), and the distractor matched neither in colour nor gender ('the flower.$_{FEM}$'). From the onset of the instruction, participants had 4500 ms to click on the target object. After each response, participants were asked to recall the owner of the target object and click on its image, with a response deadline of 1500 ms (possessor identification task). The question mark corresponded to an "*I don't know*" response.

it is unknown whether the same is true of agreement-based predictions. To directly measure the timecourse of predictions, we adopted a bootstrapping method recently proposed by Stone et al. (2020). An important advantage of this method is that it allows estimation not only of an effect onset but also of its temporal uncertainty, quantified as a confidence interval. The method is described in more detail below.

Methods

Participants

Seventy-four native German speakers participated in Experiment 1. The sample size was motivated by recent concerns about small sample sizes in psycholinguistics (Gelman & Carlin, 2014; Vasishth et al., 2018); thus, we recruited the maximum number of participants feasible during the university semester. Data from two participants were excluded due to tracker loss, leaving 72 participants for the analysis (mean age 25 years, age range 18–40 years, 47 female, 67 right-handed). Participants reported exposure solely to German until age six and no history of neurological, reading or writing disabilities. This and following experiments were approved by the University of Potsdam ethics committee.

Materials

Materials consisted of 24 experimental items and 24 fillers. In the experimental items, the possessive and the adjective agreed in gender, number and case with the noun. Auditory instructions were spliced such that the onsets of the possessive pronoun, adjective, and noun were identical across trials. The critical time window extended from the onset of the adjective to 200 ms after the onset of the noun, to account for the time taken to program and launch eye movements (Hallett, 1986; Salverda et al., 2014).

To ensure that a target-over-competitor advantage reflected the predictive use of gender information, rather than differences in visual saliency between objects, each object appeared in each of the four roles across trials. For example, the colour competitor in Example (3), "*the bottle*", was the target object in another trial, shown in Example (4), and the distractor and gender competitor in other trials. Across the experiment, half of the target nouns were masculine and half were feminine. For half of the trials, the possessive "*sein*" was used in the match condition and "*ihr*" in the mismatch condition. For the other half, "*ihr*" was used in the match condition and "*sein*" in the mismatch condition. This was done to avoid a potential effect of referential ambiguity, as the German possessive "*ihr*" can in principle refer to either a singular or plural third person antecedent. Participants were told that the pronouns always referred to either a male or female character and the practice session was used to monitor that they resolved coreference as intended. The position and colour of the objects were counterbalanced across trials.

(4) Match condition (feminine target noun)

Klicke auf **ihr**e gelbe Flasche!

Click on her.$_{FEM}$ yellow.$_{FEM}$ bottle.$_{FEM}$

Mismatch condition (feminine target noun)

Klicke auf **sein**e gelbe Flasche!

Click on his.$_{FEM}$ yellow.$_{FEM}$ bottle.$_{FEM}$

Images for the objects were taken from freely available image databases: 98% from MultiPic (Duñabeitia et al., 2018) and 2% from a previous study by Hopp and Lemmerth (2018). We used images for which a high percentage of participants had provided the target name in norming studies (mean = 90%, 95% confidence interval = [71, 100]; Duñabeitia et al., 2018). Each image measured 300 × 300 pixels and had one of four colours: red, green, blue, or yellow. Each object appeared once as a target and once as each non-target object (colour competitor, gender competitor, distractor).

The auditory instructions for the fillers had the same structure as the experimental items except that one third contained determiners instead of pronouns, and one third contained target nouns of neuter gender. For the neuter fillers, two objects on screen corresponded to neuter nouns, but only one of the objects matched the gender and colour cues of the auditory instruction. For the remaining fillers, all objects had the same gender but a different colour. These fillers served to ensure that participants were able to use colour features to make predictions, in the absence of a gender manipulation. The gender, colour and position of the filler objects was counterbalanced.

Experimental and filler trials were distributed into three lists, presented in a Latin square design such that each participant saw 8 items per condition.[1] Presentation order was randomised on a by-participant basis. A full list of materials together with their identifiability ratings and gender distribution is available online at https://osf.io/mbtcd/, alongside all data and analysis code for Experiments 1 and 2.

Procedure

Participants sat in front of a 22-inch computer monitor with a screen resolution of 1680 × 1050 pixels. The eye-to-screen distance was approximately 66 cm. The movements of the right eye were recorded at a sampling rate of 1000 Hz by an EyeLink1000 (SR Research). An adjustable chin and forehead rest served to avoid head movements during the experiment. At the beginning of the experiment, participants were introduced to two characters, *Martin* and *Sarah*, whose faces were displayed on the screen. Participants were told that their task was to help Martin and Sarah tidy up their messy house by finding their belongings before their parents arrived. They were told that they would see four images and hear an auditory instruction, and that they should click on the object mentioned in the instruction as quickly and accurately as possible.

Each trial was followed by a possessor identification task, which was used to ensure that participants resolved the coreferential relationship between the possessive and its referent (Martin or Sarah). Participants saw a question asking who the object in the auditory instruction belonged to and they were instructed to click as fast as possible on Martin's or Sarah's image, or a question mark if they were unsure about their answer. They were told that they only had 15 minutes to complete the task. The experiment started with a 4-trial practice session. The testing session lasted approximately 40 minutes.

Analysis

Raw data were initially preprocessed in DataViewer (SR Research). No cleaning of the raw data was necessary. Data were downsampled to 250 Hz and exported to R (R Core Team, 2019). The three analyses performed on the preprocessed data are described below.

Critical time-window. The goal of the time-window analysis was to establish whether the target object was predicted during the adjective time window and whether the likelihood of the predictive effect differed between the match and mismatch conditions. We fit a Bayesian generalised linear mixed-effects model with maximal random effects structure (Barr, 2008) in the package *brms* (Buerkner, 2018) to fixations in the critical window beginning at adjective onset and extending to the noun onset, adding 200 ms for saccade planning (Hallett, 1986; Salverda et al., 2014). Full random effects structure for participants and items was included. The presence of a predictive effect was operationalised as a target-over-competitor advantage, and thus we compared the relative difference in fixations to the target vs. the colour competitor—e.g. button vs. bottle in (3). To reflect the expected increase in fixations in the match vs. mismatch conditions, the effect of condition was sum contrast coded: 0.5 for match and −0.5 for mismatch.

Fitting models in a Bayesian framework allows incorporating prior expectations for the effects in the critical window. We used weakly informative priors that constrained the models to estimate psycholinguistically plausible parameter values, but ensured that the prior would not outweigh the evidence provided by the data (Gelman, 2020; Gelman et al., 2008; Schad et al., 2020). Our prior for the probability of predicting the target across conditions (i.e. the model intercept) followed a normal distribution of $N(0,\ 1)$ on the log odds scale. This prior indicated that the probability of predicting the target would fall between 12% and 88% with 95% probability. Our prior for whether predictions would be more likely in the match than in the mismatch condition (i.e. the model slope) also followed a standard normal distribution of $N(0,\ 1)$.

The above models served to estimate the effect size of our predictors of interest. We then used Bayes factors (BF) to quantify evidence for: (i) whether there was a difference in the probability of fixating the target vs. colour competitor; and (ii) whether the probability of fixating the target vs. competitor differed between the match and mismatch conditions. In both (i) and (ii), the alternative hypothesis (H_1) was that there was a difference and the null hypothesis (H_0) was that there was no difference. Bayes factors were computed using bridge sampling (Bennett, 1976; Gronau et al., 2017; Meng & Wong, 1996), which provided a ratio of the marginal likelihoods of H_1 and H_0 (BF_{10}). A BF_{10} of 1 suggests equivalent evidence for either hypothesis. In line with Lee and Wagenmaker's scale (2013; adapted from Jeffreys (1961), a BF_{10} above 1 indicates evidence for H_1: values between 1 and 3 are considered anecdotal evidence, values between 3 and 10 moderate evidence, and values over 10 strong evidence. Conversely, a BF_{10} below 1 indicates evidence for H_0: values between 1 and 0.3 and are considered anecdotal evidence, between 0.3 and 0.1 moderate evidence, and below 0.1 strong evidence. Note that these categories are only a guide and that Bayes factors should be interpreted in a continuous fashion.

