Syntactic and semantic interference in sentence comprehension: Support from English and German eye-tracking data

Daniela Mertzen, Department of Linguistics, University of Potsdam, Potsdam, Germany; University Library, University of Potsdam, Potsdam, Germany, mertzen@uni-potsdam.de

Dario Paape, Department of Linguistics, University of Potsdam, Potsdam, Germany, paape@uni-potsdam.de

Brian W. Dillon, Department of Linguistics, UMass Amherst, MA, USA, brian@linguist.umass.edu

Ralf Engbert, Department of Psychology, University of Potsdam, Potsdam, Germany, ralf.engbert@uni-potsdam.de

Shravan Vasishth, Department of Linguistics, University of Potsdam, Potsdam, Germany, vasishth@uni-potsdam.de

A long-standing debate in the sentence processing literature concerns the time course of syntactic and semantic information processing in online sentence comprehension. The default assumption in cue-based models of parsing is that syntactic and semantic retrieval cues simultaneously guide dependency resolution. When retrieval cues match multiple items in memory, this leads to similarity-based interference. Both semantic and syntactic interference have been shown to occur in English. However, the relative timing of syntactic vs. semantic interference remains unclear. In this cross-linguistic investigation of the time course of syntactic vs. semantic interference, the data from two eye-tracking during reading experiments (English and German) suggest that the two types of interference can in principle arise simultaneously during retrieval. However, the data also indicate that semantic cues are evaluated with a small timing lag in German compared to English. This cross-linguistic difference between English and German may be due to German having richer morphosyntactic marking than English, resulting in syntactic cues dominating over semantic cues during dependency resolution. More broadly, our cross-linguistic results pose a challenge for the cue-based retrieval model's default assumption that syntactic and semantic cues are used simultaneously during long-distance dependency formation. Our work also highlights the importance of collecting cross-linguistic data on psycholinguistic phenomena which can potentially advance theory development.

1. Introduction

A long-standing debate in the literature on syntactic ambiguity resolution concerns the role of syntactic and semantic constraints during initial structure building. Consider, for instance, the garden-path sentences in (1a,b), taken from Clifton et al. (2003):

(1)   a.   [[$_{NP}$ The man] [$_{RC}$ paid by the parents]] was unreasonable.
      b.   [[$_{NP}$ The ransom] [$_{RC}$ paid by the parents]] was unreasonable.

Syntax-first accounts of sentence processing would predict that comprehenders initially build an incorrect main clause analysis in which *paid* is analyzed as an active verb, regardless of the animacy status of the NP *the man/the ransom*. This results in a garden-path effect at the *by*-phrase, where the structure must be reanalyzed as a reduced relative clause (Frazier, 1979; Frazier 1987; Frazier & Clifton, 1996). Syntax-first models assume that in the earliest moments of structure building, only syntactic constraints play a role; semantic information is used only in a subsequent processing stage to interpret the sentence. By contrast, constraint-based accounts assume that syntactic and semantic constraints can be used simultaneously (e.g., MacDonald et al., 1994; McRae et al., 1998; Tabor & Hutchins, 2004; Trueswell et al., 1993). Under constraint-based accounts the inanimate NP *the ransom* is less likely to be considered as a potential subject of *paid* due to its implausible interpretation, leading the parser towards the correct relative clause analysis, and thus eliminating or reducing the garden-path effect.

The evidence relating to the syntax-first proposal is mixed: In support of the syntax-first view, Ferreira and Clifton (1986) and Clifton et al. (2003) found that for sentences such as (1a,b), both conditions caused initial processing difficulty at the *by*-phrase, regardless of the animacy status of the NP. Animate conditions such as (1a) caused additional processing difficulty only in later sentence regions and in re-reading. These results are consistent with the hypothesis that syntactic constraints precede semantic constraints during real-time ambiguity resolution (see also Frazier & Rayner, 1982; Pickering & Traxler, 1998; Rayner et al., 1983; Traxler, 2002, 2005; Trueswell et al., 1993).

By contrast, a number of other studies have found support for the assumption that semantic information is used immediately, consistent with constraint-based models of parsing (Just & Carpenter, 1992; Tabor et al., 2004; Traxler & Frazier, 2008; Trueswell et al., 1994). For example, in sentences like (1a,b), Trueswell et al. (1994) observed processing difficulty for animate but not for inanimate conditions. Given these conflicting results, the debate on the time course of syntactic and semantic information in parsing sentence structure during reading remains unresolved.

