



Principle B constrains the processing of cataphora: Evidence for syntactic and discourse predictions

Dave Kush^{a,b,*}, Brian Dillon^c

^a Department of Linguistics, University of Toronto, Canada

^b Institute for Language and Literature, NTNU Norwegian University of Science and Technology, Norway

^c Department of Linguistics, University of Massachusetts, Amherst 01003, USA

ARTICLE INFO

Keywords:

Cataphoric pronouns
Binding Theory
Syntactic processing
Self-paced reading. Bayesian analysis

ABSTRACT

We tested whether comprehenders can use Binding Principle B (Chomsky, 1981) to guide antecedent search during the processing of cataphoric pronouns. We ran two self-paced reading experiments using the gender mismatch paradigm (van Gompel & Liversedge, 2003) as an index of active prediction of coreference between a cataphor and a main clause subject. In both experiments, we find gender mismatch effects at the main clause subject when coreference with the cataphor is grammatically acceptable. We do not find comparable gender mismatch effects in conditions where coreference is ruled out by Principle B. Our results are broadly consistent with models in which grammatical constraints serve as early filters on anaphora resolution processes in comprehension. We illustrate how the parser can integrate syntactic and discourse-level information to achieve grammatical sensitivity during incremental referential processing.

Introduction

There is strong evidence that incremental sentence comprehension involves predictive processing (Kuperberg & Jaeger, 2016; Kutas, DeLong & Smith, 2011; Smith & Levy, 2013; Nieuwland, 2019; Pickering & Gambi, 2018). One context where comprehenders arguably engage in a type of active linguistic prediction is the processing of backwards anaphora or *cataphora*, pronouns whose antecedent follows rather than precedes them. An example is in (1).

- (1) While *he* sat in the living room, *Loki* kept an eye on the kitchen.

When they encounter the pronoun in (1), comprehenders must identify a suitable referent. If the preceding linguistic context does not provide an antecedent, as in (1), comprehenders seem to proactively link the cataphor to a noun phrase in a subsequent position (Ackerman, 2015; Drummer & Felser, 2018; Filik & Sanford, 2008; Giskes & Kush, to appear; Kazanina, Lau, Yoshida & Phillips, 2007; Kazanina & Phillips, 2010; Patterson & Felser, 2019; Yoshida, Kazanina, Pablos & Sturt, 2014). Many researchers propose that the cataphor triggers an ‘active search’ for an antecedent. That is, comprehenders eagerly posit potential antecedents in upcoming syntactic positions without waiting for

bottom-up confirmation that their analysis is correct (e.g., Ackerman, 2015; Giskes & Kush, to appear; Kazanina et al., 2007).

Recent experiments have focused on determining where search initially posits potential antecedents, and how grammatical constraints influence the process. One hypothesis is that grammatical knowledge plays an early role in determining where the parser predicts antecedents. Under this *early filter* hypothesis, search prospectively posits antecedents in syntactic positions where coreference would be grammatical, but it ignores positions excluded by grammatical constraints. The alternative hypothesis is that grammatical constraints do not restrict coreference relations initially considered by the parser. According to this *delayed filter* hypothesis, search can posit an antecedent in an upcoming position, regardless of whether the grammar would allow the cataphor and an NP in that position to co-refer. If the result is an ungrammatical interpretation, grammatical constraints can be applied to ‘filter out’ unwanted coreference in later stages of processing.

In an influential study, Kazanina and colleagues (2007) argued that Principle C of the Binding Theory (Chomsky, 1981) constrains early prediction of antecedent positions. However, subsequent research has raised empirical and theoretical challenges for Kazanina et al.’s hypothesis (Drummer & Felser, 2019; Patterson & Felser, 2019). Our goal in this present paper is to provide a novel experimental test of the

* Corresponding author at: Department of Language Studies, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON M1C 1A4, Canada.

E-mail addresses: dave.kush@utoronto.ca (D. Kush), brian@linguist.umass.edu (B. Dillon).

<https://doi.org/10.1016/j.jml.2021.104254>

Received 26 August 2020; Received in revised form 25 April 2021; Accepted 27 April 2021

Available online 8 May 2021

0749-596X/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

broader early filter hypothesis. In doing so, we develop a test of some key predictions and address some of the limitations of the previous studies. To preview our results, in two self-paced reading studies in English we find support for the claim that grammatical knowledge constrains cataphoric search.

Our paper is structured as follows. We first review some of the existing evidence that comprehenders engage a forward search for a cataphor's antecedent. We then review key studies that investigated the role of grammatical constraints on pronominal reference in active search, before turning to our experimental design and studies. In the General Discussion we sketch a model of how the parser achieves grammatical sensitivity by integrating syntactic and discourse information with a general preference to resolve anaphoric dependencies as quickly as possible.

Active search and cataphoric pronouns

Some early evidence for active cataphoric search comes from Van Gompel & Liversedge (2003). Van Gompel and Liversedge investigated the processing of cataphoric pronouns in fronted adverbial clauses in examples like (2). When reading an example like (2) in an out-of-the-blue context, a comprehender encounters the cataphor *he* and is unable to link the pronoun to an antecedent in the context.¹ To test whether comprehenders eagerly posit an antecedent later in a sentence, Van Gompel and Liversedge manipulated the gender of the main clause subject, such that it either matched the gender features of the cataphor (e.g. *the boy*) or did not (e.g. *the girl*). Van Gompel and Liversedge reasoned that if comprehenders anticipated a referent for the cataphor in the matrix subject position, they would expect to find a slowdown in reading times when the gender features of the matrix subject were incompatible with the cataphor. Such a slowdown - a *gender-mismatch effect (GMME)* - indexes the comprehender's surprise at finding a non-antecedent where they expected to encounter the antecedent.

(2) When *he* was at the party, *the boy/girl* visited ...

Across two eye-tracking experiments, Van Gompel and Liversedge observed a GMME in early reading time measures (*first pass* and *first pass regressions out*) at the spillover region *visited*. In a third eye-tracking experiment, the researchers showed that mismatch effects also occur when the cataphor and main subject mismatch in number features, suggesting that the effects are not driven by gender alone.

Van Gompel and Liversedge took the GMME as evidence for an active search process to resolve the cataphor's reference. They argued that comprehenders entertained the possibility of coreference between the cataphor and the NP in matrix subject position before the gender or number features of the subject had been fully processed, which in turn suggests that comprehenders consider coreference between the cataphor and the subject at a very early stage in processing. Subsequent studies that deployed the same GMME logic revealed similar effects: After encountering a cataphoric pronoun, readers are generally surprised to encounter a referent that mismatches the pronoun's gender features, implying that they are incrementally anticipating coreference and experience disruption when that expectation is foiled (Giskes & Kush, to appear; Yoshida et al., 2014).

Another source of evidence for active search comes from studies showing that antecedent search can influence participants' decisions for how to resolve syntactic ambiguities. Cowart and Cairns (1987) found that when attempting to resolve a cataphor, readers prefer to analyze ambiguous strings like *frying eggs* in (4) as NPs (which can supply a referent) rather than VPs (which cannot). Interestingly, the parser seems

to establish a cataphoric dependency even in examples like (4b) where the resulting interpretation is implausible. Cowart and Cairns' results imply that comprehenders predictively commit to placing the antecedent of the cataphor *they* in main subject position, before they have the syntactically ambiguous subject phrase.

(4) a. Even though *they* use very little oil, *frying eggs* ... (Cowart & Cairns, 1987; p. 321)

b. Even though *they* eat very little oil, *frying eggs* ...

Ackerman (2015) finds similar results in a series of eye-tracking-while-reading experiments: When presented with a temporary syntactic ambiguity, comprehenders in her experiments appear to adopt an analysis that allows them to complete the cataphoric dependency sooner, even when that analysis requires an infrequent (and therefore dispreferred) subcategorization frame for a verb.

Constraints on coreference and active search

The observation that cataphors search for their antecedent has led researchers to ask if the search proceeds unconstrained, or if it is guided by grammatical constraints on coreference.

Existing studies have focused on how one particular constraint, Principle C of the Binding Theory (Chomsky 1981), interacts with the forward search for a cataphor's antecedent. Principle C was originally formulated as a syntactic constraint that banned coreference between a pronoun and an R-expression (e.g. a name, or a lexical NP) that it precedes and c-commands,² as in (5). Example (6) shows that coreference is allowed when the pronoun does not c-command the name, because the relationship between the pronoun and the R-expression is no longer in violation of Principle C. Throughout, we use subscripts to indicate intended patterns of coreference, and asterisks to denote the impossibility of a given interpretation.

(5) *He*_{*i/j} thought that people were afraid of *Loki*_i.

(6) *His*_{i/j} owner thought that people were afraid of *Loki*_i.

Subsequent research has re-interpreted Principle C effects as reflecting semantic and pragmatic constraints on co-reference (Büring 2005; Grodzinsky & Reinhart, 1993; Johnson, 2013; Levinson, 1987; Marty, 2017; Reinhart & Reuland, 1993; Schlenker, 2005), but the general notion that the relative positions of a pronoun and an R-expression are relevant for determining their ability to co-refer remains undeniable. Following others (e.g. Drummer & Felser 2018), we use 'Principle C' as a convenient stand-in for whatever formulation of the constraint(s) should ultimately be adopted.

Kazanina and colleagues (2007) observed that if cataphoric search is constrained by Principle C, then comprehenders should *not* expect an antecedent for a cataphor in any position in the sentence c-commanded by the pronoun. That is, comprehenders should not entertain coreference between *he* and *Loki* in examples like (5) at any stage in processing. To test whether Principle C influences cataphoric search, Kazanina and colleagues investigated the processing of sentences like (7). Like Van Gompel and Liversedge, they varied gender-match between a pronoun and a subsequent noun (*quarterback*). They also varied whether coreference between the pronoun and the noun was blocked by Principle C. In (7a) coreference is blocked by Principle C because the subject pronouns *he/she* c-command *quarterback*. We refer to constructions that block coreference between a cataphor and the following NP as *Constraint* contexts (following Kazanina and colleagues). In contrast, coreference is possible in (7b) because the possessive pronouns *his/her* do not c-

¹ Filik and Sanford (2008) show that comprehenders do immediately try to find an antecedent for a pronoun in a fronted adverbial clause in the preceding context, if possible (cf. Gordon, 1997).

² C-command can be defined in a number of ways. Informally, A c-commands B if B is, or is contained inside, A's sister in the syntactic tree (see Reinhart 1983).

command *quarterback* from inside the subject NP. Correspondingly, we refer to this as a *No Constraint* context.

(7) a. Constraint:

He/She chatted amiably with some fans while the talented young quarterback ...

b. No Constraint:

His/Her managers chatted amiably with some fans while the talented young quarterback...

The authors found a GMME at *quarterback* in *No Constraint* contexts (7b), but not in *Constraint* contexts (7a). The results suggest that comprehenders considered the non-c-commanded NP in (7b) a potential antecedent for the cataphor, but not (7a), thereby providing some evidence that knowledge of Principle C guides the incremental resolution of cataphoric pronouns (see Ackerman, 2015; Kazanina & Phillips, 2010; Pablos, Doetjes, Ruijgrok, & Cheng, 2015 for similar conclusions). In response to their findings, Kazanina and colleagues concluded that 'syntactic constraints immediately restrict active search processes.'

The effect does not seem to be limited to reading studies. Clackson and Clahsen (2011; discussed in Drummer & Felser, 2019) used the visual world paradigm to test whether adults and children anticipated coreference between a cataphor and the main subject in contexts structurally similar to (7). Adult looking data suggests that participants never entertained coreference between a cataphor and a subsequent NP if Principle C blocked the relation.