Onset times. To determine the onset of the predictive effect and its possible divergence between conditions, we used a bootstrapping approach for time-series data proposed by Stone and colleagues (2020). This procedure estimates the onset of an effect and provides a measure of its temporal variability using bootstrapping, i.e. resampling the original data and reapplying the onset estimation procedure multiple times. We adapted the procedure to create a Bayesian version that allowed us to incorporate knowledge about plausible effect sizes and to estimate a continuous posterior probability distribution of onset times. The Bayesian version also allowed us to quantify evidence for our hypotheses using Bayes factors.

Two components were necessary to compute posterior distributions for prediction onset times. First, we needed priors to encode our expectations about the expected onsets. For the onset of predictive effects in the match and mismatch conditions, we reasoned that these could only arise in the 700 ms time window between the onset of the adjective and the onset of the noun (adding 200 ms for saccade planning). We thus specified a normal distribution centred in the middle of this critical window, with a 95% probability of falling between 200 and 900 ms: $N(550,\ 175)$. This prior centred the probability of the match-mismatch difference on zero—consistent with no difference between conditions—with a symmetric probability of being positive or negative.

Second, we needed onset time data to inform posterior inference via a likelihood function. This involved two steps: First, a likelihood was approximated using a distribution of onset times estimated via the bootstrap procedure proposed by Stone et al. (2020). The bootstrap procedure involved fitting a linear model of weighted empirical logits to compare fixations to the target vs. competitor at each timepoint in the empirical data. The first significant test statistic in a run of five consecutive significant tests was considered the onset of predictive looks, corresponding to a target-over-competitor preference sustained for at least 100 ms. A distribution of plausible divergence points was obtained by resampling the original data 2000 times and repeating the procedure after each resample. Second, a function for the likelihood of the bootstrapped onset time data was approximated using a normal distribution (i.e. Laplace approximation), under the assumption that the population distribution underlying the bootstrap distribution was normal.[2]

We were then able to use the normal distributions of the respective priors and likelihoods to estimate the posteriors for the match/mismatch conditions as the product of two Gaussians (e.g. Smith, 2011). The posterior for each condition was then used to estimate the between-condition differences by computing a normal difference distribution (Weisstein, 2020). Finally, we computed Bayes factors using the Savage-Dickey method to assess the degree of evidence for the hypothesis of a difference in the match vs. mismatch condition (Dickey & Lientz, 1970).

Possessor identification. Participants' task during the experiment was to click on the target object (possessee) and then on its owner (possessor). Accuracy in selecting the possessee object was used as an attention check, such that we only analysed trials in which the correct object was selected. Participants' accuracy and response time to the possessor served to assess whether the (mis)-match manipulation affected the recall of the referent of the possessive. Accuracy was at ceiling in both conditions and thus was not analysed further. For the response times, we fit a linear mixed-effects model with a lognormal likelihood and maximal random effects structure. We used weakly informative priors of $N(0,1)$ for all fixed effects (Gelman, 2020; Gelman et al., 2008). At an average response time of 1000 ms, this corresponds to a plausible range of response times between 150 and 7400 ms. The effect of condition was sum contrast coded: 0.5 for mismatch and −0.5 for match.

Results

We report the posterior probability distribution of each fixed effect, reflecting the probability of different effect sizes given the statistical model and the observed data. We present the mean of each posterior distribution and its 95% credible interval (CrI), which represents the interval in which the population mean is estimated to lie with 95% probability. All results correspond to the adjective time-window, excluding samples with blinks or saccades (17%), samples in which none of the four objects was fixated (23%) and trials in which participants clicked on the wrong object (0.17%). The analysis focused on fixations to the target vs. colour competitor, excluding fixations to the other objects. Note that "competitor" always refers to the colour competitor. For ease of interpretability, estimates are back-transformed from logits to percentages (for the fixation probability analyses) and from log milliseconds to milliseconds (for response time analyses).

Critical time-window

Figure 2 shows the fixation patterns during the entire experimental trial, averaged across participants. At the beginning of the trial, fixations to each object were at chance level—25% given the four objects on-screen. After the onset of the adjective, looks to the target and colour competitor increased sharply, with a steep decline of fixations to the colour non-matching objects.

The time-window analysis showed clear evidence of a predictive effect, but not of a between-condition difference. The mean of the posterior distribution for the target-over-competitor advantage, i.e. the predictive effect, was 59% and its 95% CrI ranged from 53% to 66%. A Bayes factor of 6 indicated moderate evidence in favour of the predictive effect. With respect to whether the target-over-competitor advantage was more likely in the match than in the mismatch condition, the posterior estimate was 50 [43, 57]%. The Bayes factor was 0.17, indicating moderate evidence against a difference between conditions.

Onset times

In the match condition, the posterior estimate of the prediction onset was 548 [392, 704] ms, while in the mismatch condition it was 654 [611, 697] ms. The posterior estimate of the difference between conditions was 106 ms [−56, 268] ms. Most of the probability mass covered positive values, consistent with a

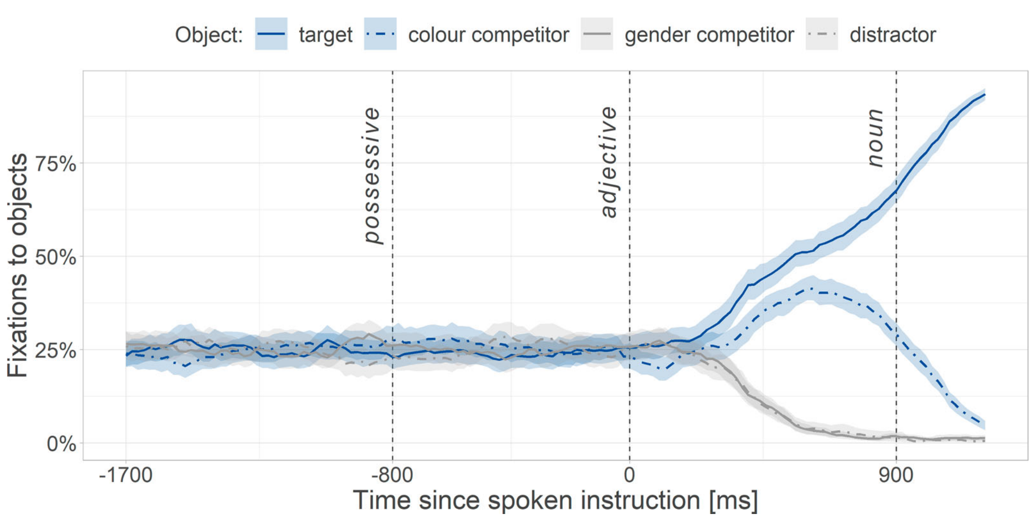

Figure 2. Fixation curves and bootstrapped 95% confidence intervals for the four objects on screen in Experiment 1. The predictive window extended from the onset of the colour adjective to the onset of the noun shifted 200 ms to the right. The x-axis is time-locked to the onset of the adjective.

delayed onset in the mismatch vs. match condition. However, the CrI was wide and also included values consistent with no difference, or with an earlier onset in the mismatch condition. A Bayes factor of 0.69 indicated anecdotal evidence against a difference between conditions. However, a sensitivity analysis using priors with a smaller standard deviation indicated moderate evidence in favour of a small between-condition difference (Appendix 1.2 Supplemental data). Figure 3 shows mean fixation proportions to the target and competitor objects, together with the onset of the predictive effect in each condition and the distribution of the between-condition difference.