The open question regarding the relative timing of syntactic and semantic influences on parsing is also of crucial interest outside of garden-path configurations. Within the cue-based parsing framework (e.g., McElree, 2000; Van Dyke & Lewis, 2003; Van Dyke, 2007; Van Dyke &

McElree, 2011), the time course of syntactic and semantic information has been studied in long-distance dependency resolution.[1]

For example, in order to comprehend the sentence (2), a dependency must be established between *the child* and *loved*:

(2)   **The child** who the mother saw in the garden **loved** the rich chocolate cake.

Cue-based parsing assumes that the subject *the child* is encoded in memory, and subsequently retrieved at the matrix verb *loved*. This retrieval process is guided by retrieval cues, such as {*grammatical subject*} and {*animate*}, which are matched against memory representations of the nouns to seek out the correct grammatical subject (henceforth, the target). In (2), the syntactic and semantic retrieval cues {*grammatical subject*} and {*animate*} match not only the correct target noun *the child,* but also the intervening distractor noun phrase (NP) *the mother.* Cue-based retrieval theory assumes that when retrieval cues match multiple similar items in memory, it is more difficult for the processor to identify the target noun. The resulting processing difficulty is known as similarity-based interference. Occasionally, interference can also lead to a misretrieval of the distractor, which results in misinterpretation. For instance, in (2), *the mother* would be misinterpreted as the subject of *loved*.

The default assumption of the cue-based retrieval theory is that syntactic and semantic cues are used simultaneously (Lewis & Vasishth, 2005). However, several researchers have explored the possibility that syntactic cues may be weighted more strongly than other cues during retrieval (Dillon et al., 2013; Engelmann et al., 2020; Parker & Phillips, 2017; Sturt, 2003; Yadav et al., 2022), compatible with the syntax-first view described above. Although such a differential weighting in favor of syntactic cues still assumes simultaneous use of cues, it implies no or at least weaker effects of semantic (or other) interference manipulations compared to syntactic interference manipulations. Differential cue weighting can in principle be extended to also allow for differential "cue lag", such that syntactic cues are evaluated before semantic cues. This evaluation lag would be in line with the proposal that syntactic cues may serve a "gating" function, ruling out syntactically mismatching chunks from being considered as retrieval targets in an early processing stage (Nicol & Swinney, 1989; Sturt, 2003; Van Dyke & McElree, 2011).

Many empirical studies have reported interference from syntactically similar distractor items (e.g., Arnett & Wagers, 2017; Glaser et al., 2013; Van Dyke & Lewis, 2003; Van Dyke, 2007) or semantically similar distractor items during retrieval (e.g., Cunnings & Sturt, 2018; Glaser et al., 2013; Gordon et al., 2004; Gordon et al., 2006; Laurinavichyute & von der

---

[1] There are at least two different instantiations of cue-based retrieval; the Lewis and Vasishth (2005) model based on ACT-R (Anderson & Lebiere, 1998), and the direct-access model (McElree, 2000). We follow Parker et al. (2018) in referring to these instantiations as the cue-based retrieval framework.

Malsburg, 2022; Lowder & Gordon, 2014; Rich & Wagers, 2020; Tabor et al., 2004; Van Dyke & McElree, 2011). However, only a small subset of studies has manipulated syntactic and semantic interference at the same time, which is necessary to uncover the relative timing of the two types of constraints.

An important study in this context is Van Dyke (2007). In two eye-tracking reading experiments (Experiment 2 and 3), the subjecthood and the animacy of a distractor were manipulated in a 2 × 2 repeated-measures design. In sentence (3), the distractor *seat/man* intervenes between the critical verb *moaned* and the target subject *the lady*. In (3a,b), the distractor is not a grammatical subject whereas in (3c,d), the distractor is the subject of a complement clause. In (3c,d), the {*grammatical subject*} cue on the verb matches the target as well as the distractor, which should lead to syntactic interference in these conditions. By contrast, in (3a,b), no syntactic interference is expected because the {*grammatical subject*} cue only matches the target. Analogously, in (3b,d), semantic interference is expected because the animacy cue matches the target *the lady* as well as the distractor *the man*. By contrast, in (3a,c), no interference is expected because the animacy cue only matches the target but not the distractor *the seat*. In Van Dyke (2007)'s Experiment 2, the adverbial phrase *yesterday afternoon* was not present in the embedded clause. This was only added in Experiment 3 to remove a potential confound in the stimuli of Experiment 2: observed reading time slowdowns may not be the consequence of syntactic interference but rather due to reading two adjacent verb phrases in conditions like (3c,d). Given the potential confound of Experiment 2, Experiment 3 offered an important check of the reading time patterns found in Experiment 2.