Although the above studies appear to support the early filter hypothesis, some subsequent work has raised empirical and theoretical challenges. One empirical challenge regards the strength of existing experimental evidence (Drummer & Felser, 2019). For example, Kazanina et al.'s results offer somewhat mixed evidence for the claim that Principle C limits early forward search. Kazanina and colleagues found a GMME effect on the first word of the potential antecedent only in one of three experiments (Experiment 1), but in this experiment the interaction between constraint and gender match was significant only in the by-participants ANOVA. This pattern suggests that the experiment may have been underpowered to detect the critical effect, which in turn indicates that caution is warranted before concluding that this is a replicable finding (Jäger, Mertzen, Van Dyke & Vasishth, 2020; Mertzen, Laurinavichyute, Dillon & Vasishth, 2020; Vasishth, Mertzen, Jäger & Gelman, 2018; see Drummer & Felser, 2019, for discussion).

Drummer and Felser (2019) also raise an important conceptual challenge to a strong interpretation of Kazanina and colleague's findings. They note that the relevance of Principle C for ruling out coreference between a cataphor and an item it c-commands depends on the *form* of that item. If the item is an R-expression, coreference is unacceptable (8a). If the item is a pronoun or expressive epithet as in (8b,c), however, coreference is allowed.

(8) a. R-EXPRESSION: He_i chatted with some fans while the quarterback_i ...

b. PRONOUN: He_i chatted with some fans while he_i ...

c. EPITHET: He_i chatted with some fans while the conceited jerk_i ...

Given that the possibility of coreference between a cataphor and c-commanded position depends on the form of the item in that position, a strategic parser should not predictively exclude coreference with the position itself. Comprehenders could entertain coreference with the upcoming position until they find an R-expression, and only then rule out coreference. The results of Experiments 2 and 3 from Kazanina et al. (2007) are potentially consistent with late exclusion, as GMME effects were observed relatively late. For example, in Experiment 3 (example 6 above), the GMME was observed at the head noun *quarterback*, three words downstream from the definite article that announces the subject NP. It is possible that comprehenders first entertained coreference between the cataphor and the NP, only to rule out coreference before

reaching the head noun *quarterback*. Later studies arguing for immediate sensitivity to Principle C (Kazanina & Phillips, 2010; Pablos et al., 2015) have similarly failed to provide clear evidence of early effects.

To test the hypothesis that there is an early stage of processing where comprehenders entertain coreference between a cataphor and a grammatically illicit NP, Drummer & Felser (2019) monitored L1 and L2 German participants' eye-movements as they read the German equivalents of the sentences in (9), where gender-match and c-command relation between the pronoun *he/his* and the name *Daniel/Annika* were manipulated.

(9) a. Constraint:

He fed the animals, as **Daniel/Annika** a loud noise heard ...

b. No Constraint:

His friend fed the animals, as **Daniel/Annika** a loud noise heard ...

Drummer and Felser found that native German participants showed GMMEs at the name *Daniel/Annika* in early reading measures (first-fixation and first-pass times), irrespective of whether Principle C allowed coreference between the pronoun and the name (see also Patterson & Felser 2019). The researchers interpreted their results as evidence that Principle C operates as a *late filter* on cataphor resolution: The parser first posits coreference with the nearest noun-phrase in the linear string and subsequently filters out dependencies that violate grammatical constraints.³

Reassessing the role of grammatical constraints on forward search

Stepping back, we see that studies using Principle C provide mixed empirical support for the early filter hypothesis. The mixed results could imply that grammatical constraints do not strongly guide active search, but they could just as well indicate that Principle C does not provide a clear test of the early filter hypothesis. Since the application of Principle C is form-dependent, it is unclear whether the parser should rule out predicted coreference from the earliest stages of processing in experimental stimuli like those tested.

A stronger test of the early filter hypothesis requires a constraint that definitively precludes coreference between a cataphor and a specific position. We provide such a test by investigating if knowledge of Binding Principle B (Büring, 2005; Chomsky 1981; Reinhart & Reuland 1993) constrains forward search in cataphoric processing. Principle B blocks coreference between a subject and a pronoun that are co-arguments of the same predicate, as in (10).

(10) a. Loki_i scratched him_j.

b. People were worried [after Loki_i scratched him_j].

Previous studies have investigated how Principle B influences the resolution of anaphoric pronouns. Overall, these studies have concluded that the constraint is used in the earliest stages of antecedent retrieval (e.g. Badecker & Straub 2002; Clackson, Felser & Clahsen, 2011; Clifton, Kennison & Albrecht, 1997; Cummings & Sturt, 2018; Chow, Lewis & Phillips, 2014; Sturt, 2013). However, there remains a debate on whether only Principle B-compliant antecedents are considered at the point of retrieval (e.g. Chow et al., 2014; Nicol & Swinney, 1989), or if Principle B is deployed as one constraint on antecedent selection alongside other constraints like appropriate gender features (Badecker & Straub, 2002; Cummings & Sturt, 2018). No studies have, to our knowledge, tested how the constraint interacts with forward-looking antecedent search in cataphor resolution. Probing prospective application of

³ This late time course, they suggest, is consistent with the view of Principle C as a semantic or pragmatic filter on coreference (see, e.g., Schlenker, 2005; Marty, 2017).

the binding principles permits a test of constraint sensitivity free from the potentially confounding effects of noisy memory retrieval (Kazanina et al., 2007).

To investigate whether Principle B is deployed as an early filter on cataphor resolution, we tested whether comprehenders attempt to link an object pronoun in a fronted participial clause with the main subject. That is, we tested whether they fleetingly consider coreference between *him* and *Loki* in sentences like (11), even though such coreference is ruled out by the grammar.

- (11) After scratching *him*_{*j/k}, *Loki*_j growled at Jorge_k.

To illustrate why coreference between the pronoun and main subject is ruled out we first note that the non-finite verb *scratching* in the fronted clause has an implicit subject ('the scratcher'). Following Chomsky (1981) we represent the implicit subject as the (null) pronoun PRO as in (12a).⁴

- (12) After PRO_j scratching *him*_{*j/k}, ...

Because PRO and *him* are co-arguments of the same predicate, Principle B makes coreference impossible. As a consequence, *him* cannot refer to any other NP that PRO co-refers with later in the sentence. This consequence is important, because *Control Theory* (Chomsky, 1981) forces PRO to be co-interpreted with the main subject. As shown in (13), the referent of PRO is obligatorily interpreted as the main subject, whether the adjunct is fronted or not (see Gerard, Lidz, Zuckerman & Pinto, 2018; Kwon & Sturt, 2015, for experimental data on the interpretation of PRO in adjunct control constructions).

- (13) a. *Loki*_j growled [after PRO_j/_{*k} scratching *him*].
b. [After PRO_j/_{*k} scratching *him*], *Loki*_j growled.

With these facts in hand, we have the foundation for our test of Principle B in cataphor processing. Co-reference between the cataphor in object position in sentences like (11) is ruled out transitively via the interaction of two grammatical requirements: (i) that PRO and a cataphor co-argument be disjoint in reference, and (ii) that PRO and the main subject be co-referential.

Important for our purposes, the block on coreference is not form-dependent in cases like (14). The object pronoun must not corefer with the matrix subject no matter what form the subject takes, R-expression, pronoun, or epithet:

- (14) a. R-EXPRESSION: After PRO_i scratching *him*_{*k/j}, *Loki*_i barked at Jorge_k.
b. PRONOUN: After PRO_i scratching *him*_{*k/j}, *he*_i barked at Jorge_k.
c. EPITHET: After PRO_i scratching *him*_{*k/j}, *that miserable mutt*_i barked at Jorge_k.

Thus, unlike the Principle C contexts tested in previous studies, the parser has no reason to consider coreference between the cataphor inside the fronted adjunct and the matrix subject at any point in processing. A parser that uses grammatical constraints as early filters is therefore licensed to ignore the main subject entirely when looking for an antecedent for a direct object cataphor in a fronted non-finite adjunct clause. We conducted two self-paced reading experiments to determine whether the parser actually does so.

Experiment 1

Experiment 1 tested whether comprehenders can use the Principle B

to constrain the forward search for a cataphor's antecedent by comparing the processing of sentences like (15) and (16):

- (15) Constraint:

While / driving / {*him* | *her*} / to school / on Friday, / ...

- (16) No Constraint:

While / driving / {*his* | *her*} daughter / to school / on Friday, / ...
... Christopher / casually / told / {Juan|Hannah} / that / he / would / pick up / everyone / early / for / a surprise.

The *Constraint* examples in (15) block coreference between the cataphor *him/her* and the matrix subject due to the interaction of Control and Principle B. In *No Constraint* control conditions (16) we modified the pronoun to be a possessor, rather than a direct object. When the pronoun is possessive, coreference is allowed with PRO. We reasoned that comprehenders would preferentially interpret the possessive as co-referent with PRO within the adjunct, given that past work has shown that comprehenders actively link adjunct-internal pronouns and PRO when possible (Kreiner, Sturt & Garrod, 2008). The pronoun should thus be co-interpreted with the main subject, as a consequence of Control.

In both sentence types, we manipulated whether the gender features of the pronoun matched the matrix subject. Following previous studies, we expect a GMME in *No Constraint* conditions as an indication of active search. If knowledge of Principle B acts as an early filter on search, then we do not expect a similar GMME in the *Constraint* conditions. Conversely, if Principle B does not guide search, then we expect a GMME in the *Constraint* conditions. To ensure that there was always a potential within-sentence antecedent for the cataphor, the direct object of the matrix verb was always a name that matched the gender features of the pronoun (e.g. *Juan* or *Hannah* in 17).

Participants

83 self-reported native English-speaking participants were recruited through the *Prolific Academic* online platform. Of these, 16 were excluded prior to statistical analysis due to answers on their debriefing questions and to performance on comprehension questions. Participants provided informed consent and were compensated at a rate roughly equivalent to 9GBP/hr.

Materials

We created 24 test items following the format in (15) and (16). The critical experimental items contained a fronted participial adjunct clause containing a PRO and a pronoun. We implemented a 2 × 2 within-subjects crossed factorial design with two factors: Match, which controlled whether the pronoun *matched* or *mismatched* the main clause subject (e.g., *Cristopher*) in gender, and Constraint, which controlled whether the pronoun could grammatically co-refer with the subject. In *Constraint* sentences, the pronoun was the direct object of the infinitival verb. In these conditions the pronoun is a cataphor, because there is no acceptable antecedent for the pronoun earlier in the sentence. In *No Constraint* sentences, the pronoun was a possessor embedded inside the NP direct object.

Each item was divided into presentation regions that were either one or two words: prosodically weak words such as definite determiners or short prepositions were combined with the following word. In the *No Constraint* conditions, the possessive cataphor was presented together with the following noun in order to ensure that the same number of presentation regions intervened between the cataphor and the antecedent in both the *Constraint* and *No Constraint* conditions. Across items, the cataphor always occupied the third presentation region, and the matrix subject the sixth region. Following the main clause subject, there was a spillover region that contained an adverb (*casually* in 17). Across items, three different subordinating expressions were used (*before*, *while*, *after*).

⁴ We adopt the PRO analysis for ease of exposition, but our experimental logic does not rest on this assumption.

These 24 critical items were combined with 56 grammatical filler sentences of comparable overall complexity and length. None of the fillers contained cataphors, but they did contain various syntactic ambiguities, long-distance dependencies, and anaphoric pronouns.

Method

The experiment was deployed on the IxetFarm web-based experimental presentation platform (Drummond, 2013). Participants performed the experiment remotely on their own computers via a link distributed via Prolific Academic.

Test sentences were presented in a self-paced phrase-by-phrase manner (phrase boundaries are shown in the example sentences above). Each phrase was center-aligned and presented non-cumulatively. Participants pressed the spacebar to move to the next region, and reading times were measured on each phrase. Each trial began with a fixation cross.