Possessor identification

Participants' accuracy in identifying the object's possessor was near ceiling in both conditions (match condition: 99.7%; mismatch condition: 100%). The "*I don't know*" option—visually depicted by question mark—was selected in only 0.17% of trials, showing that participants were able to successfully identify the referent of the pronoun and did not process the pronoun "*ihr*" as ambiguous (i.e. referring to a plural third person antecedent, see *Materials*). The Bayes factor indicated strong evidence of slower responses in the mismatch than in the match condition (Table 1).

Discussion

Experiment 1 examined whether German speakers used the agreement information encoded in the suffix of a possessive pronoun to inform their predictions about upcoming words. We also examined whether the retrospective dependency encoded by the stem of the possessive (possessor agreement) influenced participants' predictive computations, consistent with an interaction between the two agreement dependencies. The time-window analyses showed that participants were able to predict the noun using the gender-marked suffix of the possessive. During presentation of the adjective, fixations to the target object matching the suffix gender increased more than those to a competitor object of different gender. Thus, our results are consistent with previous research on gender-marked determiners and adjectives (Brouwer et al., 2017; Dahan et al.,

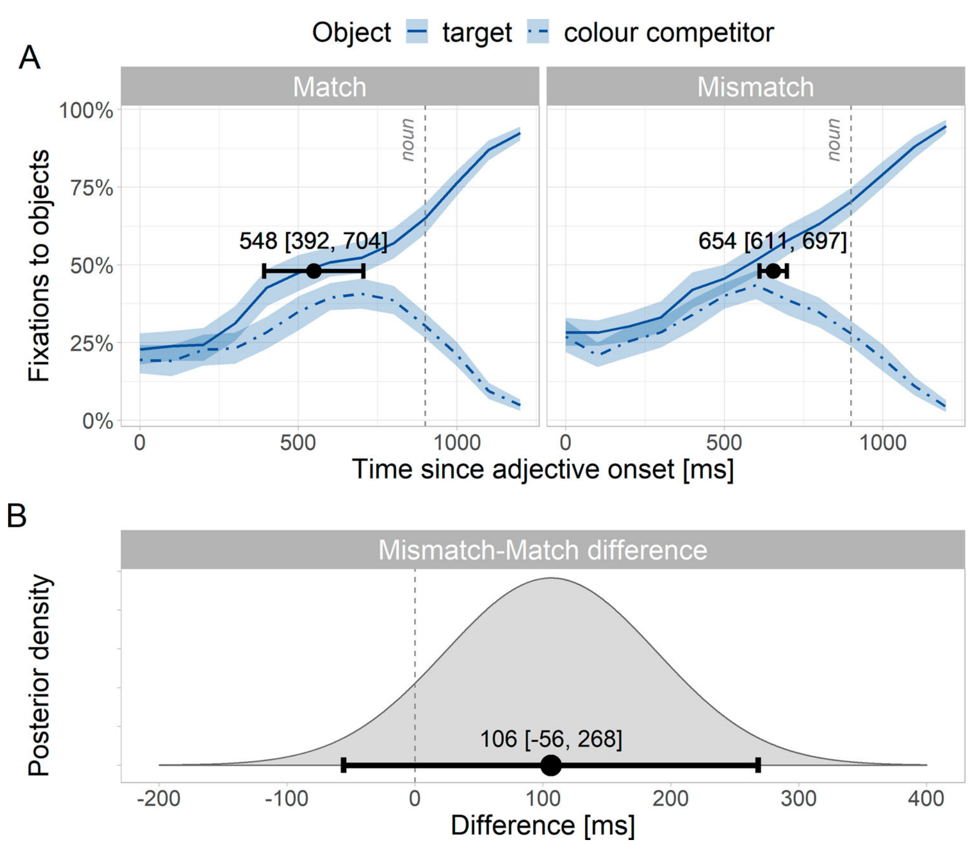

Figure 3. Onset time results in Experiment 1. **A.** Estimated predictive onsets overlaid on the fixation curves in each condition. Black points and error bars indicate onset means and 95% credible intervals respectively. **B.** Posterior of the difference in onset times between the match and mismatch conditions. While the difference distribution suggests a delayed onset in the mismatch condition, its wide 95% credible interval shows a large amount of uncertainty.

Table 1. Model-estimated response time differences between the mismatch vs. match condition in the possessor identification task in Experiments 1 and 2.

	Experiment 1			Experiment 2		
	Estimate (ms)	95% CrI	BF_{10}	Estimate (ms)	95% CrI	BF_{10}
Mismatch effect	82	[39, 124]	18	15	[−25, 55]	0.07

2000; Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018; Lew-Williams & Fernald, 2010). We extend these results by showing that gender predictions also occur when participants have to concurrently resolve an additional dependency—i.e. the retrieval of a referent.

With regard to whether the syntactically irrelevant possessor gender feature influenced the target noun prediction, our findings were inconclusive. We hypothesised that if predictions were not fully grammatically restricted, the conflict between gender features in the mismatch condition should negatively impact predictions as overriding the conflict may require additional time and processing effort. The time-window analysis suggested that when fixations in the critical time-window were considered as a whole, there was evidence against this hypothesis.

However, our specific hypothesis concerned the onset of predictions, and whether it would be affected by the gender alignment between the prospective and retrospective dependencies. To directly assess this hypothesis, we used a novel Bayesian bootstrapping analysis to diagnose the onset of predictive effects. The results showed that the target-over-competitor advantage started on average 106 ms later in the mismatch than in the match condition. However, the Bayes factor analysis—which used weakly informative priors—showed anecdotal evidence against a difference between conditions. Because the Bayes factor is sensitive to the informativity of the prior and weakly informative priors can bias the Bayes factor towards supporting the null hypothesis (Lee & Wagenmakers, 2013), we performed a sensitivity analysis to assess the effect of different priors on our conclusions. This analysis showed that a more informative (narrower) prior yielded moderate evidence in favour of a between-condition difference, supporting a small mismatch effect (Appendix 1.2, see Supplemental data).

One explanation for the lack of conclusive evidence for a difference between the match and mismatch conditions is that the gender features of the possessive may not have interacted in a uniform way across all trials or across participants, giving rise to individual processing variability. Some support for this hypothesis may be found in the wide credible interval of the onset of the between-condition difference, which spanned 324 ms. This variability could be related to the design of the task, which may have rendered gender cues less diagnostic and/or harder to use in real time. Each display contained four objects, requiring participants to store the gender specifications of four different nouns in their working memory. These different gender encodings may have taxed participants' memory by increasing the amount of interference, and thus reduced the diagnosticity of gender cues. Moreover, in order to predict the target object, participants needed to attend to both gender and colour information, as only their combination rendered the target fully predictable. If both types of information had the same weight in prediction formation, we would expect as many looks to the gender-matched competitor as to the colour-matched competitor. However, the colour competitor was fixated more often than the gender competitor during the predictive window, as shown in Figure 1. This pattern suggests that visual cues were prioritised over linguistic cues during predictive processing, as found in other studies (Coco & Keller, 2015; Kukona et al., 2014). If participants prioritised colour information, either due to the visual nature of the task or because the gender features often conflicted, this may have reduced their capacity to integrate the gender of the possessive stem and suffix, decreasing the magnitude of the mismatch effect (for visual world studies on cue integration see DeDe, 2010, 2012; Henry et al., 2017).