(3)   The pilot remembered that **the lady**
   a.   who was sitting [$_{PP}$ in the smelly *seat*]
   b.   who was sitting [$_{PP}$ near the smelly *man*]
   c.   who said that [$_{NP}$ the *seat*] was smelly
   d.   who said that [$_{NP}$ the *man*] was smelly
         (yesterday afternoon) **moaned** about a refund …

In both Experiments 2 and 3, Van Dyke (2007) found reading time patterns consistent with syntactic and semantic interference effects. However, across the two experiments, the effects were observed at different sentence regions: Van Dyke reported statistically significant syntactic effects that are compatible with syntactic interference effects occurring at the critical point of retrieval, and statistically significant semantic effects only in later sentence regions. The study provides further evidence for a cue-based retrieval mechanism that is employed during online sentence processing. Van Dyke proposes that the semantic effects occur later, either because it takes "longer for the inconsistent assignment of one NP in two thematic roles to be recognized",

or because semantic effects "may be part of sentence wrap-up processing" (Van Dyke, 2007, p. 427).

In **Figure 1**, we summarize the estimates (with 95% confidence intervals) for the effects of Subjecthood (syntactic interference) and Animacy (semantic interference) of Van Dyke's experiments (E2 and E3). When inspecting the effect estimates at the critical and post-critical regions, the reading time patterns do not clearly suggest that syntactic effects occur at the verb while semantic effects occur post-critically, as discussed in Van Dyke (2007). Across the two experiments, multiple sentence regions show reading time slowdowns consistent with syntactic and semantic interference. In Experiment 2, the critical region shows reading time slowdowns in regression-path durations (RPD) and total reading times (TFT) consistent with syntactic interference, and a reading time slowdown in RPD consistent with semantic interference. At the post-critical regions, the reading time slowdowns consistent with syntactic effects are observable in RPD, but the confidence intervals for semantic interference are centered on zero. In Experiment 3, surprisingly, the added pre-critical adverbial region shows slowdowns consistent with syntactic and semantic interference in RPD and TFT. In the critical region, the reading time patterns are consistent only with syntactic interference in RPD. The final region shows a reading time slowdown consistent with semantic interference in RPD. While the patterns in the Van Dyke study are consistent with syntactic and semantic interference affecting reading time, it remains unclear whether both types of effects arise during retrieval, that is, at the critical verb. Moreover, the relatively wide 95% confidence intervals in several of the estimates from Van Dyke (2007) suggest that more data is needed for drawing firmer conclusions: it is well-known that underpowered studies will have wide confidence intervals, and that the statistically significant estimates are likely to be overestimates (Gelman & Carlin, 2014; Jäger et al., 2020; Nicenboim et al., 2018; Vasishth et al., 2023; Vasishth et al., 2018).

In a subsequent study, Van Dyke and McElree (2011) carried out two eye-tracking studies in which, inter alia, semantic interference effects were investigated. In one experiment (Experiment 1B), the distractor was in subject position whereas in another experiment (Experiment 2B), it was in a non-subject position. This study offers equivocal results with regard to where semantic interference effects occur. In Experiment 1B, in which the distractors were grammatical subjects, a pattern consistent with semantic interference was observed in total fixation time at the critical verb (95% CI [–16, 116] ms). By contrast, in Experiment 2B, which had distractors in non-subject position, the corresponding estimate for semantic interference didn't show any clear pattern (95% CI [–45, 87] ms).[2]