The experimental items were distributed into four Latin Square lists, and each was combined with the same set of experimental fillers. Participants were randomly assigned to a list. Upon loading the experiment, participants were presented with informed consent and instructions. Participants had to answer comprehension questions about the experiment instructions before they could proceed to the experiment. They were instructed to minimize distractions, read sentences at a natural pace, and to respond to the comprehension questions as accurately as possible. Participants were given a self-timed break after every 12 sentences.

All sentences in the experiment were followed by a comprehension question. Comprehension questions were accompanied by two possible answers from which the participants had to choose using the *f* and *j* keys to choose the answer on the left or right, respectively. The use of the mouse or trackpad was disabled during the experiment. Participants received feedback if they answered incorrectly. Comprehension questions on filler trials asked about various aspects of the previous sentence, including objects mentioned (e.g., *What did Pauline celebrate with? Donuts – Champagne*), locations (e.g., *Where were the prisoners taken? The town square – The Jail*), event duration (*How long was the trip? One day – Three days*), and the interpretation of potential ambiguities. We provide more detail about the questions following test items and report participants' performance on them separately below.

Following the experiment, participants were given an open-ended debrief so they could supply qualitative feedback about the experiment. One debriefing question asked if they had any problems during this experiment. One participant said that their cat pestered them throughout the experiment. Another participant indicated that they had a reading disorder. These two participants were excluded prior to statistical analysis.

Another debriefing prompt elicited open-ended, imaginative responses so as to identify bots and increase data validity (Chmielewski & Kucker, 2020). The prompt read: *Imagine you drove from your house to the nearest shopping mall. Describe the most boring and the most interesting thing you would see on the way.* Responses with language that seemed bot-generated or non-native-like were rejected. Five responses met these criteria (e.g. *none*, or *people the road*). These five participants were excluded prior to statistical analysis. Notably, four of the five rejected participants were also rejected for chance performance on the comprehension questions.

Analysis

Prior to analysis we set 75% accuracy on comprehension questions as a minimum threshold for inclusion. 13 participants failed to meet this threshold and were excluded. Combined with the other exclusions described above, data from 67 participants remained for analysis. Furthermore, reaction times of less than 100 ms, or greater than 3000 ms, were trimmed from the dataset prior to statistical analysis. This led

to the loss of 496 data points (0.7% of the overall data).

We adopted a Bayesian approach to the statistical analysis of our data. In adopting a Bayesian approach, our main goal was to estimate both the magnitude and the probability of a GMME in *Constraint* and *No Constraint* contexts. This inferential approach stands in contrast to null hypothesis significance testing, which asks the dichotomous question of whether we do or do not have evidence for an effect in our data (Nicenboim & Vasishth, 2016). For our experiments, we fit Bayesian linear mixed effects regression models using the brms package (Bürkner, 2017), which is a front end to the Stan language for Bayesian estimation of model parameters (Gelman, Lee & Guo, 2015). All analyses were conducted in the R statistical computing environment (R Core Team, 2013). For each region of interest, we fit two models: a crossed model and a nested model. In the crossed model, each experimental factor was sum-coded, and the fixed effects specification of the model included both of these main effects and their interaction. The interaction term in crossed model tests whether the GMME effect interacts with syntactic construction (i.e. the *Constraint* factor). In addition, we fit a nested model, which had separate fixed effects parameters for the GMME within the *Constraint* conditions, and within the *No Constraint* conditions. Table 1 provides the contrast coding for all fixed effects in both models, for all experiments.

For each model, participants and items were treated as random grouping factors. We implemented 'maximal' mixed-effects regression models (Barr et al., 2013): The regression model contained random intercepts and random slopes for all fixed-effects predictors. We modeled RT in milliseconds as our dependent variable using generalized linear models. RT data are characteristically rightward skewed, which means that untransformed RT data violate the assumption of normality inherent in linear regression models. Generalized linear models address this shortcoming by adopting a link function that specifies how the regression equation is related to the data's distribution. For RT data, one common choice is the log-normal link function, which is equivalent to analyzing log-transformed RT data with a linear model. For our analyses, we opted to use a shifted log-normal link function (Nicenboim et al., 2018), which is a log-normal distribution that is offset by a constant value. The decision to use a shifted log-normal, rather than an unshifted log-normal distribution, was based on a visual inspection of how well each model's posterior predictive distribution matched the overall RT distributions in the data. The shifted log-normal distribution yielded a posterior predictive distribution that matched the major distributional features of the experimental data. There are also good theoretical reasons to favor shifted log-normal distributions for reaction time data: See Rouder and Lu (2005) and Nicenboim et al. (2018) for a more in-depth discussion.

We set normal priors over all fixed effects and the intercept. All priors had a mean value of 0; the variance on the prior distribution was set to 1 for all fixed effects, and 10 for the intercept. These are mildly uninformative priors that do not place strong *a priori* constraints on the model's predictions, and incorporate very little knowledge about what makes (e.g.) a plausible RT distribution. The prior on the random effects correlation matrix was an LKJ prior with $\eta = 2$. This is known as a *regularizing* prior, because this setting for the η hyperparameter assigns lower *a priori* plausibility to large correlation values (e.g. +1 or -1). In the context of our experimental data, this prior encodes an *a priori* belief that it is relatively unlikely that participant reading times will be strongly correlated with their susceptibility to the experimental manipulation. Such regularizing priors are recommended for complex Bayesian models (Vasishth, Nicenboim, Beckman, Li & Kong, 2018). For each model, we ran four Monte Carlo Markov Chains in parallel, with 6500 samples each. The first 3250 samples were always discarded as part of the model 'warmup' period, leaving a total of 13,000 post-warmup samples altogether for each model. For all models reported below, the potential scale reduction factor (*R-hat*) statistic was at or near 1.0 for all fixed effects parameters of interest. This value indicates that there was little between-chain variance, which in turn indicates

Table 1

Table of contrast coding coefficients used for the crossed and nested statistical models described in the text. Experimental conditions are given in rows, model contrasts given in columns.

	Crossed model			Nested model		
	Constraint	Match	Constraint:Match	Constraint	Match: Constraint	Match: No Constraint
Constraint, Match	0.5	−0.5	−0.25	0.5	−0.5	0
Constraint, No match	0.5	0.5	0.25	0.5	0.5	0
No constraint, Match	−0.5	−0.5	0.25	−0.5	0	−0.5
No constraint, Mismatch	−0.5	0.5	−0.25	−0.5	0	0.5

satisfactory convergence of the posterior estimates across chains. No divergences were observed.

We modeled RTs at two regions of interest: critical and spillover regions. The critical region was the matrix subject position (e.g. *Christopher*), and the spillover region was the adverb that followed this region (e.g. *casually*).

Results

Comprehension questions. Experimental comprehension questions were split into three groups of 8 questions. Each group probed a different aspect of how participants interpreted the test sentences. 8 comprehension questions targeted the interpretation of PRO in the fronted adjunct (e.g. *Who drove someone to school on Friday? Christopher – Juan/Hannah* for example 17). The next 8 questions directly targeted participants' interpretation of the cataphor (e.g. *Who was driven to the school on Friday?*). The interpretation of the cataphor is ambiguous in the *No Constraint Match* condition - either the main subject or object are grammatical potential antecedents. In all other conditions the matrix subject is not a grammatical antecedent: in *Mismatch* conditions coreference is blocked because of gender-mismatch; in the *Constraint Match* condition coreference is ruled out by Principle B. The last 8 comprehension questions targeted the argument roles in the main clause (e.g. *Who was told something?* for example 17). Performance on the three types of comprehension questions for the experimental items is summarized in Table 2.

Responses indicated that participants largely interpreted the fronted adjunct as anticipated by control theory, treating the implicit PRO subject of adjunct as coreferent with the overt matrix subject. The one exception was in the *No Constraint Mismatch* condition, where participants only offered this interpretation 66% of the time. A logistic linear mixed effects model fit to these response data revealed a main effect of *Constraint* (more subject control responses in *Constraint* conditions; $z = 2.2$), and an interaction of *Constraint* and *Match* ($z = 2.8$).

On the questions that targeted the cataphor's interpretation, there was more coreference between the cataphor and the matrix subject in the *No Constraint* conditions ($z = 4.3$), and more coreference when the cataphor and the matrix subject matched in gender features ($z = 4.8$). The interaction was not reliable. On the questions that targeted the comprehension of the matrix clause, we saw generally high

comprehension performance. Still, comprehension was less accurate in the *Match* conditions ($z = 3.5$).

Self-paced reading results. Average region-by-region raw RTs are plotted in Fig. 1. Tables 3 and 4 summarize the mean of the posterior distribution over all experimental fixed-effect parameters of interest in the critical region, along with the 95% highest posterior density interval (HPDI). These Bayesian credible intervals indicate where the most plausible parameter values for these fixed effects parameters lie, given the data.

The analysis reveals evidence that RTs were overall slower in the *No constraint* conditions than in the *Constraint* conditions ($\Pr(\beta < 0) = .99$), but somewhat unclear evidence for a main effect of *Match* ($\Pr(\beta > 0) = .85$). The crossed model revealed evidence for a *Constraint* \times *Match* interaction in the predicted direction ($\Pr(\beta < 0) = .99$). The nested model gives further insight into the source of this interaction. There was evidence for a GMME in the *No Constraint* conditions ($\Pr(\beta > 0) = .98$), but not for a GMME in the *Constraint* conditions ($\Pr(\beta > 0) = .19$).

Turning to the spillover region, we again observed that *No constraint* conditions were read more slowly than the *Constraint* conditions ($\Pr(\beta < 0) > .999$). In the crossed model, there was evidence for a main effect of *Match* ($\Pr(\beta > 0) = 0.98$), but this was qualified by clear evidence that these two factors interacted ($\Pr(\beta < 0) = 0.99$). In the nested model, we saw clear evidence for a GMME effect in the *No Constraint* conditions ($\Pr(\beta > 0) = 0.99$). In the *Constraint* conditions, however, the probability of there being a GMME in the same direction was very low: Only 5% of posterior samples for this parameter revealed a GMME in the predicted direction ($\Pr(\beta > 0) = .05$).

To aid in interpretability, we back-transformed parameter estimates from the nested model to milliseconds by calculating the predicted log-RT in the *Match* and *Mismatch* conditions, exponentiating these estimates to yield an estimated reading time in milliseconds, and taking the difference between these estimates. Figs. 2 and 3 present the marginal posterior distribution over the resulting GMME effect in milliseconds, for *Constraint* and *No Constraint* conditions. The model for the critical subject region estimates an average GMME of +44 ms in the *No Constraint* conditions. The mean posterior estimate of this same effect in the *Constraint* conditions is roughly −14 ms. This estimated difference is in the opposite direction of the predicted GMME, and the HPDI overlaps with zero. A similar pattern is seen in the spillover.

Discussion

Qualitatively, the reading-time results suggest that participants exhibit a pronounced GMME effect when the cataphor was a possessive (*No Constraint* conditions), but not when it was a direct object pronoun (*Constraint* conditions). The reading time data thus suggest that search is constrained as predicted by the early filter hypothesis: The incremental reading-time record provides no evidence that comprehenders entertained coreference between a fronted direct object pronoun and the matrix subject in violation of Principle B.

Turning to the offline question responses, we see a more complex pattern, which might initially seem at odds with the self-paced reading results: In the subset of questions that probed whether participants interpreted the cataphor and the matrix subject as coreferent, participants chose responses consistent with coreference more often in *Match*

Table 2

Response data on the three types of questions for experimental sentences across the experiment.