To address these concerns, we conducted a follow-up experiment that attempted to simplify the experimental task in order to focus on the use of gender information. We modified our design by presenting only two objects on the screen and by having both objects match in colour but differ in gender. We speculated that a reduced number of objects would decrease encoding interference during the task and would also circumvent the need to integrate gender and colour cues, as only gender information was necessary to predict the target object. Finally, we sought to increase the chance of detecting experimental effects by increasing the number of trials per condition.

Experiment 2

Experiment 2 focused on the effect of conflict between gender features by removing the need for colour information to predict the target object. We used the same experimental design as Experiment 1, except that only two objects, the target and colour competitor, were displayed on-screen. Both objects were of the same colour but only the target matched the gender in the instruction (Figure 4). The colour adjective was still present in the auditory instruction to give listeners more time to process the gender cues of the possessive. The critical comparison was between fixations to the target vs. competitor, with a target-over-competitor advantage reflecting the predictive use of gender information. As for Experiment 1, we hypothesised that the conflict in gender features might derail predictions, resulting in less likely or delayed target predictions in the mismatch vs. match condition.

Methods

Participants

Seventy-four native German speakers participated in Experiment 2. The sample size for Experiment 2 was kept consistent with Experiment 1 and again reflected the maximum feasible number of participants recruitable during the academic semester. Data from four participants were excluded due to tracker loss, leaving 70 participants for the analysis (mean age 26 years, age range 18–53 years, 52 female, 59 right-handed). Participants reported exposure solely to German until age six and no history of neurological, reading or writing disabilities.

Materials, procedure and analysis

The materials consisted of 96 experimental items, resulting in 48 trials in each experimental condition. The experimental items always featured two objects of the same colour. Each object was repeated twice throughout the experiment, once as a target and once as a competitor. There were 12 filler trials, in which the two objects on-screen had the same gender but differed in colour. The fillers were presented in a block after the experimental trials, in order to remove the need to attend to colour information during the experimental trials. The onset of the target noun in the auditory instruction was 100 ms later than in Experiment 1 and there was no "*I don't know*" option in the possessor identification task. We removed this option because it was seldom used in Experiment 1. All other aspects of the presentation were identical to Experiment 1.

The testing procedure was similar to Experiment 1, but participants additionally completed an object naming test prior to the eye-tracking session. This was done because while all objects were highly identifiable, individual participants may have used different names for the same object (e.g. *die Blüte* 'the blossom' instead of *die Blume* 'the flower'). Participants were presented with grey-coloured target objects and asked to name them using a determiner and a noun, e.g. "*der Knopf*". Trials in which participants gave an incorrect gender for the target object or used a synonym with a different gender were excluded from the analysis. The experimental session lasted approximately 40 mins.

For the time window analysis, the critical window ranged from the possessive onset to the noun onset plus 200 ms for saccade planning. The critical window

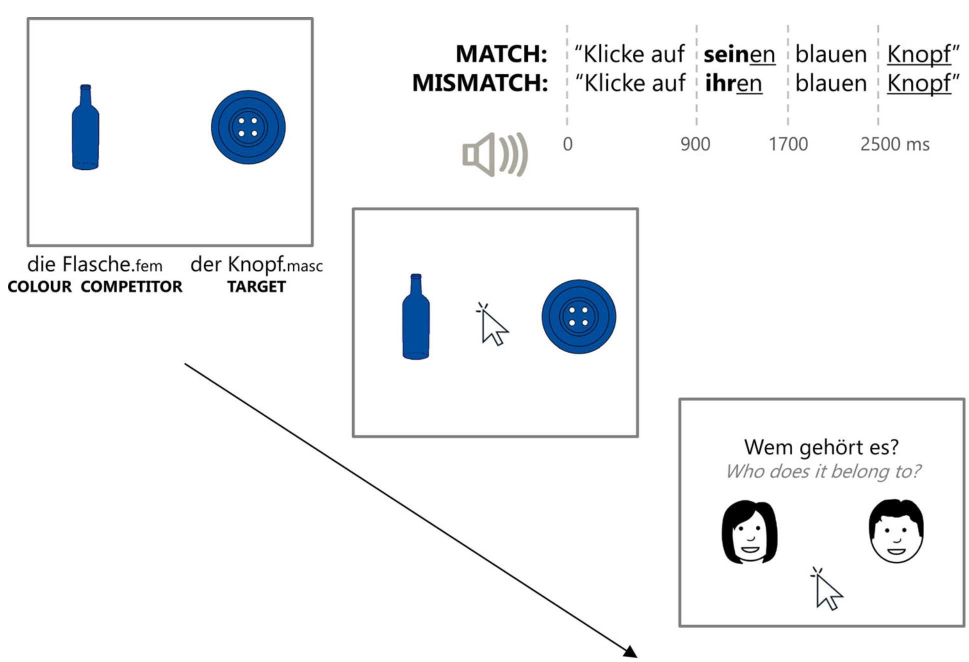

Figure 4. Sample experimental trial in Experiment 2. Trials were identical to those of Experiment 1, except that there were only two objects on screen, the onset of the noun in the auditory instruction was 100 ms later, and there was no longer an "*I don't know*" option in the possessor identification task.

started at the possessive because the target object was already predictable at the possessive offset (in contrast with Experiment 1) and we were interested in the effect of possessor match/mismatch. Since Experiment 1 provided us with some knowledge about prediction probabilities, we used its posterior estimates to calibrate our priors for Experiment 2 (e.g. Nicenboim et al., 2020). The priors were calibrated using the means of the posterior parameters in Experiment 1. Uncertainty given the change in experimental design was incorporated by specifying prior standard deviations that were larger than the posterior standard deviation in Experiment 1, but still narrower than the original prior standard deviation. The prior for the target-over-competitor advantage was $N(0.37, 0.25)$, indicating that the probability of predicting the target was between 47% and 70% with 95% probability. The prior for whether predictions would be more likely in the match than in the mismatch condition was $N(0.01, 0.25)$, reflecting a 95% probability range of 38% to 62%.

The onset analysis was as for Experiment 1, but we also used the posterior estimates from Experiment 1 to inform the priors of Experiment 2. We took into account that the predictive effect needed to be time-locked to the possessive offset in Experiment 2, as opposed to the adjective onset in Experiment 1, since the adjective was no longer necessary for predicting the noun. Because each of the four possessive forms had a different offset—"*seine*", "*seinen*", "*ihre*", "*ihren*"—we took the average offset across forms (612 ms) and added the relevant posterior mean onset from Experiment 1 to specify the prior mean. As for the window analysis above, the prior standard deviation was larger than the posterior standard deviation from Experiment 1, but narrower than the original prior standard deviation. The prior for the match condition was thus $N(1160, 100)$, corresponding to an onset time of 1160 ms with a 95% probability range between 960 and 1360 ms. The prior for the mismatch condition was $N(1266, 100)$, corresponding to an onset time of 1266 ms and a 95% probability range between 1066 and 1466 ms.

Response times in the possessor identification task were analysed as for Experiment 1, but again we used the posterior means from Experiment 1 to specify prior means and standard deviations that were larger than the posterior standard deviations for Experiment 1 but more informative than the original prior standard deviations.