---

[2] There was no information available for computing the standard error of this estimate in the paper, so we took the corresponding standard error from Experiment 1B as an approximation.
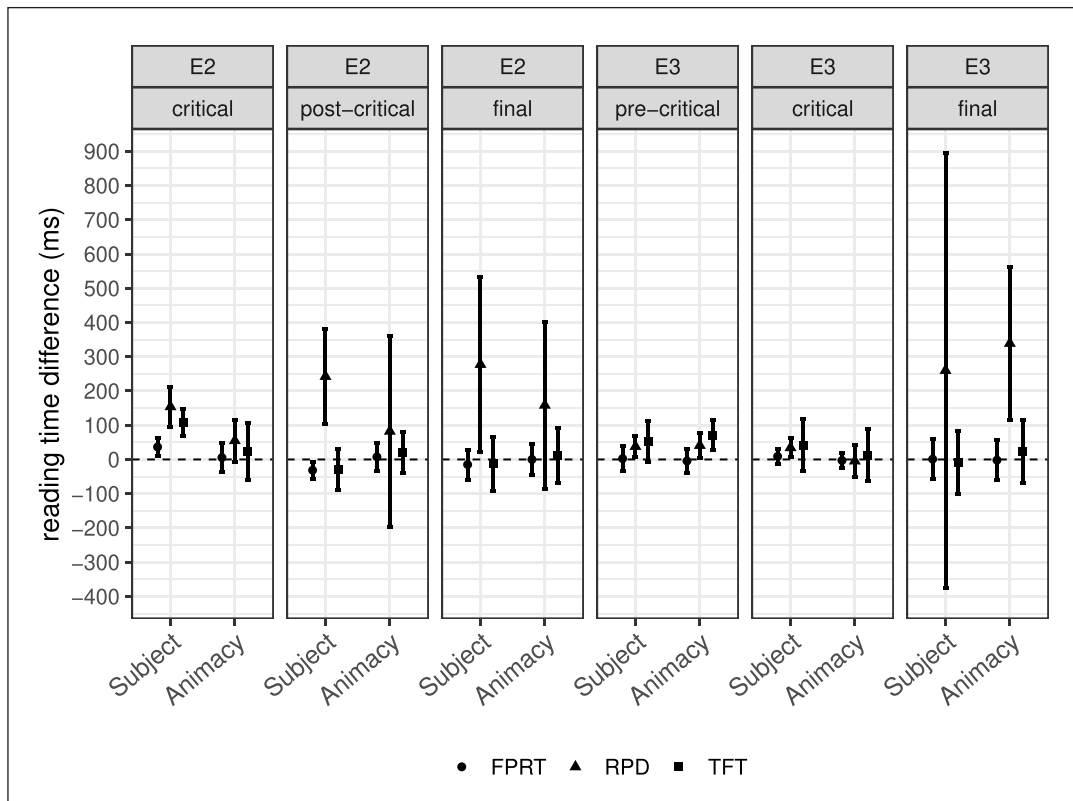
Figure 1: Estimated means and 95% confidence intervals (CIs) extracted from the reported statistics in Van Dyke (2007) for syntactic and semantic interference. *Subject* refers to the manipulation of the distractor's subjecthood (main effect of syntactic interference). *Animacy* refers to the manipulation of distractor animacy (main effect of semantic interference). A positive sign means that there is a slowdown for [+subject] distractor conditions, or for [+animate] distractor conditions. The means and CIs are shown for the critical, post-critical, and final regions of Experiment 2, and the pre-critical, critical, and final sentence regions of Experiment 3. The effects were observed in first-pass reading times (FPRT), regression-path durations (RPD) and/or total fixation time (TFT). For the effects that were reported as non-significant, there was no information available to compute the standard error. Following Jäger et al. (2017), we took the largest standard error of the untransformed reading times for a given measure and region that were reported in Van Dyke (2007).

An interesting proposal from this study is that the difference in the patterns in Experiments 1B vs. 2B may suggest that the syntactic status (subject or non-subject) modulates or "gates" semantic interference such that a distractor—even if it is a semantic match—is only considered as a potential retrieval target when it shares syntactic features with the target noun. However, a limitation of this study is that the contrast between subject and non-subject distractors was not directly tested in a within-subjects design. For such a conclusion to be drawn, a cross-experiment analysis is required, testing the interaction between semantic interference and experiment. This

interaction would need to have a clearly positive sign (Nieuwenhuis et al., 2011). The interaction estimate derived from the published statistics in Van Dyke and McElree (2011), however, spans a broad range of negative and positive values (95% CI [–64, 122] ms).