	Interpretation of PRO: % Matrix subject control of adjunct clause	Cataphor interpretation: % Matrix subject interpreted as antecedent	Main clause arguments: % Error
Constraint, Match	81%	25%	13%
Constraint, Mismatch	88%	13%	4%
No constraint, Match	84%	77%	21%
No constraint, Mismatch	66%	41%	6%

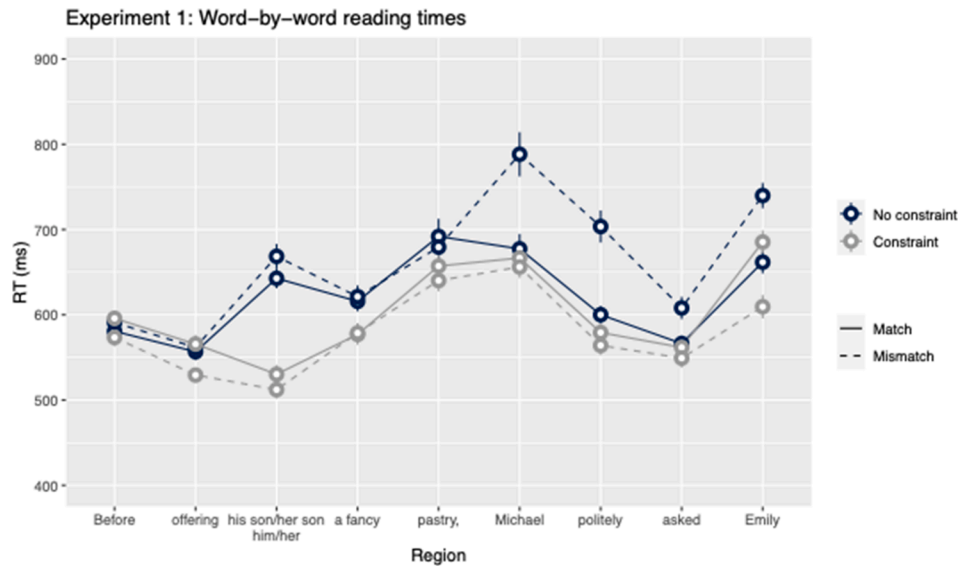


Fig. 1. Average raw reading times for the first 9 regions of test sentences in Experiment 1. Error bars represent standard error of the mean corrected for between participant variance (Bakeman & McArthur, 1996).

Table 3

. Evidence for an interaction of constraint and GMME in target region. Mean and 95% HPDIs for experimental and fixed effects in the nested and spillover models, for the critical region.

Crossed model	Mean	95% HPDI	Nested model	Mean	95% HPDI
<i>Constraint</i>	−0.08	[−.14, −.03]	<i>Constraint</i>	−0.08	[−.14, −.03]
<i>Match</i>	0.03	[−.02, .08]	<i>Match:</i>	−0.03	[−.10, .04]
<i>Constraint × Match</i>	−0.12	[−.22, −.02]	<i>Constraint</i>		
			<i>Match: No constraint</i>	0.09	[.01, .17]

Table 4

. Clear evidence for an interaction of constraint and GMME in spillover. Mean and 95% HPDIs for experimental and fixed effects in the nested and spillover models, for the spillover region.

Crossed model	Mean	95% HPDI	Nested model	Mean	95% HPDI
<i>Constraint</i>	−0.12	[−.16, −.07]	<i>Constraint</i>	−0.12	[−.16, −.07]
<i>Match</i>	0.05	[.00, .09]	<i>Match:</i>	−0.05	[−.12, .01]
<i>Constraint × Match</i>	−0.21	[−.32, −.09]	<i>Constraint</i>		
			<i>Match: No constraint</i>	0.15	[.07, .23]

conditions in both *Constraint* and *No Constraint* conditions. Comprehenders answered as if they considered coreference between the cataphor and the matrix subject 25% of the time in the *Constraint Match* condition, even though coreference is predicted to be grammatically impossible. One possible interpretation is that these responses reflect a ‘lingering misinterpretation’ that results from earlier parser error. Under this interpretation comprehenders initially co-interpret the cataphor and main subject in the *Constraint Match* condition and the erroneous initial interpretation occasionally ‘lingers’ in memory, even if the parser later revises its analysis (Christianson, Hollingworth, Halliwell & Ferreira, 2001; Slattery, Sturt, Christianson, Yoshida & Ferreira, 2013). This lingering misinterpretation explanation would be at odds with the early filter hypothesis. However, we note that a similar pattern occurs even in questions that do not directly probe the cataphor’s interpretation: When the comprehension questions queried the subject of the main clause, we again saw more errors in *Match* conditions than *Mismatch* conditions.

This pattern suggests that the increased error rate might not be directly the result of misinterpretation of the anaphor, but instead of more general similarity-based interference processes that might lead to confusion about who did what to whom at the point of answering the question. We take up this issue in greater depth in the discussion.

We note that 8 out of 24 of the critical items in our experiment directly probed the interpretation of the cataphor. This is a departure from Kazanina et al (2007) and Drummer and Felsner (2019) who asked *yes/no* comprehension questions that did not directly probe the antecedent of the pronoun. This raises the possibility that our results here reflect a strategic adaptation to this feature of the experimental set-up. We address this in Experiment 2 by reducing the number of questions that probe the interpretation of the pronoun.

Finally, we acknowledge an additional, unexpected effect in the question response data: Participants responded as if PRO was not interpreted as bound by the matrix subject on a non-trivial portion of trials, in apparent violation of obligatory subject control. The pattern is most apparent in the *No Constraint, Mismatch* condition, where participants chose the matrix object as the controller of PRO 34% of the time. We do not interpret these numbers as evidence that control constraints were applied inconsistently, given previous experimental reports of obligatory subject interpretation of adjunct control constructions (Gerard, Lidz, Zuckerman & Pinto, 2018; Sturt & Kwon, 2015). We suspect that the especially low accuracy on subject control questions may have arisen due to an interaction of (i) retrieval interference at question time and (ii), in the *No Constraint* conditions, reanalysis prompted by an initial preference to treat the possessive and the PRO as coreferent inside the adjunct. In any event, we do not replicate this effect in Experiment 2, and so do not interpret this pattern further.

Questions of how to interpret participants’ offline responses notwithstanding, the results of Experiment 1 are consistent with the early filter hypothesis: Participants appear not to expect coreference between a cataphor and the main subject when such a relation would violate Principle B. Although the results are consistent with the early filter hypothesis, they are also consistent with an alternative reductive hypothesis: The absence of a GMME in the *Constraint* cases could have arisen if object pronouns in fronted adjunct clauses simply do not trigger active search for an antecedent. Experiment 2 teases these two alternatives apart, while providing an attempt to replicate the key finding from Experiment 1: No GMME effect in the *Constraint* conditions.

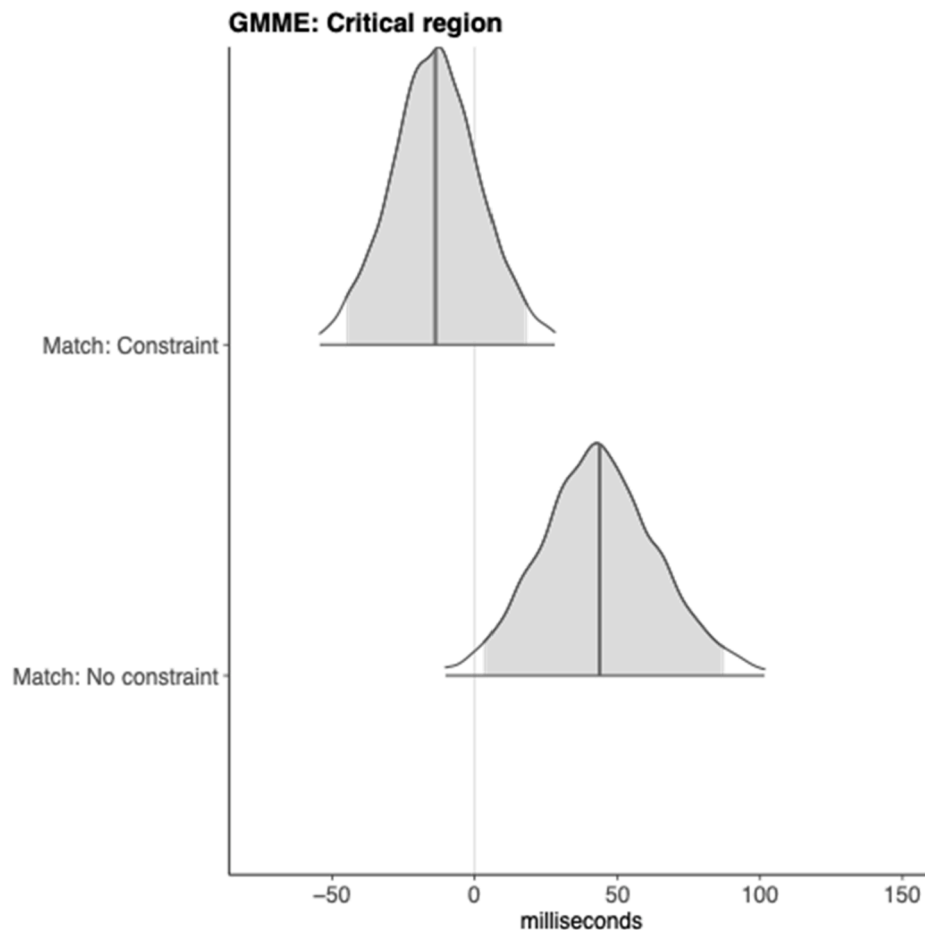


Fig. 2. Marginal posterior distributions over the gender-mismatch effect in the critical main subject region for Experiment 1, in *Constraint* and *No Constraint* conditions. Posterior distributions back-transformed to millisecond scale using the estimates in the nested model.

Experiment 2

We failed to see a GMME in the *Constraint* conditions in Experiment 1, a finding that we ascribed to the active use of grammatical constraints to filter the search for a cataphor's antecedent. But it is also possible that direct object pronouns simply do not trigger a search for an antecedent. To test this alternative hypothesis, Experiment 2 replicates Experiment 1 but with a novel *No Constraint* baseline:

(17) Constraint:

While / driving / {him | her} / to school / on Friday, / ...

No Constraint:

While / a parent / drove / {him | her} / to school / on Friday, / ...

... Christopher / casually / told / {Juan|Hannah} / that / he / would / pick up / everyone / early / for / a surprise.

In the modified *No Constraint* conditions, the fronted adjunct clause was made finite and PRO replaced with an overt indefinite subject. The nature of this indefinite subject varied across item sets. For some it was a generic indefinite noun phrase (*someone*, *anyone*), in others, it was a lexically specified indefinite NP (e.g. *a parent*). Without PRO there is no longer a control relationship between the adjunct clause and the matrix subject, and therefore no requirement that the object pronoun not corefer with the matrix subject. The early filter hypothesis therefore predicts a GMME in the *No Constraint* conditions, but not the *Constraint* conditions. The *delayed filter* hypothesis predicts a GMME in both *Constraint* and *No Constraint* pairs.

Participants

80 self-reported native English-speaking participants were recruited via *Prolific Academic*. 12 were excluded for failing to meet a 75% minimum accuracy threshold, 2 more were removed for reporting a native language other than English, and 1 further subject was removed for an unacceptable debrief response, leaving 65 participants for analysis.

Materials

As in Experiment 1, materials were constructed in 2×2 within-subjects crossed factorial design with the factors *Match* and *Constraint*. *Constraint* items were identical to Experiment 1. In *No Constraint* items, the cataphor was changed to an object pronoun, as in *Constraint* sentences. The adjunct-internal non-finite verb (*driving*) was replaced with a finite verb (*drove*) with an indefinite NP (*someone*) subject. An example item is found in (18–19) above. All other features of the experimental materials, including fillers, were identical to Experiment 1.

In Experiment 2, we changed the proportion of comprehension questions that targeted the critical cataphoric dependency. Out of 24 critical items in Experiment 2, 14 had questions that targeted argument roles in the matrix clause, 6 had questions that targeted the interpretation of PRO, and only 4 directly probed the interpretation of the cataphor. Changing the distribution of question targets in Experiment 2 has the benefit of helping to minimize the risk that the results of Experiment 1 reflect a strategic effect driven by a high proportion of comprehension questions that target the critical dependencies.