Results

Critical time-window

Samples with blinks or saccades (14%) were excluded, as well as samples in which none of the four objects was fixated (18%). Trials were excluded if the wrong target object was selected (0.47%) or if an incorrect answer was given in the naming test (1%). For one participant, 58% of fixations had to be vertically adjusted due to issues with eye tracker calibration. Figure 5 shows the fixation patterns to the two objects during the entire experimental trial, averaged across participants. At the beginning of the trial, fixations to each object were at chance level—50% given the two objects on-screen.

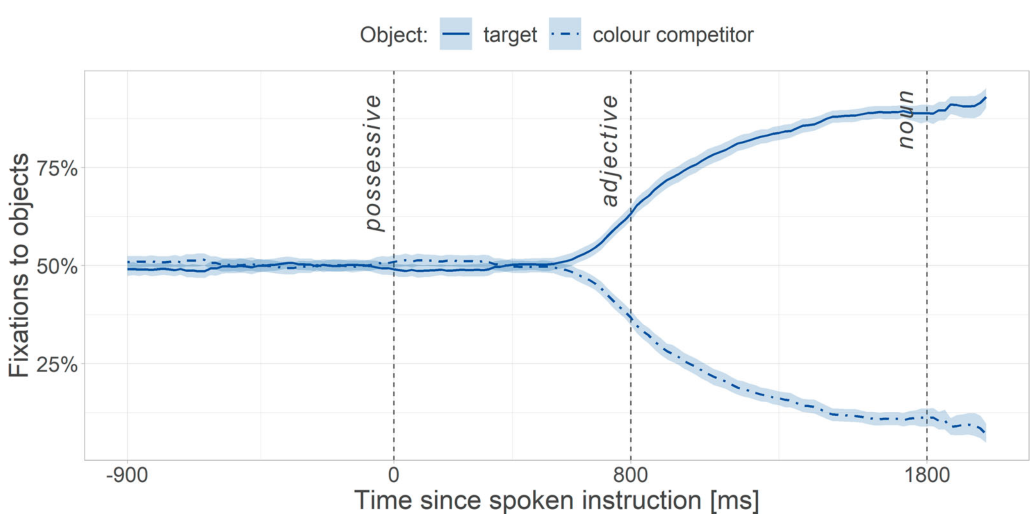

Figure 5. Fixation curves to the two objects averaged across items and participants and their bootstrapped 95% confidence intervals in Experiment 2. The predictive window extended from the onset of the possessive to the onset of the noun, shifted 200 ms to the right. The x-axis is time-locked to the possessive onset.

Shortly before the adjective onset, looks to the target sharply increased.

The time-window analysis showed clear evidence of a predictive effect. The posterior estimate of the probability of a predictive effect was 66 [64, 68]% and a Bayes factor of over 7×10^{18} indicated strong evidence in favour of a target-over-competitor advantage. In contrast to Experiment 1, there was also evidence that predictions were more likely in the match than in the mismatch condition. The estimated probability of predicting the target in the match vs. mismatch condition was 52 [50, 54]% and the Bayes factor of 3 indicated moderate evidence in favour of a difference between conditions.

Onset times

The estimated prediction onset was 402 [361, 444] ms in the match condition and 709 [692, 726] ms in the mismatch condition. The estimate of the between-condition difference was 307 [262, 352] ms, consistent with a later predictive onset in the mismatch condition (Figure 6). A Bayes factor of over 1×10^{38} indicated strong evidence of a between-condition difference.

Possessor identification

Accuracy in identifying the possessor was at ceiling in both conditions. The Bayes factor indicated strong evidence against a difference in response times between the match and mismatch conditions, thus failing to replicate Experiment 1 (Table 1). We therefore do not discuss this task further.

Discussion

Experiment 2 sought to clarify the existence of a difference in onset times between the match and mismatch conditions. We used a simpler design with a larger number of items to assess whether the retrospective dependency encoded in the possessive stem influenced the prediction of an upcoming noun. This hypothesis was supported by the onset analysis, which this time convincingly demonstrated that target predictions were on average 307 ms slower in the mismatch condition. Further evidence came from the time-window analysis, which showed that participants not only predicted the target noun, but also that the predictive effect was reduced when possessive stem and suffix mismatched in gender.

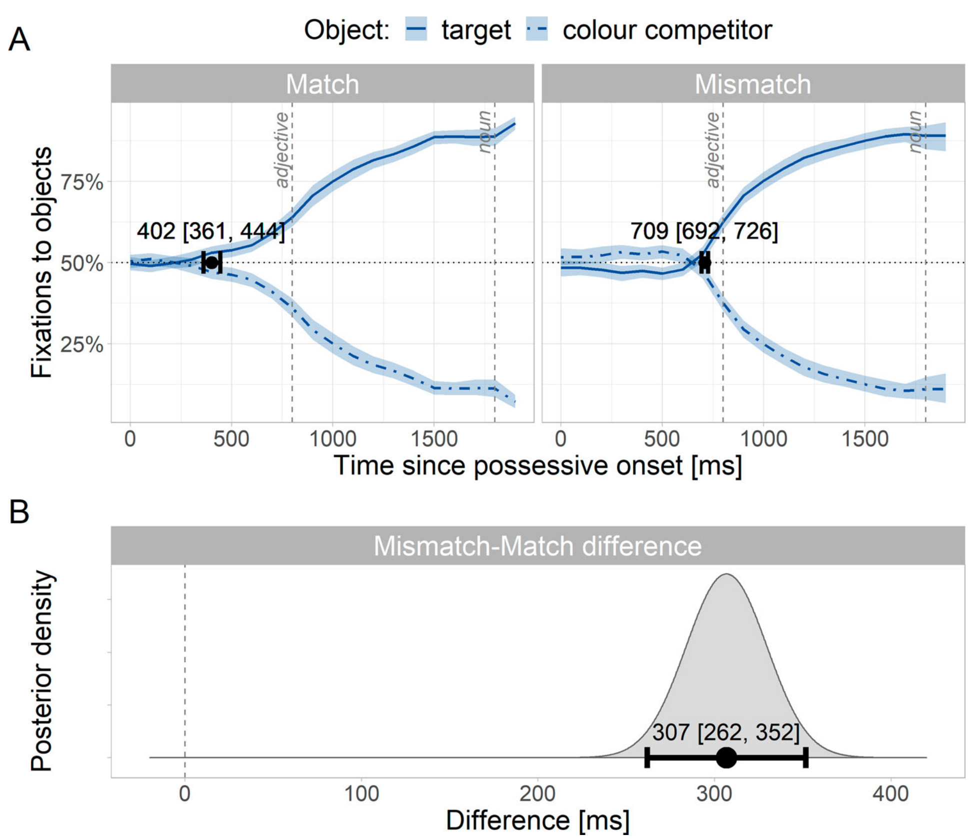

Figure 6. Onset time results in Experiment 2. **A.** Estimated predictive onsets are overlaid on the fixation curves in each condition. **B.** Posterior of the difference in predictive onsets between the match and mismatch conditions, with the point and errorbar indicating the mean and 95% credible interval. The difference distribution was consistent with a delayed prediction onset in the mismatch condition.

The analysis of predictive onsets in Experiment 2 showed not only a between-condition difference, but also surprisingly early onsets with respect to the possessive offset, which was the earliest point in time in which the target could be anticipated. While the possessive offset occurred on average at 612 ms, the predictive onset in the match condition arose around 402 ms. This raises the question: how could participants start looking at the target object before hearing the possessive suffix? One explanation may be phonetic coarticulation: our sentences were produced in a naturalistic manner, and thus the sounds in the possessive varied as a function of their local phonetic environment. For example, the vowel in the suffix (either "*-en*" or "*-e*") was shorter and more nasal when it appeared before a nasal, consistent with how coarticulation is used in spontaneous contexts (e.g. Zellou et al., 2016). Comprehenders may have used coarticulatory information to start predicting before the possessive offset. Alternatively, comprehenders may have based their noun prediction on the stem of the possessive ("*sein-*" or "*ihr-*"), before even hearing the suffix. This explanation has important implications for questions about the grammatical faithfulness of predictions (see *General Discussion*).