Overall, the Van Dyke studies reported syntactic and semantic interference effects. However, these were observed at different sentence regions across experiments. Given the equivocal results in the Van Dyke studies, more research is necessary on the timing of semantic and syntactic interference effects.

In the present work, we use the design of Van Dyke (2007) to investigate the time course of syntactic and semantic constraints during retrieval in English. Our study uses new stimuli to eliminate a potential confound that was present in the materials of Van Dyke (2007) (see Section 2.1). To test whether the interference patterns can be observed cross-linguistically, a second, larger-sample experiment investigates the relative timing of syntactic and semantic interference in another language, German.

Given the cue-based theory's default assumption of simultaneous use of syntactic and semantic retrieval cues, it is hypothesized that both syntactic and semantic interference effects arise at the retrieval point. This would also be compatible with constraint-based models of sentence parsing that assume both types of information can be used simultaneously. However, if only syntactic interference arises during retrieval, and semantic interference arises at a later sentence region, this would speak in favor of syntactic information preceding semantic information during online dependency resolution, in line with the claim of syntax-first models. Additionally, the within-subjects design of our larger-sample experiments can address the syntactic gating proposal, that is, the question of whether semantic interference only occurs when a distractor additionally matches a verb's syntactic cue.

To anticipate our results, in English, at the critical verb, where memory retrieval is assumed to occur, we found reading time patterns consistent with syntactic and semantic interference. In German, at the critical verb, we observed reading time patterns consistent only with syntactic interference. However, the post-critical region showed reading time slowdowns consistent with semantic interference. The divergent patterns in English and German suggest that while syntactic and semantic cues can be used simultaneously during dependency resolution, there may be cross-linguistic differences. A possible cause of differences in cue timing could be the amount of morphosyntactic information available in a given language. An additional, surprising result was that both languages showed unexpected reading time slowdowns in the pre-verbal modifier region. In Section 6, we explain that these pre-critical slowdowns are compatible with encoding interference effects and/or predictive processing effects. It is also possible that both encoding interference and predictive processing act in tandem to give rise to the observed slowdowns on the pre-verbal modifier.

2. The present eye-tracking study

2.1 Experimental design and materials

Our eye-tracking-while-reading experiments used a 2×2 fully-crossed factorial design. Syntactic interference was tested by manipulating DISTRACTOR SUBJECTHOOD [–subject, +subject], and semantic interference was tested by manipulating DISTRACTOR ANIMACY [–animate, +animate]. Here, [+] denotes that the distractor is a subject/animate, and [–] means that it is not a subject/ not animate. In the following we refer to the factors SUBJECTHOOD and ANIMACY to abbreviate DISTRACTOR SUBJECTHOOD/ANIMACY. The two factors were manipulated within-subjects and within-items, resulting in the four conditions shown in **Table 1**.

Table 1: English example sentences. Factor 1 (SUBJECTHOOD) manipulated whether the distractor (underlined) was a subject (+subj) or not a subject (–subj). Factor 2 (ANIMACY) manipulated whether the distractor was animate (+anim) or inanimate (–anim).
Conditions a,b: *It turned out that the attorney whose secretary had forgotten about the important meeting/visitor frequently complained about the salary at the firm.*
Conditions c,d: *It turned out that the attorney whose secretary had forgotten that the meeting/the visitor was important frequently complained about the salary at the firm.*

It turned out that **the attorney**$_{+anim}^{+subj}$

a. [–subj; –anim]

whose secretary had forgotten about the important <u>meeting</u>$_{-anim}^{-subj}$

b. [–subj; +anim]

whose secretary had forgotten about the important <u>visitor</u>$_{+anim}^{-subj}$

c. [+subj; –anim]

whose secretary had forgotten that the <u>meeting</u>$_{-anim}^{+subj}$ was important

d. [+subj; +anim]

whose secretary had forgotten that the <u>visitor</u>$_{+anim}^{+subj}$ was important

frequently **complained** $\{_{anim}^{subj}\}$ about the salary at the firm.