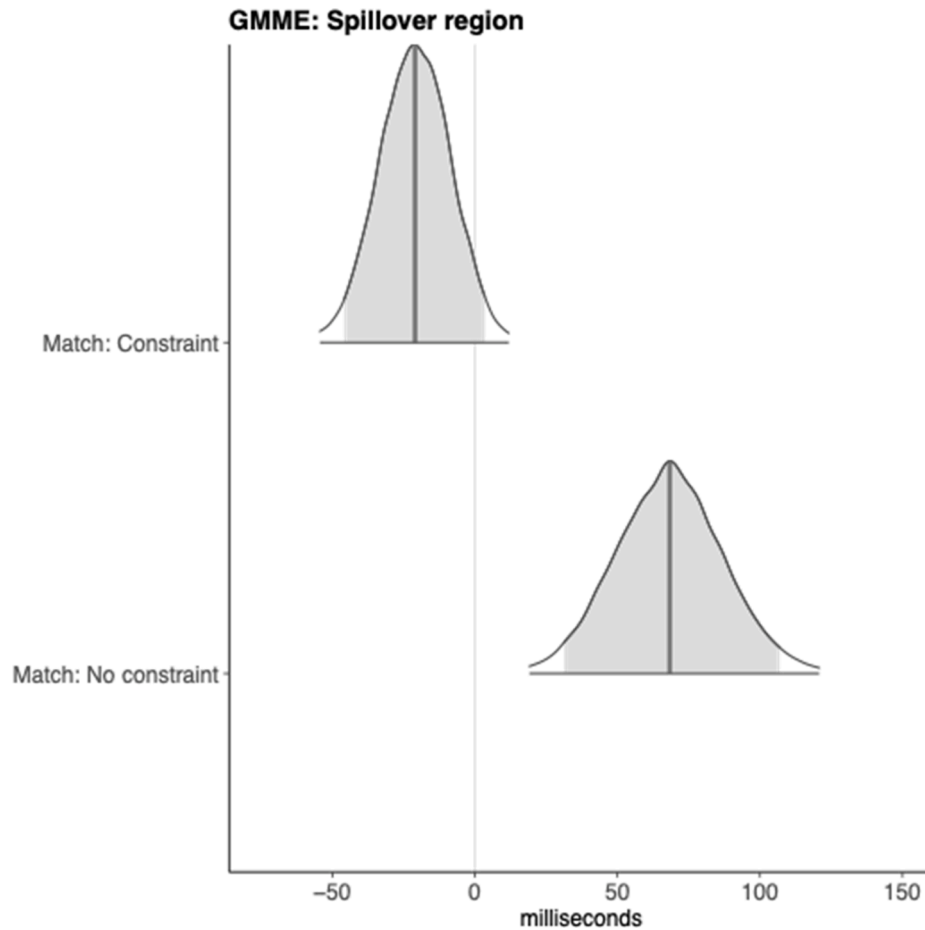


Fig. 3. Marginal posterior distributions over the gender-mismatch effect in the spillover region for Experiment 1, in *Constraint* and *No Constraint* conditions. Posterior distributions back-transformed to millisecond scale using the estimates in the nested model.

Method

The experimental method was identical to Experiment 1.

Analysis

All statistical analysis procedures, including exclusion criteria, were identical to Experiment 1.

Results & discussion

Comprehension questions. Performance on the three types of comprehension questions for the experimental items is summarized in Table 5. The types of comprehension questions asked were largely identical to Experiment 1, with the exception of the questions that

Table 5

. Response data on the three types of questions for experimental sentences across the experiment.

	Interpretation of adjunct clause: % error	Cataphor interpretation: % Matrix subject interpreted as antecedent	Main clause arguments: % Error
<i>Constraint, Match</i>	20%	28%	12%
<i>Constraint, Mismatch</i>	18%	14%	2%
<i>No constraint, Match</i>	15%	58%	10%
<i>No constraint, Mismatch</i>	15%	23%	5%

probed the interpretation of PRO. In Experiment 2, the question subset was changed so that it probed the subject of the fronted adjunct clause (e.g. *Who drove to school?* in (17) and (18)) across all conditions. In *Constraint* conditions, these probed the interpretation of PRO, and the response options were the same as in Experiment 1. In *No constraint* conditions, the response options were either the indefinite subject of the adjunct clause (e.g. *A parent* in 18), or the matrix subject (*Christopher*). We coded the matrix subject as the correct response in the *Constraint* conditions, and the indefinite subject of the adjunct clause as correct in the *No Constraint* conditions, and simply present percent error below.

As in Experiment 1, participants predominantly interpreted PRO and the main subject as coreferent: There was relatively little error in the *Constraint* conditions. There were no significant differences between conditions in response accuracy to the questions probing the interpretation of fronted adjunct clause. We did not replicate the unexpected finding of more incorrect responses in the *No Constraint Mismatch* condition seen in Experiment 1. This is consistent with the hypothesis that the higher error rate in Experiment 1 relates to the use of a possessive pronoun, but we do not speculate further on possible interpretations of that effect.

On the questions that targeted the cataphor's interpretation, we again observed more responses that indicated coreference between the cataphor and the matrix subject in the *No Constraint* conditions ($z = 2.7$), and when the cataphor and the matrix subject matched in gender features ($z = 4.0$). As in Experiment 1, the interaction was not reliable. Again, on the questions that targeted the comprehension of the matrix clause, we saw generally high comprehension performance. Still, comprehension was less accurate in the *Match* conditions ($z = 3.6$). These results replicate response behavior seen in Experiment 1, with the

exception noted above. We return to the implications of this pattern of results in the General Discussion.

Self-paced reading results. Average region-by-region raw reading times are plotted in Fig. 4. Table 6 summarizes the mean of the posterior distribution over experimental fixed-effect parameters of interest in the critical region and the 95% HPDI. Table 7 provides the same information for the spillover region.

In the critical region, there was clear evidence for a main effect of *Constraint* ($\text{Pr}(\beta < 0) = .92$), as well as a main effect of *Match* ($\text{Pr}(\beta > 0) = .91$). This main effect was qualified by a clear *Constraint* \times *Match* interaction in the predicted direction ($\text{Pr}(\beta < 0) = .99$). According to the nested model, the interaction is driven by differential GMMEs within *Constraint* and *No Constraint* pairs. There was clear evidence for a GMME in the *No Constraint* conditions ($\text{Pr}(\beta > 0) = .99$), but no clear evidence for any effect in the *Constraint* conditions ($\text{Pr}(\beta > 0) = .10$).

In the spillover region there was evidence for a main effect of *Constraint* ($\text{Pr}(\beta < 0) = .98$), and some evidence for a main effect of *Match* ($\text{Pr}(\beta > 0) = 0.98$). Once again, this main effect was qualified by a clear *Constraint* \times *Match* interaction ($\text{Pr}(\beta < 0) > .999$). The nested model shows very clear evidence for a GMME effect in the *No Constraint* conditions ($\text{Pr}(\beta > 0) > .999$). The probability of a GMME in the same direction was very low in the *Constraint* conditions, though there was evidence of an effect in the opposite direction ($\text{Pr}(\beta < 0) = .98$).

Figs. 5 and 6 plot the marginal posterior distributions over the GMME effect back-transformed to milliseconds from estimates in the nested model. The model for the critical subject region estimated an average GMME of +63 ms in the *No Constraint* conditions. In the *Constraint* conditions, the average estimate was -21 ms.

One notable feature of the results in both Experiment 1 and Experiment 2 is that we observe evidence for a small effect in the *Constraint* conditions that runs opposite to the predicted direction, and opposite to what is observed in the *No Constraint* conditions. This is potentially noteworthy. We refrain from interpreting this effect too strongly, as it was not predicted, and we did not observe very strong evidence for this effect in either experiment. Nonetheless, we note that previous work has interpreted similar effects as indications of *inhibition* (e.g. Badecker & Straub 2002) of grammatically inaccessible antecedents. We return to this ‘reverse GMME’ in the General Discussion.

General discussion

In two self-paced reading experiments, we tested whether Principle B operates as an early or late filter on active antecedent search in cataphor processing. We focused on antecedent search in constructions where Principle B definitively blocks coreference between a cataphor in a preposed adjunct clause and the main subject position. Our experiments manipulated (i) gender-match between a cataphor and the main subject and (ii) whether coreference between the two was grammatical according to Principle B. Following previous work, we used a gender-mismatch effect (GMME) as an indication that coreference was actively considered.

In both experiments, we observed a GMME at the main subject when coreference with the cataphor was grammatically permitted. The model-estimated GMME was between 44 (in Experiment 1) and 63 (in Experiment 2) milliseconds slower at a subject NP that mismatched the cataphor in gender. The effect persisted to the spillover region, resulting in an average model-estimated GMME of approximately 68 ms in both Experiment 1 and 2. In contrast, we saw no convincing GMME in contexts where Principle B blocked coreference between the cataphor and the subject in either experiment.

We conducted a quantitative Bayesian analysis to estimate the strength of our results. The analysis reveals a slightly more nuanced picture than would a traditional analysis. The probability that the ‘true’ (i.e. population-level) value of the GMME in the *No Constraint* conditions was greater than zero at the main subject was very high (98% in Experiment 1, 99% in Experiment 2) and even higher in the spillover

region (99% in Experiment 1, 100% in Experiment 2). Given our data, the probability that there is a GMME effect in the same direction in the *Constraint* conditions was very low at the critical region, though not zero: It was 19% in Experiment 1, 10% in Experiment 2. At the spillover region, these probabilities dropped to 5% and 2% in Experiment 1 and 2 respectively. Taken together, the data suggest that if there is a positive GMME effect in the *Constraint* conditions, it is both very modest and significantly smaller than the GMME effect in the *No Constraint* conditions. Indeed, if anything the data seem to suggest a ‘reverse GMME’ effect in the *Constraint* conditions, a possibility we take up in detail below.

On the now standard assumption that GMMEs index active consideration of coreference between the cataphor and a syntactic position, the reliable GMMEs seen in the *No Constraint* conditions indicate that comprehenders entertain coreference between the cataphor and the main subject position when permitted by Principle B. The absence of reliable GMMEs in *Constraint* conditions indicates that comprehenders are very unlikely to predict coreference between a fronted cataphor and the following main subject position when Principle B precludes it. Knowledge of Principle B seems to apply to early antecedent search, either outright blocking (or at least dramatically reducing the probability of) comprehenders predictively positing illicit coreference between the cataphor and main subject position.

Our results are most consistent with the hypothesis that grammatical constraints can be used as early, predictive filters on the search for the cataphor’s antecedent. NPs in grammatically excluded positions are unlikely to be considered as potential antecedents during the course of processing of backwards anaphora. The predictions of the late filter hypothesis were not supported: If comprehenders initially considered coreference with the main subject in violation of Principle B, then we should have seen comparable GMMEs in both pairs of experimental conditions.

Incremental referential processing

Our results suggest that comprehenders are able to immediately recognize the consequences of i) Principle B of the Binding Theory and ii) obligatory subject control of non-finite adjunct phrases such that they almost never entertain coreference between the cataphor and the subject in examples like (19):

(19) While driving *him* to school on Friday, **Christopher** ...

Applying the relevant constraints early in the processing of examples like (19) suggests that the parser integrates different features of the linguistic context and, arguably, different levels of linguistic representation to make active predictions about where an antecedent can show up. In the interest of better understanding cataphor processing and how different types of information interact in the process, we outline below a model of how the parser might implement Principle B as a constraint on active antecedent search.