The early onset of the predictive effect in Experiment 2 additionally rules out the possibility that participants based their noun predictions solely on the colour adjective. Predictions based on the adjective were possible because the adjective's suffix also agreed in gender with the upcoming noun. However, the target-over competitor advantage in Experiment 2 arose (and remained present) prior to this suffix. Thus, while it is likely that the adjective suffix contributed to the predictive effect, it cannot have been its only source, such that predictions were clearly driven by the possessive suffix.

Finally, one potential disadvantage of the two-object design in Experiment 2 is that since each object appeared twice, participants may have noticed that once an object had been a target, its next appearance would be as a competitor (and vice versa). This could have led to a prediction strategy that had no relation to morphosyntactic gender. However, if this were the case, we would expect to see increased fixations to the target object before the possessive, as participants would have been able to predict the target from the trial onset. Figure 5 confirms that this was not the case. A similar strategy may have been possible in Experiment 1, although this would have required participants to recall which of the four objects had been previously seen and in which role, an option that seems unlikely.

General discussion

Our study addressed whether speakers use the prospective agreement relationship encoded in possessive pronouns to predict upcoming nouns, and also whether these predictions were influenced by the computation of a grammatically distinct dependency. Because German possessives encode not only a prospective dependency but also a retrospective dependency with a previously mentioned referent (possessor agreement), we were able to examine predictions when the agreement features of the prospective and retrospective dependency conflicted. We reasoned that if predictive computations were not fully grammatically constrained, then their resolution may be affected by the gender of the possessor, suggesting an interaction between the two agreement dependencies. This result would support the claim that situations of conflict can give rise to grammatically unlicensed predictions, as found for other linguistic phenomena (Kamide & Kukona, 2018; Knoeferle & Crocker, 2006, 2007; Kukona et al., 2011, 2014; Rommers et al., 2013, 2015). We assessed the effect of conflict in the timecourse of predictions by creating a Bayesian version of an onset analysis (Stone et al., 2020), which allowed comparing prediction onsets between conditions.

The time-window analysis of the two visual world experiments showed that comprehenders used the gender-marked suffix of the possessive to predict an upcoming noun, as evidenced by anticipatory fixations to the target object. These results are consistent with previous studies showing that speakers use agreement-marked adjectives and determiners to anticipate nouns (Brouwer et al., 2017; Dahan et al., 2000; Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018; Lew-Williams & Fernald, 2010). We extend these findings by demonstrating that agreement-based predictions also occur with pronouns, and that speakers can make gender predictions even when the pronoun encodes an additional dependency.

Additionally we found that possessor agreement affected the timecourse of predictions. In Experiment 1, predictions were delayed by approximately 100 ms in the mismatch condition, although the Bayes factor analysis was inconclusive. Using a simpler design and more experimental items, Experiment 2 showed a reliable delay of approximately 300 ms. Taken together, both experiments are consistent with a delayed onset of predictions in the mismatch condition, which suggests that conflict between possessor and possessee agreement—two grammatically distinct dependencies in German—can temporarily derail predictions. Our findings resemble previous visual world studies reporting that conflicting information can lead comprehenders to entertain interpretations unlicensed by global sentence constraints. However, while previous studies focused on cases of conflict between different information types (e.g. grammatical vs. lexical), we demonstrate that conflict may also arise between features at the same level of representation. This suggests that unlicensed predictions are not just the result of, for example, conflict between different levels of processing, but rather that the process of conflict resolution itself takes time to complete. Note that our results should not be interpreted as evidence that comprehenders predict syntactically unlicensed words. Their scope is more specific: they demonstrate that predictions are temporarily sensitive to grammatically illicit information and, crucially, that this can affect the timecourse of lexical expectations, even if the appropriate target is eventually predicted.

While we believe that the results of Experiment 2 are more informative than those of Experiment 1 due to the simpler design and a larger number of experimental trials, the results of Experiment 1 are still informative and consistent with Experiment 2. However, the fact that the mismatch effect was weaker in Experiment 1 highlights the potential saliency of colour as a predictive cue in visual tasks, as has been noted elsewhere (Kukona et al., 2014). We have speculated that attention to the colour cue decreased participants' capacity to attend to and/or integrate gender features, dampening the mismatch effect in Experiment 1. If so, the difference between experiments is potentially informative about the role of feature integration and attention demands in predictive processing, which is an interesting avenue for future investigation.

A final result that merits discussion concerns the surprisingly early onset of predictions in Experiment 2. In the match condition a target-over-competitor advantage was observed after only 400 ms post possessive onset, which is earlier than the offset of the possessive forms. As suggested above, one explanation is that participants exploited coarticulatory information to infer the nature of the suffix prior to its offset. But the early onset may also suggest a different interpretation of the match-mismatch difference: Given that the onsets precede the point in the possessive where any conflict could have arisen (i.e. the suffix), a conflict account of the mismatch condition may not be appropriate. Instead, it is possible that participants relied on the gender of the possessive stem alone to predict the target object. This would have resulted in an early preference for the target object in the match condition, which would then have been reinforced by the congruent gender feature of the possessive suffix. The reverse would have been true in the mismatch condition. The match-mismatch difference could therefore instead be characterised as a facilitatory boost in the match condition. This "facilitation account" is similar to a conflict resolution account in that both appeal to the congruency between the two agreement dependencies to explain timecourse differences. However, they imply different underlying mechanisms, which we discuss below.

Cognitive mechanisms underlying the timecourse of predictions

A conflict-based explanation of our results would posit that the two conflicting gender features in the mismatch condition delayed the prediction of the target noun because comprehenders required more time to arbitrate between the conflicting features. This would resemble a Stroop effect, in which slower behavioural responses are observed when participants have to name incongruent stimuli, e.g. the word "*red*" printed in green colour (Cohen et al., 1990; MacLeod, 1991). According to this conflict-based account, in an instruction like "*Klicke auf* ***ihr****en* … (*Knopf*)" the two different gender features in the possessive may have concurrently activated two different elements in participants' memory: the noun associated with the target object (the masculine button), which matched the gender of the possessive suffix, and the noun associated with the competitor object (the feminine bottle), which matched the gender of the possessive stem. Competition between the two coactivated representations may have led to lower activation of the target noun, increasing the time needed for its selection. Participants would therefore have taken longer to start looking at the image of the target object in the mismatch than in the match condition.

Alternatively, processing may have been facilitated in the match condition, if the gender of the possessive stem (incorrectly) preactivated gender-congruent nouns. This would resemble a facilitatory priming effect, which occurs when multiple inputs activate the same representation such that it reaches a response threshold more quickly (Collins & Loftus, 1975; McNamara, 2005; Meyer & Schvaneveldt, 1971; Neely., 1991; Neely & Keefe, 1989). Under a facilitation account, the masculine stem in the instruction "*Klicke auf* ***sein****en* … (*Knopf*)" could preactivate the memory representation of "*Knopf*", facilitating its selection as the prediction target if a gender-matching possessive suffix was heard. Participants would thus predict the target object more quickly in the match condition.