Our study tested subject-verb dependencies similar to those used in Van Dyke (2007). In all conditions, the critical dependency is between the verb *complained* (the point of retrieval) and its subject NP *the attorney* (the target of retrieval). Retrieval cues at the verb, {*grammatical subject*} and {*animate*}, always fully match the features [+subject], [+animate] of the target noun phrase.

The critical manipulation concerns the subjecthood and animacy of the distractor noun *meeting/visitor*. In the [–subject] conditions (a, b), the distractor is the direct object of the relative clause. Therefore, it does not match the retrieval cue {*grammatical subject*}. By contrast, in the

[+subject] conditions (c, d), the distractor is the subject of the complement clause, matching the {*grammatical subject*} cue. In the [+animate] conditions (b, d), the distractor matches the animacy cue, while in the [–animate] conditions (a, c), it mismatches the animacy cue at the verb. Whenever there is a (partial) match between the retrieval cues at the verb and the features of the distractors, this should lead to similarity-based interference during the retrieval of the target.

Like the materials in Van Dyke (2007), our stimuli contain an additional animate distractor (e.g., *secretary*) across all conditions. We added the additional distractor to increase the strength of the manipulation (Nicenboim et al., 2018; Parker & Phillips, 2017). This additional distractor in our materials was added to the relative clause, that is, it intervenes between the critical verb and target; this configuration is called retroactive interference in the memory literature. By contrast, in the Van Dyke (2007) study, the additional distractor preceded the critical dependency; this configuration is called proactive interference.[3] We made this change because previous research suggests that retroactive configurations may be stronger than proactive configurations (Van Dyke & McElree, 2011; also see the meta-analysis in Jäger et al., 2017).

Our stimuli address a potential confound in the original Van Dyke (2007) study (previously discussed in Arnett & Wagers, 2017; Wagers, 2008): While Van Dyke's Experiment 3 had an adverbial phrase in the embedded clause to avoid slowed reading times due to reading two adjacent verb phrases, syntactic effects on the critical verb could be a consequence of clause boundary wrap-up. Specifically, in [+subject] distractor conditions, such as sentence (3c,d), the critical verb region immediately follows two clause boundaries, while in the [–subject] distractor conditions, such as sentence (3a,b), it follows only one clause boundary. The processing of the additional clause boundary may create more difficulty, raising the possibility that the observed effects were the result of clause boundary wrap up, and not the result of syntactic interference from subject distractors. We addressed this potential issue by adding an adverb, not in the embedded clause prior to the clause boundary like the original study, but as a pre-verbal modifier.

There is one remaining potential confound in the English stimuli: it is possible that a semantic effect in [–subject] distractor conditions is a consequence of a syntactically illicit but locally coherent parse (Tabor et al., 2004), such as in sentence (3a) of Van Dyke (2007)'s Experiment 2 (*near* [*the smelly man moaned*]), or in our stimuli (*about* [*the important visitor complained*]), shown in **Table 1b**. This potential confound in the pre-critical region can be ruled out entirely by our German stimuli. Since commas are obligatory at clause boundaries in German, the possibility of a locally coherent parse is eliminated. Beyond this, the overall word order in our stimuli is largely similar in the two languages, with one notable exception: Unlike English, German word order in

---

[3] To the best of our knowledge, the terms proactive and retroactive interference from the memory literature were first invoked in the context of sentence processing theories by Lewis (1996).

subordinate clauses is always verb final. This means that the linear position of the distractor vis-à-vis the critical verb is not identical in the two [–subject] conditions across languages.

The design of the German experiment closely matched that of the English experiment. **Table 2** shows example sentences. Here, a dependency must be established between the verb *log* (*'lied'*) and the target NP *der Journalist* (*'the-NOM journalist'*). As in the English example, the distractor is either the direct object of the embedded clause in conditions (a, b), or the grammatical subject as in conditions (c, d). The SUBJECTHOOD factor is crossed with ANIMACY such that the distractor is either inanimate (*Skandal, 'scandal'*; a, c), or animate (*Mafiaboss, 'mafia boss'*; b, d).