As discussed in the introduction, the parser must use the following syntactic information to process our test sentence: The parser must recognize, upon encountering the direct object cataphor, that the pronoun is necessarily disjoint in reference with PRO, the implicit subject of the infinitival verb, based on the local c-command/co-argument relation between the two items. The parser must also recognize that disjoint reference with PRO entails disjoint reference with the upcoming main subject, given that PRO is necessarily co-interpreted with the subject. Importantly, the position of the cataphor in relation to the main subject does not rule out coreference between the two, as evidenced by the ability to cointerpret an object cataphor and main subject in *No Constraint* conditions in our Experiment 2. Blocked coreference is contingent on the presence of PRO.

The anaphoric dependency between PRO and its antecedent is a syntactically-mediated binding dependency (Chomsky, 1981), though the interpretive consequences of this dependency are reflected in the discourse representation. The dependency between a cataphor and its

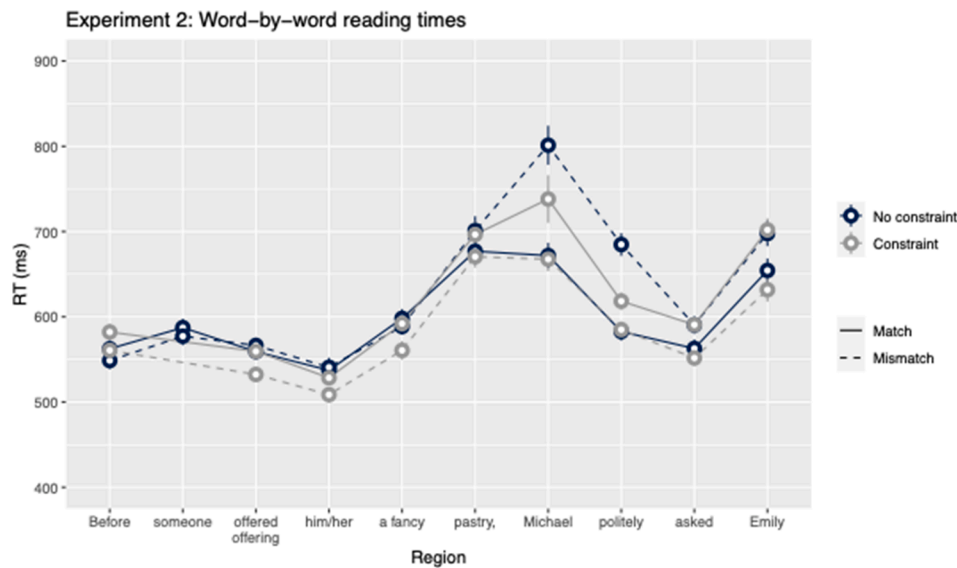


Fig. 4. Average raw reading times for the first 9 regions of test sentences in Experiment 2. Error bars represent standard error of the mean. Error bars represent standard error of the mean corrected for between participant variance (Bakeman & McArthur, 1996).

Table 6

. Clear evidence for an interaction of constraint and GMME in target region. Mean and 95% HPDIs for experimental and fixed effects in the nested and spillover models, for the critical region.

Crossed model	Mean	95% HPDI	Nested model	Mean	95% HPDI
Constraint	-0.04	[-.10, .02]	Constraint	-0.04	[-.10, .02]
Match	0.04	[-.02, .11]	Match: Constraint	-0.05	[-.12, .03]
Constraint × Match	-0.19	[-.29, -.07]	Match: No constraint	0.13	[.04, .23]

Table 7

Clear evidence for an interaction of constraint and GMME in spillover. Mean and 95% HPDIs for experimental and fixed effects in the nested and spillover models, for the spillover region.

Crossed model	Mean	95% HPDI	Nested model	Mean	95% HPDI
Constraint	-0.05	[-.10, .00]	Constraint	-0.05	[-.10, -.00]
Match	0.05	[.01, .10]	Match: Constraint	-0.07	[-.14, .00]
Constraint × Match	-0.24	[-.34, -.14]	Match: No constraint	0.17	[.10, .24]

antecedent can, on the other hand, be either binding or coreference. It is commonly assumed that coreference is represented in the discourse representation, but not the syntax (Bosch, 1983; Evans, 1980; Grodzinsky & Reinhart, 1993, a.o.), since cross-sentential coreference is possible. Thus, coreferential cataphor-antecedent dependencies are only represented at the discourse level.

Since the discourse representation is arguably the only level of representation where both binding and coreference dependencies are represented, we model active search for the antecedent of the cataphor as involving prediction at the discourse level.⁵ For concreteness, we show

⁵ We wish to point out that positing prediction at the level of discourse is compatible with simultaneous syntactic prediction (in the case of binding dependencies). In fact, we consider it likely that prediction of the PRO-antecedent relation in the discourse representation is the consequence of syntactic prediction propagating up to the higher level.

how this can be done within a simplified version of Discourse Representation Theory framework (DRT; Kamp, 1981; Kamp & Reyle, 1993), which has been used as a framework for modeling incremental referential processing (Gordon & Hendrick, 1998; Brasoveanu & Dotlačil, 2019; Kush & Eik, 2019; Kush & Eik, 2019). A much lengthier discussion replete with an explicit computational implementation of cataphor resolution within DRT can be found in Brasoveanu and Dotlačil (2019, chapter 9). When appropriate, we note differences between our model and Brasoveanu and Dotlačil's.

For present purposes the key features of DRT are that listeners actively construct a discourse representation that tracks individuals, called *discourse referents*, and information about those individuals, represented as predicates, that accrues over the course of the incremental processing of linguistic input.

Fig. 7a offers a pared-down sketch of how the construction of a discourse representation would proceed for the processing of example (19). Encountering the infinitival *driving* triggers the creation of a discourse referent, *x*, corresponding to the referent of PRO. Syntactic knowledge of control theory requires PRO to be co-interpreted with the subject of the upcoming main clause predicate. For simplicity, we represent the prediction via the addition of an underspecified main predicate, *P*, with *x* as its subject.⁶ To our knowledge, prediction of the matrix predicate is not required by DRT at this point. Thus, we are making a theoretical claim that the parser chooses to make abstract predictions about upcoming predicates when licensed by syntactic knowledge (a claim we share with many models of sentence processing: Aoshima, Weinberg & Phillips, 2004; Lewis, Vasishth & Van Dyke, 2006; Konieczny, 2000). At the cataphor *him*, the parser must postulate a new discourse referent, *y*, because the pronoun cannot refer to *x* according to Principle B; the inference that *x* and *y* are disjoint in reference guaranteed by Principle B can also be represented in the discourse representation. The presuppositions of masculine gender and singular number associated with *him* are also accommodated and entered into the discourse representation. Thus, before encountering the matrix subject phrase, the parser has constructed the discourse representation in step 2, in which the pronoun is explicitly represented as disjoint in reference from the matrix subject. When the matrix subject *Mary* is encountered, it is identified with a variable in the discourse representation, and

⁶ The simultaneity of predicates *drive* and *P* is represented by passing their events to the *while* function.

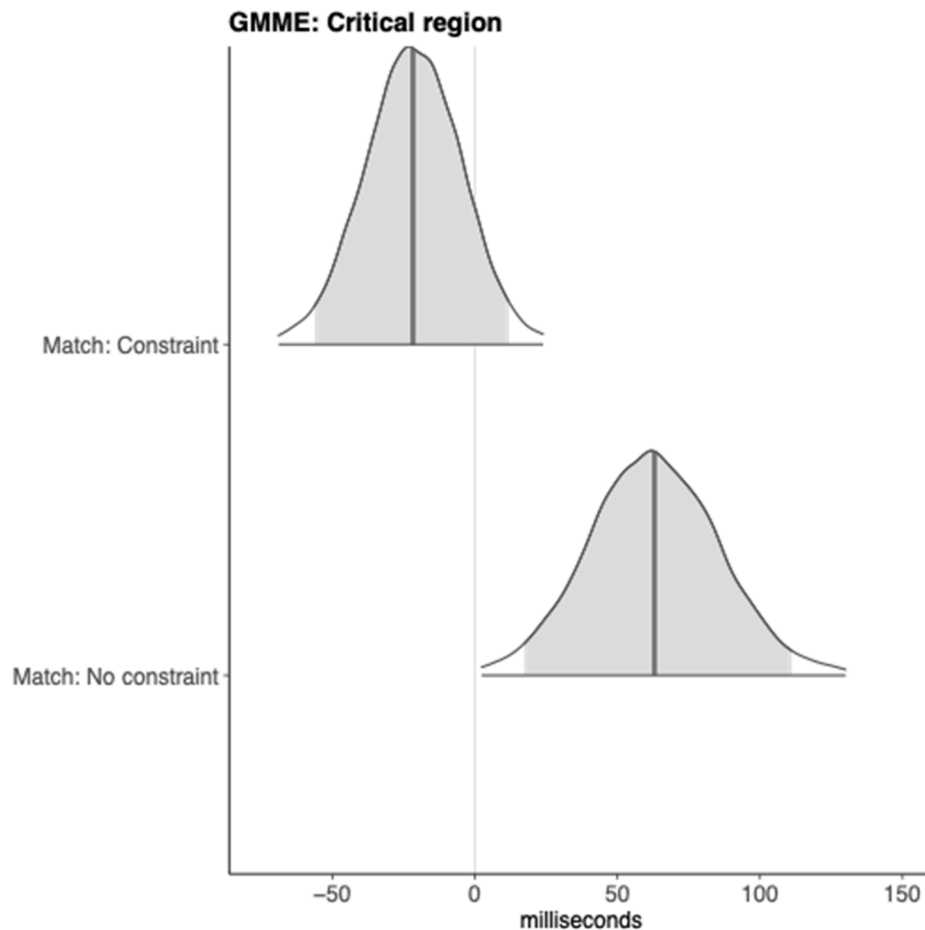


Fig. 5. Marginal posterior distributions over the gender-mismatch effect in the critical main subject region for Experiment 2, in *Constraint* and *No Constraint* conditions. Posterior distributions back-transformed to millisecond scale using the estimates in the nested model.

integration proceeds without error, even though *Mary* and *him* do not match in gender features.

Fig. 7b shows the processing of a sentence where PRO in (19) is replaced by the indefinite NP *someone*. In the absence of PRO, the key difference is that the discourse referent linked to the cataphor is not blocked from coreferring with the subject of the main clause predicate. If we suppose, with past work and our results as a guide, that the parser predictively posits that the antecedent for the cataphor is the matrix subject, we can represent this prediction by inserting the discourse referent y in the subject position of predicate P . We represent the provisional nature of this assignment with a '?' in Fig. 7b. Once again, we point out that such predictive assignment of an argument role is not forced by the DRT framework. This should be treated as one of multiple possible proposals for how to implement active search within DRT.⁷ The predictive interpretation of y as the subject of the matrix predicate yields a clash when a matrix subject with mismatching gender features is processed, because the discourse representation that results will

(perhaps temporarily) contain apparently conflicting information about the gender associated with an individual, resulting in the GMME.

The model we have articulated goes beyond our data, but it is useful in illustrating what we take to be the primary theoretical conclusions licensed by our findings: Prediction of coreference must at least be sensitive to established discourse relations (to capture the effect of obligatory control between PRO and the matrix subject), and those relations are constructed in accordance with grammatical constraints (to capture the effect of Principle B).

The model can also accommodate results from Kreiner, Sturt & Garrod (2008). In a larger eye-tracking study, Kreiner, Sturt & Garrod (2008) found that readers predictively co-interpret the main subject with a reflexive in a pre-posed adjunct as in (20): The researchers observed a GMME after the subject *minister* when it mismatched the reflexive (*himself/herself*).

- (18) After reminding *himself/herself* about the letter, **the minister** immediately ...

Such predictive co-interpretation follows if the PRO subject of *reminding* is predictively linked to the main subject, the reflexive is automatically co-interpreted with PRO in accordance with Principle A (Chomsky, 1981), and the features of the reflexive become predicated of the corresponding discourse referent.