While the present results do not allow us to arbitrate between a conflict- and a facilitation-based mechanism, these alternatives could be empirically dissociated in future studies. For example, participants' processing after a possessive pronoun in the match and mismatch conditions could be compared with their processing after a determiner, which in German encodes only the gender of the upcoming noun: "*Klick auf den Knopf*" (*Click on the.$_{MASC}$ button*). Since the determiner has only one gender feature and thus cannot elicit the same gender-based conflict or facilitation as the possessives, fixation patterns in a determiner condition could be considered a baseline to be compared against the possessive conditions. If conflict slows predictions in the mismatch condition, then the match condition should pattern closely with the determiner baseline. If predictions are facilitated in the match condition, then the mismatch condition should pattern closely with the determiner baseline.

Alternatively, both conflict- and facilitation-based mechanisms may be involved. A computational account incorporating both mechanisms is offered by the cue-based memory retrieval framework (for a recent review see Vasishth et al., 2019). Under the LV05 cue-based retrieval model (Lewis & Vasishth, 2005), the morphosyntactic features in the possessive stem should trigger the memory retrieval of a feature-matching antecedent (Engelmann et al., 2019; Jäger et al., 2017; Patil et al., 2016). For example, the antecedent retrieval cues at "*sein-*" in the match condition (his) would include [+masculine] and [+animate] features, which would best match the memory representation of "Martin". However, being also a partial match for the [+masculine] feature, the memory representation of "*Knopf*" (button.$_{MASC}$) would also receive some activation. This would facilitate prediction of the button when the masculine suffix "*-en*" is heard. In contrast, the antecedent retrieval triggered by "*ihr-*" (her) in the mismatch condition would partially activate the feminine noun "Flasche" (bottle.$_{FEM}$), thus increasing the competition between the bottle and button representations and delaying the prediction of the button. Thus, in a cue-based retrieval framework, both facilitation (preactivation) and conflict (interference) result from feature overlap.

Finally, it should be noted that the retrospective dependency encoded by the possessive in our experiments did not involve a linguistic antecedent in the auditory instruction, but rather a previously mentioned discourse referent, which had been introduced at the beginning of the experiment. An interesting question for future research is how the (mis)match effect would be affected if the auditory instruction had involved a linguistic antecedent, either within the critical sentence, as in "Give *Martin* his blue button", or in the preceding linguistic context, e.g. "Now *Martin* needs your help. Click on his blue button". This distinction may have processing consequences, as previous work has proposed that some types of features, such as biological gender, may be stored in both the lexicon and the discourse, while others, such as grammatical gender, may be stored solely in the lexicon (Cacciari et al., 1997; Frazier et al., 1996; Garnham, 2001; Garnham et al., 1995; Lago et al., 2017).

Conclusion

Two visual world studies examined the interaction of agreement constraints on prediction, finding that comprehenders were able to rapidly integrate the gender suffix of a possessive pronoun to inform their noun predictions. However, the timecourse of predictions was influenced by a second, grammatically distinct agreement dependency encoded in the pronoun, which conveyed a retrospective relationship. We conclude that the two grammatically distinct dependencies interacted, temporarily leading to syntactically unlicensed predictions. We propose that our results may be due to the interaction of memory retrieval and predictive mechanisms, suggesting that these operations are not encapsulated during online processing. Our findings highlight the utility of examining not only whether predictions occur, but also *when* they occur. By measuring timecourse differences, we can generate novel hypotheses about the use of grammatical constraints in predictive processing and the nature of predictive mechanisms.

Notes

1. There were three lists because our experiment included an additional set of items with definite determiners instead of possessive pronouns, e.g., "*Klicke auf den blauen Knopf!*". Since gender predictions have not previously been tested with possessive pronouns, we included the determiner items to confirm that our experimental design was able to replicate the predictive gender effects observed in previous research with articles and adjectives (e.g., Hopp, 2012, 2015; Dahan et al., 2000; Lew-Williams & Fernald, 2010; Dussias et al., 2013; Brouwer et al., 2017; Hopp & Lemmerth, 2018). This was indeed the case, and the results of the determiner items have been reported in Stone et al. (2020).
2. Using a normal distribution for the likelihood assumes that, with sufficient observations, the bootstrap distribution will approach a normal distribution, consistent with the central limit theorem (Hesterberg, 2002). In some cases however, a normal distribution may not be a good approximation of a likelihood function, for

example, if the bootstrap distribution is itself not normally distributed. In this case, a likelihood function could be defined using a kernel density estimator. We demonstrate this approach in Appendix A1 (see Supplemental data) and show that it yields similar posterior estimates for our data as the approach that assumes a normally distributed likelihood.

Acknowledgements

We would like to thank Pienie Zwitserlood, Umesh Patil, Dorothea Pregla and two anonymous reviewers for their valuable feedback. We also thank Lisa Becker for her assistance with materials creation and data collection.

Data accessibility

Full data and code for the manuscript are available at https://osf.io/mbtcd/

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Sol Lago and Kate Stone were supported by a German Research Council project awarded to Sol Lago (grant number LA 3774/1-1).

ORCID

Kate Stone http://orcid.org/0000-0002-2180-9736
João Veríssimo http://orcid.org/0000-0002-1264-3017
Daniel Schad http://orcid.org/0000-0003-2586-6823
Shravan Vasishth http://orcid.org/0000-0003-2027-1994
Sol Lago http://orcid.org/0000-0002-4966-1913

References

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in Japanese. *Syntax (oxford, England)*, *12*(2), 93–134. https://doi.org/10.1111/j.1467-9612.2009.00123.x

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26. https://doi.org/10.1016/S0010-0277(00)00073-1

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769. https://doi.org/10.1037/0278-7393.28.4.748

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–153. https://doi.org/10.1162/0898929052880101

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*(2), 245–268. https://doi.org/10.1016/0021-9991(76)90078-4

Bourdages, J. S. (1992). Parsing complex NPs in French. In H. Goodluck, & M. Rochemont (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 61–87). Springer Netherlands. https://doi.org/10.1007/978-94-017-1980-3_3

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, *159*, 50–65. https://doi.org/10.1016/j.jecp.2017.01.007

Buerkner, P.-C. (2018). Brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), https://doi.org/10.18637/jss.v080.i01

Cacciari, C., Carreiras, M., & Cionini, C. B. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, *37*(4), 517–532. https://doi.org/10.1006/jmla.1997.2528

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and spanish. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 639–663. https://doi.org/10.1080/713755647

Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience* 1–26. https://doi.org/10.1080/23273798.2018.1427878

Chow, W.-Y., Momma, S., Smith, C., Lau, E., & Phillips, C. (2016). Prediction as memory retrieval: Timing and mechanisms. *Language Cognition & Neuroscience*, *44*, 5. https://doi.org/10.1080/23273798.2016.1160135

Coco, M. I., & Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Quarterly Journal of Experimental Psychology*, *68*(1), 46–74. https://doi.org/10.1080/17470218.2014.936475

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480. https://doi.org/10.1006/jmla.1999.2688

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–1870. https://doi.org/10.1016/j.neuropsychologia.2012.04.011

DeDe, G. (2010). Utilization of prosodic information in syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, *39*(4), 345–374. https://doi.org/10.1007/s10936-009-9139-x

DeDe, G. (2012). Lexical and prosodic effects on syntactic ambiguity resolution in aphasia. *Journal of Psycholinguistic Research*, *41*(5), 387–408. https://doi.org/10.1007/s10936-011-9191-1

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov Chain. *The Annals of Mathematical Statistics*, *41*(1), 214–226. https://doi.org/10.1214/aoms/1177697203

Drummer, J.-D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language*, *101*, 97–113. https://doi.org/10.1016/j.jml.2018.04.001

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). Multipic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, *35*(02), 353–387. https://doi.org/10.1017/S0272263112000915

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: A computational account. *Cognitive Science*, *43*(12). https://doi.org/10.1111/cogs.12800

Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220. https://doi.org/10.1017/S0142716410000342

Frazier, L., Henstra, J., & Flores d'Arcais, G. B. (1996). Finding candidate antecedents: Phrases or conceptual entities. *University of Massachusetts Occasional Papers in Linguistics*, *22*(8), 47. https://scholarworks.umass.edu/umop/vol22/iss1/8

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

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*, 340–351. https://doi.org/10.1016/j.cognition.2016.10.003

Garnham, A. (2001). *Mental models and the interpretation of Anaphora*. Psychology Press.

Garnham, A., Oakhill, J., Ehrlich, M. F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language*, *34*(1), 41–62. https://doi.org/10.1006/jmla.1995.1003

Gelman, A. (2020). *Prior choice recommendations*. Stan Development Repository. https://github.com/stan-dev/stan

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (Sign) and Type M (Magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*(6), 715–728. https://doi.org/10.1023/A:1023277213129

Hallett, P. (1986). Eye movements and human visual perception. *Handbook of Perception and Human Performance*, *1*, 10–11.

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

Hesterberg, T. (2002). Bootstrap methods and permutation tests. In S. Monaghan, D. S. Moore, A. Clipson, & R. Epstein (Eds.), *The practice of business statistics* (Ch18, 4–25). New York, NY: WH Freeman and Co.

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. *Proceedings of the 36th Annual Boston University Conference on Language Development*, 226–245.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

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

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PloS One*, *9*(7), e100986. https://doi.org/10.1371/journal.pone.0100986

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*. https://doi.org/10.1016/j.jml.2015.10.007

Jäger, L. A., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, *79-80*, 97–120. https://doi.org/10.1016/j.jml.2014.10.005

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. https://doi.org/10.1016/j.jml.2017.01.004

Jeffreys, H. (1961). *Theory of probability*. Clarendon.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The timecourse 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., & Kukona, A. (2018). The influence of globally ungrammatical local syntactic constraints on real-time sentence comprehension: Evidence from the visual world paradigm and reading. *Cognitive Science*, *42*(8), 2976–2998. https://doi.org/10.1111/cogs.12694

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

Kazanina, N., Lau, E., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384–409. https://doi.org/10.1016/j.jml.2006.09.003

Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology*, *63*(2), 371–400. https://doi.org/10.1080/17470210902974120

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative spanish: An eye-movement investigation. *Language Learning*, *59*(3), 503–535. https://doi.org/10.1111/j.1467-9922.2009.00516.x

Kennison, S. M. (2003). Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language*, *49*(3), 335–352. https://doi.org/10.1016/S0749-596X(03)00071-8

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

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*(1), 1–25. https://doi.org/10.1207/s15473341lld0201_1

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 326–347. https://doi.org/10.1037/a0034903

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. https://doi.org/10.1016/j.cognition.2010.12.002

Kuperberg, G., & 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

Lago, S., Sloggett, S., Schlueter, Z., Chow, W. Y., Williams, A., Lau, E., & Phillips, C. (2017). Coreference and antecedent representation across languages. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 795–817. https://doi.org/10.1037/xlm0000343

Laurinavichyute, A., Jäger, L. A., Akinina, Y., Roß, J., & Dragoy, O. (2017). Retrieval and encoding interference: Cross-linguistic evidence from Anaphor processing. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00965

Lee, M., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, *146*, 349–370. https://doi.org/10.1016/j.cognition.2015.10.012

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 432–460. https://doi.org/10.1037/0278-7393.24.2.432

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, *6*(4), 831–860. https://www.jstor.org/stable/24306045

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

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

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, *1228*, 161–176. https://doi.org/10.1016/j.brainres.2008.06.064

Momma, S., Sakai, H., & Phillips, C. (2015). *Give me several hundred more milliseconds: The temporal dynamics of verb prediction*. Talk given at the 28th annual CUNY Conference on human sentence processing, Los Angeles, USA.

Neely, J. H. (1991). Basic processes in reading: Visual word recognition. In D. Besner & G. Humphreys (Eds.), *Semantic priming effects in visual word recognition: A selective review of current findings and theories* (pp. 264–336).

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective-retrospective processing theory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 24, pp. 207–248). Academic Press. https://doi.org/10.1016/S0079-7421(08)60538-1

Ness, T., & Meltzer-Asscher, A. (2019). When is the verb a potential gap site? The influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience*, *34*(7), 936–948. https://doi.org/10.1080/23273798.2019.1591471

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research*, *1164*, 81–94. https://doi.org/10.1016/j.brainres.2007.05.058

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words preactivated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*. https://doi.org/10.1016/j.neuropsychologia.2020.107427

Omaki, A., Lau, E., White, I. D., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *06*, 1–18. https://doi.org/10.3389/fpsyg.2015.00384

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563–588. https://doi.org/10.1017/S0272263111000313

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773. https://doi.org/10.1006/jmla.1995.1033

Özge, D., Kornfilt, J., Münster, K., Knoeferle, P., & Küntay, A. C. (2016). *Predictive use of case markers in German children*. Proceedings of the 40th Annual Boston University conference on language development.

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00329

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*(4), 795–823. https://doi.org/10.1353/lan.2006.0217

Phillips, C., Wagers, M. W., & Lau, E. (2011). 5: Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces*, *37*, 147–180. https://doi.org/10.1108/S0092-4563(2011)0000037009

Pickering, M. J., Barton, S., & Shillcock, R. C. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 199–224). Lawrence Erlbaum London.

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

Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00590

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, and Psychophysics*, *77*(3), 720–730. https://doi.org/10.3758/s13414-015-0873-x

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447. https://doi.org/10.1016/j.neuropsychologia.2012.12.002

Ross, J. R. (1967). *Constraints on variables in syntax*. https://eric.ed.gov/?id=ED016965

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrase. *Cognitive Science*, *30*(2), 193–241. https://doi.org/10.1207/s15516709cog0000_58

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Schad, D. J., Betancourt, M., & Vasishth, S. (2020). Toward a principled Bayesian workflow: A tutorial for cognitive science. *Psychological Methods*. https://doi.org/10.1037/met0000275

Smith, J. O. (2011). Product of two Gaussian PDFs. *Spectral Audio Signal Processing*. https://ccrma.stanford.edu/~jos/sasp/Product_Two_Gaussian_PDFs.html

Stone, K., Lago, S., & Schad, D. J. (2020). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, 1–9. https://doi.org/10.1017/S1366728920000607

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. https://doi.org/10.1006/jmla.1996.0025

van Berkum, J., Zwitserlood, P., Bastiaansen, M. C. M., Brown, C. M., & Hagoort, P. (2004). So who's 'he' anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension. *Supplement to the Journal of Cognitive Neuroscience*, *16*. https://hdl.handle.net/11245/1.275264

van Gompel, R. P. G., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 128–139. https://doi.org/10.1037/0278-7393.29.1.128

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175. https://doi.org/10.1016/j.jml.2018.07.004

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*. https://doi.org/10.31234/osf.io/e4jds

Wagers, M. W., & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension1. *Journal of Linguistics*, *45*(2), 395–433. https://doi.org/10.1017/S0022226709005726

Weisstein, E. W. (2020). *Normal difference distribution* [Text]. Wolfram Research, Inc. https://mathworld.wolfram.com/NormalDifferenceDistribution.html

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32. https://doi.org/10.1016/j.cortex.2015.03.014

Zellou, G., Scarborough, R., & Nielsen, K. (2016). Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America*, *140*(5), 3560–3575. https://doi.org/10.1121/1.4966232

Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*. Article 34.