Early versus late constraint application

Although we have characterized our results as support for a parser that uses grammatical constraints to preemptively exclude

⁷ Brasoveanu & Dotlačil (2019) do not use explicit prediction of argument positions in their model of cataphor processing. If we understand their account correctly, their model would not attempt to establish coreference between a cataphor and the matrix subject until the matrix subject was actually encountered in the input. The authors do not consider how to model GMMEs and active search, but we assume that their model could accommodate GMMEs if the parser were automatically required to posit coreference between the cataphor and the matrix subject before checking the subject's gender features. In this regard, the model would implement something akin to van Gompel and Liversedge's (2003) proposal for cataphor processing.

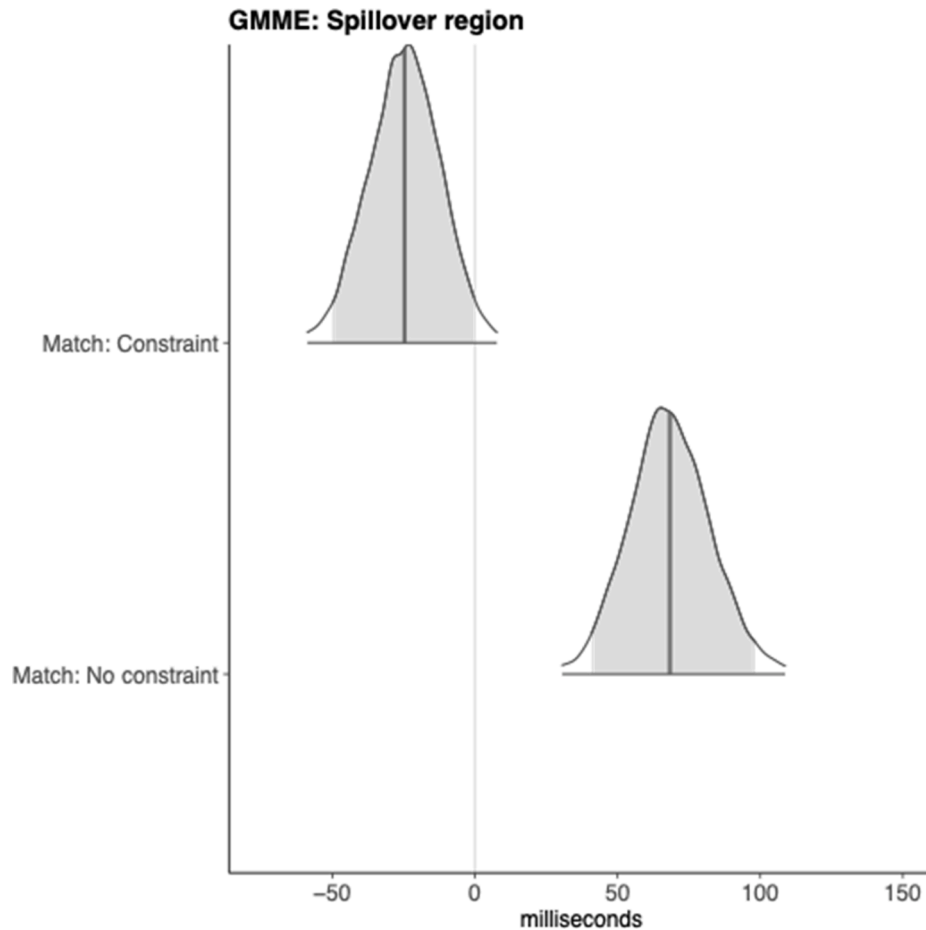


Fig. 6. Marginal posterior distributions over the gender-mismatch effect in the spillover region for Experiment 2, in *Constraint* and *No Constraint* conditions. Posterior distributions back-transformed to millisecond scale using the estimates in the nested model.

grammatically illicit antecedents during the search for a cataphor's antecedent, there remain several open issues that should be considered before accepting a strong version of this hypothesis. Most importantly, two studies mentioned above have reported early GMME effects during cataphor processing where a different grammatical constraint - Principle C - should rule them out (Drummer & Felser, 2019; Patterson & Felser, 2019). Felser and colleagues' findings might appear at first blush inconsistent with the model described above.

We see several ways to reconcile our results and these studies. The first concerns the strength of the constraints at play across studies. Drummer and Felser investigated contexts where Principle C rules out coreference between a cataphor and an R-expression that it c-commands (see 21, repeated from 17 above). As they note, even though the cataphor cannot corefer with an R-expression in a c-commanded position, it is nevertheless possible for the cataphor to corefer with a pronoun or epithet in the same position. Thus, the possibility of coreference between a cataphor and a later item depends on the specific form of the expression that occupies that position, which cannot be ascertained in advance. In contrast, coreference between the cataphor and the target position in our materials is ruled out no matter what type of noun phrase occupies that position (see 22):

- (19) a. R-EXPRESSIONS: He_i chatted with some fans while the young quarterback_{i/j} ...
 b. PRONOUN: He_i chatted with some fans while he_{i/j} ...
 c. EPITHET: He_i chatted with some fans while the conceited jerk_{i/j} ...
 (20) a. R-EXPRESSION: After PRO_i scratching him_{i/j}, Loki_i barked at Dave_j..
 b. PRONOUN: After PRO_i scratching him_{i/j}, he_i barked at Dave_j..
 c. EPITHET: After PRO_i scratching him_{i/j}, that miserable mutt_i barked at Dave_j..

c. EPITHET: After PRO_i scratching him_{i/j}, that miserable mutt_i barked at Dave_j..

This contrast may underlie the difference between these results and ours. In the Principle C contexts (21) from Drummer and Felser's studies, the parser cannot categorically rule out coreference between the cataphor and the underlined syntactic position. It can only rule out coreference once it has done enough bottom-up analysis on this position to recognize that it contains an R-expression. In our experiments, it is possible to exclude the underlined position as coreferential with the cataphor before any bottom-up analysis is performed on that region. Indeed, the differences across studies might even be expected if the early application of grammatical constraints is grounded in how effectively they allow comprehenders to predict upcoming syntactic and semantic dependencies, as originally suggested by Kazanina et al. In the original Principle C contexts, coreference cannot be categorically ruled out entirely in advance of processing the critical noun phrase. In our Principle B contexts, it can, which should allow the parser to more effectively exclude this position during active search.

It is also possible that the choice of experimental methodology underlies the differences between our results and Felser and colleagues'. Patterson and Felser (2019) investigated the time-course of constraint application in cataphoric processing using a head-to-head comparison of self-paced reading and eye-tracking-while-reading. The researchers found a simple main effect of pronoun-antecedent gender mismatch in early eye-tracking measures (first-pass reading times and regression path on the critical region), but they found an immediate interaction of constraint and gender mismatch in their self-paced reading study. Based on these apparently conflicting results, Patterson and Felser argue that

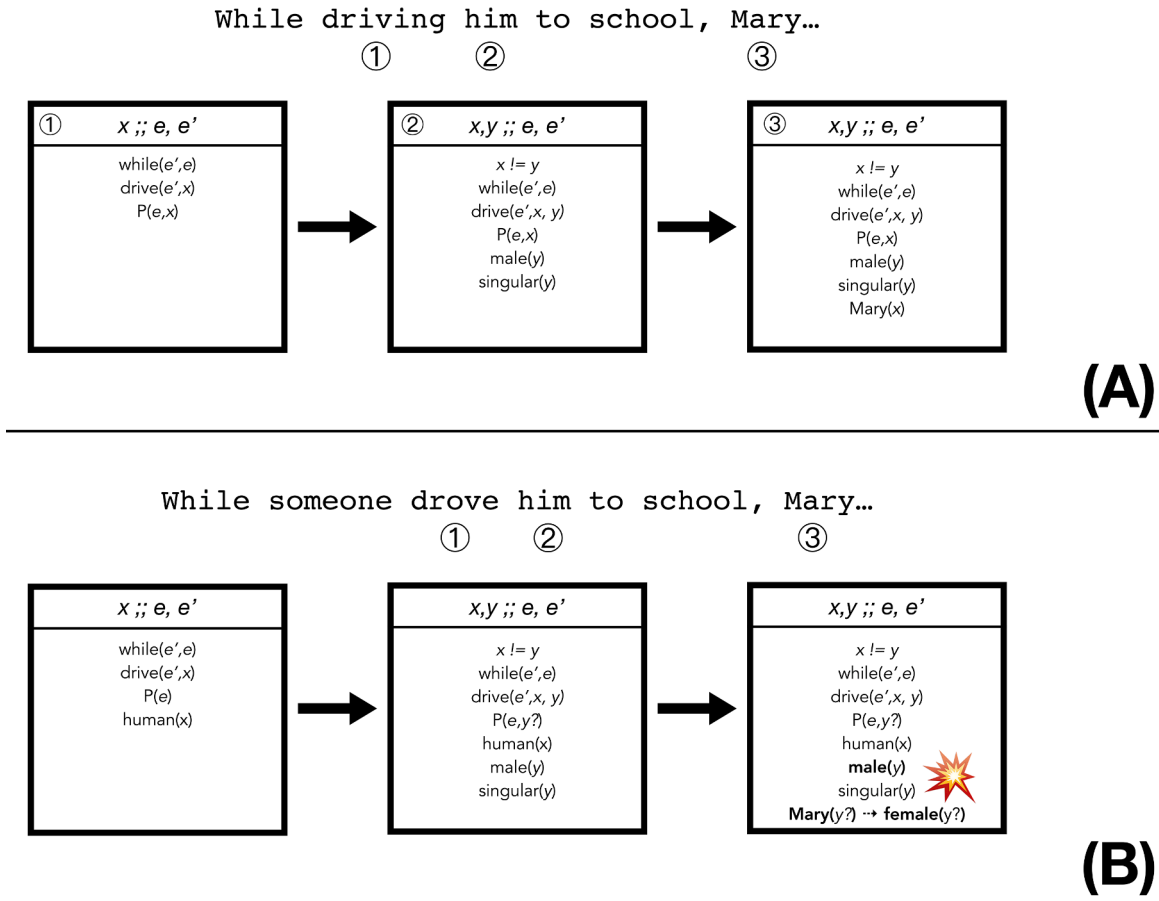


Fig. 7. Incremental processing and discourse-level prediction for cataphor sentences with gender-mismatching main subject. (A) *Constraint* sentence, where Principle B rules out coreference between cataphor and main subject. (B) *No Constraint* sentence, where coreference between cataphor and main subject is grammatically possible and gender-mismatch effect observed in Experiment 2.

self-paced reading may not have the temporal resolution to detect early, fleeting GMMEs in *Constraint* contexts, which is only seen in eye-tracking-while-reading and other measures with fine-grained temporal resolution.

We note that Patterson and Felser (2019) did not follow up the main effect of gender match in their eye-tracking study and test the pairwise differences in *Constraint* and *No Constraint* conditions. Given their relatively low sample size, it is possible that an interaction was present, but they simply lacked the power to detect it. We do not dispute that there likely is a 'true' GMME in their *Constraint* conditions, though it is possible that the 'true' GMME in *Constraint* conditions is smaller than the GMME in the *No Constraint* conditions. Such an outcome could be interpreted as supporting an alternative to the presumed dichotomy between categorical early constraint application versus categorical late application.

Given Patterson and Felser (2019)'s results, we cannot at present draw strong conclusions about the time-course of constraint application in our studies. Their study raises the possibility there was an early, fleeting GMME that is obscured by later processing in our *Constraint* contexts. At the same time, it is not clear that Patterson and Felser's methodological claims extend to the present study because of the linguistic differences between our materials and theirs discussed above (see also Drummer & Felser, 2019). Either way, this uncertainty could be resolved by conducting an eye-tracking-while-reading version of the experiments reported here. If grammatical constraints are applied as a filter on the earliest stages of processing, then we predict an interaction of gender match and constraint from the earliest point where any GMME is seen. If, instead, they are deployed as a late filter, then we expect to see a comparably-sized early GMME even in *constraint* contexts. Finally,

if all constraints apply as probabilistic/graded early filters, we would expect a non-negligible, but smaller GMME in *constraint* conditions. We leave resolving this question to future research.

Limitations of the current study

We have argued that the simplest interpretation of our data is that comprehenders avoid entertaining coreference between the cataphor and the matrix subject position in our *Constraint* conditions because Principle B is applied as an early filter in processing to block this interpretation. Still, there were some empirical and theoretical challenges to this conclusion that bear mentioning.

First, our experimental task differed from previous studies (Kazanina et al., 2007; Drummer & Felser, 2019) in that some comprehension questions directly probed the cataphor's interpretation. In principle, this could have influenced the results by introducing a strategic processing strategy whereby participants treated the task as deciding which of two names was the antecedent for a cataphor. However, in Experiment 1, these questions constituted 10% of all total questions in the experiment (33% of critical trials), and in Experiment 2, only 5% (17% of critical trials). It is unclear whether these low proportions are sufficient to introduce a task-specific strategy for resolving the cataphors. In addition, the difference in the proportion of cataphor-related questions across experiments was not reflected in any clear modulation of the key reading time effects. For these reasons we do not believe that the key findings reflect a task-specific strategy.

Second, the comprehension questions that targeted the cataphoric pronoun in both Experiment 1 and Experiment 2 suggest that

comprehenders interpreted the pronoun as coreferent with the matrix subject more often when the pronoun and the matrix subject matched in features. Importantly, we saw this pattern even in the *Constraint* conditions, where Principle B should have ruled this interpretation out. This pattern might be taken to suggest that Principle B (and/or Control) is a violable constraint. However, we find this conclusion difficult to reconcile with the reading-time results: In neither experiment did we see evidence of a GMME that would unambiguously index consideration of the illicit coreference relation in the *Constraint* conditions. An alternative possibility is that the higher error rate reflects retrieval interference that arises at the point of answering the question, caused by the presence of two gender-matching names in memory. Similarity-based interference could lead to confusion about which answer is correct even if participants only constructed the correct coreference relations during online processing. We note that this possibility is also consistent with the decreased accuracy observed in the questions that targeted the argument roles assigned the matrix predicate. Since these questions simply targeted who did what to whom in the matrix clause, they provide a baseline measure of how much interference was caused by having two gender-matching individuals in an event. In both experiments, we saw that error rates were generally higher in the *Match* conditions in these question types as well, as would be expected if the effect seen in the offline question-answering data reflects general similarity-based interference between multiple similar referents in memory. For this reason, we do not interpret the increase in ungrammatical responses in the *Match* conditions as bearing directly on the real-time search process.

There is one final feature of our data that may challenge our conclusion: the apparent ‘reverse GMME’ observed in both Experiment 1 and Experiment 2. In both experiments, we see some evidence for a small slowdown in reading times in the *Constraint*, *Match* conditions. We are reluctant to draw very strong conclusions from the reverse GMME: We did not predict this effect, and the statistical evidence for this effect is arguably limited. Only in the spillover region in Experiment 2 does the 95% credible interval for this effect not overlap 0. Still, it is worth considering what such a reverse GMME could mean. While it is generally agreed that the standard GMME arises when comprehenders interpret a pronoun as coreferent with a feature-mismatched referent, the interpretation of the reverse GMME is less clear. However, one important possibility is that the reverse GMME reflects a processing time slowdown rooted in a competitive constraint evaluation process of the sort proposed by Badecker & Straub (2002). At the point of processing the matrix subject, comprehenders may evaluate a dependency between the cataphor and the matrix subject with respect to (at least) two constraints: A feature-matching constraint, and Principle B. Under this model, the relationship between the cataphor and the matrix subject satisfies the feature-matching constraint, but clashes with Principle B. This conflict would lead to increased competition between different interpretations of the cataphor, which in turn slows processing. In contrast, the *Mismatch* conditions present the parser with a potential referent that mismatches both constraints. As a result, there would be relatively little competition from the (illicit) interpretation where the matrix subject and cataphor corefer. This would reduce processing time. A related but distinct possibility is that processing the matrix subject involves searching memory for a potentially coreferent noun phrase, which may include previously encountered pronouns. There is broad consensus that such a memory retrieval process involves cue-based reactivation of potential referents in memory (see Lewis et al., 2006 for a review). The processing dynamics of a cue-based retrieval process are roughly similar to the constraint satisfaction process sketched above (Badecker & Straub, 2002), which would imply that any memory retrieval process triggered by the matrix subject would likely engender more retrieval interference in the *Match* condition (see Badecker & Straub, 2002, for further details).⁸

Again, this speculative interpretation of the reverse GMME should be treated with caution, as this effect was not predicted in our experiments, and to our knowledge, a theoretical model of competitive constraint evaluation during forward search for a cataphor’s antecedent has not been explicitly articulated. But if this speculation is on the right track, it suggests an important qualification to our broader theoretical conclusions. In particular, it raises the possibility that Principle B is deployed as one constraint among many on forward search for a cataphor, rather than a single categorical filter. On this view, the dependency between the cataphor and the matrix subject is evaluated—but not routinely adopted—as part of this constraint satisfaction process, which would account for the ‘reverse’ GMME in place of the standard GMME effect in the *Constraint* conditions. Further research is necessary to address this possibility.

Conclusion

In two self-paced reading experiments, we tested whether comprehenders use grammatical constraints in early stages of processing to rule out coreference between a cataphor and a grammatically illicit antecedent (Ackerman, 2015; Drummer & Felser, 2019; Kazanina et al., 2007; Patterson & Felser, 2019, a.o.). We found evidence that the search for an antecedent for a cataphor displays immediate sensitivity to Principle B of the Binding Theory, when it imposes disjoint reference between the cataphor and a subsequent referent. Our results are broadly consistent with a parser that uses both syntactic and discourse-level information to predictively anticipate coreference during incremental processing.

CRediT authorship contribution statement

Dave Kush: Conceptualization, Formal analysis, Project administration, Writing - original draft, Writing - review & editing. **Brian Dillon:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgments

We are grateful to Claudia Felser and two anonymous reviewers for their helpful feedback on earlier versions of this manuscript. All errors remain our own. We also thank members of the Øyelab/Eyelands Lab (NTNU), the Psycholinguistics Workshop at the University of Massachusetts Amherst, and the University of Toronto Linguistics community for early feedback on this project. Special thanks go to Anna Giskes and Kyle Johnson for their discussion of the material in this paper. This work was supported in part by National Science Foundation Grant #BCS-1941485 to BD.

References

- Ackerman, L. M. (2015). *Influences on parsing ambiguity*. Unpublished doctoral dissertation. Northwestern University.
- Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, 51(1), 23–54.
- Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers*, 28(4), 584–589.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bosch, P. (1983). *Agreement and anaphora: A study of the role of pronouns in syntax and discourse* (Vol. 3). Academic Press.

⁸ We are grateful to an anonymous reviewer for suggesting this possibility.

- Brasoveanu, A., & Dotlačil, J. (2019). Formal linguistics and cognitive architecture. Language, Cognition, and Mind (LCAM) Series. Cham: Springer Open.
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>.
- Büring, D. (2005). *Binding theory*. Cambridge University Press.
- Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464–473.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in psychology*, 5, 630.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407.
- Clackson, K., & Clahsen, H. (2011). Online processing of cataphoric pronouns by children and adults: Evidence from eye-movements during listening. In N. Danis, K. Mesh, & H. Sung (Eds.). *Proceedings of BUCLD 35* (Vol. 1, pp. 119–131). Somerville, MA: Cascadia Press.
- Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, 65(2), 128–144.
- Clifton, C., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words her, his, him: implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, 36(2), 276–292.
- Cowart, W., & Cairns, H. S. (1987). Evidence for an anaphoric mechanism within syntactic processing: Some reference relations defy semantic and pragmatic constraints. *Memory & Cognition*, 15(4), 318–331.
- Cummings, I., & Sturt, P. (2018). Coargumenthood and the processing of pronouns. *Language, Cognition and Neuroscience*, 33(10), 1235–1251.
- Drummer, J. D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language*, 101, 97–113.
- Evans, G. (1980). Pronouns. *Linguistic Inquiry*, 11(2), 337–362.
- Filik, R., & Sanford, A. J. (2008). When is cataphoric reference recognised? *Cognition*, 107(3), 1112–1121.
- Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5), 530–543.
- Gerard, J., Lidz, J., Zuckerman, S., & Pinto, M. (2018). The acquisition of adjunct control is colored by the task. *Glossa*, 3(1).
- Giskes, A., & Kush, D. (in press). Active antecedent search triggered by cataphors persists past the subject: evidence from Norwegian and English. *Memory & Cognition*.
- Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, 22(4), 389–424.
- Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry*, 24(1), 69–101.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, 111, Article 104063.
- Johnson, K. (2013). Pronouns vs. definite descriptions. In Becker, Grinstead, and Rothman (Eds.), *Generative linguistics and acquisition: Studies in honor of Nina M. Hyams* (pp. 157–182).
- Kazanina, N., Lau, E. F., 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.
- Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology*, 63(2), 371–400.
- Kamp, H. (1981). A theory of truth and semantic representation. In C. Rohrer, A. Rossdeutscher, & H. Kamp (Eds.), *Formal methods in the study of language* (pp. 277–322). Amsterdam: Mathematical Centre Tracts.
- Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and Discourse Representation Theory*. Dordrecht: Kluwer.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6), 627–645.
- Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58(2), 239–261.
- Kush, D. W., & Eik, R. (2019). Antecedent accessibility and exceptional covariation: Evidence from Norwegian Donkey Pronouns. *Glossa: A Journal of General Linguistics*, 4(1).
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.
- Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, 23(2), 379–434.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- Marty, P. (2017). *Implicatures in the DP domain*. Unpublished doctoral dissertation. MIT.
- Mertzen, D., Laurinavichyute, A., Dillon, B., Engbert, R., & Vasishth, S. (2020). A cross-linguistic investigation of proactive, similarity-based retrieval interference in sentence comprehension: No support from English, German and Russian eye-tracking data. Retrieved from <https://doi.org/10.31234/osf.io/t2j8v>.
- Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, 10(11), 591–613.
- Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, 42, 1075–1100.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of psycholinguistic research*, 18(1), 5–19.
- Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience and Biobehavioral Reviews*, 96, 367–400. <https://doi.org/10.1016/j.neubiorev.2018.11.019>.
- Pablos, L., Doetjes, J., Ruijgrok, B., & Cheng, L. L. S. (2015). Active search for antecedents in cataphoric pronoun resolution. *Frontiers in psychology*, 6, 1638.
- Patterson, C., & Felser, C. (2019). Delayed application of binding condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research*, 48(2), 453–475.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002.
- R Core Team. (2013). R: A language and environment for statistical computing.
- Reinhart, T. (1983). Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy*, 6(1), 47–88.
- Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24(4), 657–720.
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4), 573–604.
- Schlenker, P. (2005). Minimize restrictors! (Notes on Definite Descriptions, Condition C and Epithets). In E. Maier, C. Bary & J. Huitink (Eds.), *Proceedings of Sinn und Bedeutung* (Vol. 9, pp. 385–416). Nijmegen: The Nijmegen Centre of Semantics (NCS).
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Linger misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104–120.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319.
- Sturt, P. (2013). Syntactic constraints on referential processing. In Van Gompel, R. (Ed.), *Sentence Processing* (pp. 136–159). Psychology Press.
- Sturt, P., & Kwon, N. (2015). The processing of raising and nominal control: An eye-tracking study. *Frontiers in Psychology*, 6, 331.
- Van Gompel, R. P., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 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.
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71, 147–161.
- Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2014). On the origin of islands. *Language, Cognition and Neuroscience*, 29(7), 761–770.