Table 2: German example sentences. Factor 1 (SUBJECTHOOD) manipulated whether the distractor (underlined) was a subject (+subj) or not a subject (–subj). Factor 2 (ANIMACY) manipulated whether the distractor was animate (+anim) or inanimate (–anim).
Conditions a,b: *It turned out that the journalist whose colleague had reported on the gruesome scandal/mafia boss in fact lied to obtain information.*
Conditions c,d: *It turned out that the journalist whose colleague had reported that the scandal/mafia boss was gruesome in fact lied to obtain information.*

> Es stellte sich heraus, dass **der Journalist**$^{+subj}_{+anim}$,
>
> *It turned out that the journalist,*
>
> **a. [–subj; –anim]**
>
> dessen Kollege von dem grauenhaften <u>Skandal</u>$^{-subj}_{-anim}$ berichtet hatte,
>
> *whose colleague of the gruesome scandal reported had,*
>
> **b. [–subj; +anim]**
>
> dessen Kollege von dem grauenhaften <u>Mafiaboss</u>$^{-subj}_{+anim}$ berichtet hatte,
>
> *whose colleague of the gruesome mafia boss reported had,*
>
> **c. [+subj; –anim]**
>
> dessen Kollege berichtet hatte, dass der <u>Skandal</u>$^{+subj}_{-anim}$ grauenhaft war,
>
> *whose colleague reported had, that the scandal gruesome was,*
>
> **d. [+subj; +anim]**
>
> dessen Kollege berichtet hatte, dass der <u>Mafiaboss</u>$^{+subj}_{+anim}$ grauenhaft war,
>
> *whose colleague reported had, that the mafia boss gruesome was,*
>
> tatsächlich **log**$\{^{subj}_{anim}\}$, um Informationen zu erhalten.
>
> *indeed lied, to obtain information.*

A possible important difference between English and German is that German has overt morphological case marking. There are some previous findings that indicate that overt case marking may modulate interference in the production literature (e.g., Badecker & Kuminiak,

2007; Nicol & Antón-Méndez, 2009) as well as in the comprehension literature (e.g., Slioussar, 2018) (cf. Avetisyan et al., 2020; Turk & Logacev, 2022). German has overt case marking (nominative, accusative, dative, or genitive) on determiners, nouns, and adjectives. Masculine nouns have unambiguous case marking, while feminine and neuter nouns show syncretism between nominative and accusative case. In our German experimental items, the grammatical roles of all noun phrases in the sentence were disambiguated prior to the critical verb either by (a) unambiguous morphological case marking (masculine nouns) and/or by (b) the noun phrases being dependents of case-assigning embedded verbs or prepositions. Half of the items had feminine nouns in NP1 position (that is, directly following the complementizer of the outermost clause), which could, in principle, be accusative up to the critical verb, but would canonically be interpreted as nominative.

Forty experimental items were created for each of the two experiments. For both languages, we carried out an online plausibility rating experiment[4] in order to check that all animate NPs were similarly plausible subjects of the critical verb, and that the inanimate distractor NP was implausible across items. The ratings also helped ensure that any differences between English and German are not the result of different plausibility judgements between languages. For each experimental item, each noun phrase was combined with the critical verb, resulting in four conditions as shown in example (4).

Participants rated these sentences on a scale from one ('1', *very implausible*) to seven ('7', *very plausible*). Plausibility ratings in **Figure 2** show that the animate noun phrase received very high plausibility ratings in both languages, whereas the inanimate conditions received very low ratings.

(4)   a.   The attorney complained.
         b.   The secretary complained.
         c.   The meeting complained.
         d.   The visitor complained.

The English study had 92 fillers and the German study had 90 fillers. All experimental sentences and half of the filler sentences were followed by a comprehension question. For experimental items, the questions targeted one of the three NPs in the sentence (e.g., *Who complained?* in **Table 1**). Each question had four response choices: one of the three NPs, or 'I don't know'. For instance, the example sentences in **Table 1** had the response choices *an attorney, a secretary, a visitor* or *'?'* (*'I don't know'*). For [–animate] conditions, instead of the inanimate NP, the animate distractor from the [+animate] conditions was used as a response choice. This question-response design is more demanding for the participant than the two-choice response in Van Dyke (2007);

---

[4] The experiment was implemented in PC Ibex, https://farm.pcibex.net/; participants were recruited via Prolific, https://www.prolific.co/.

our question-response design was designed to encourage deeper engagement with the target sentences.
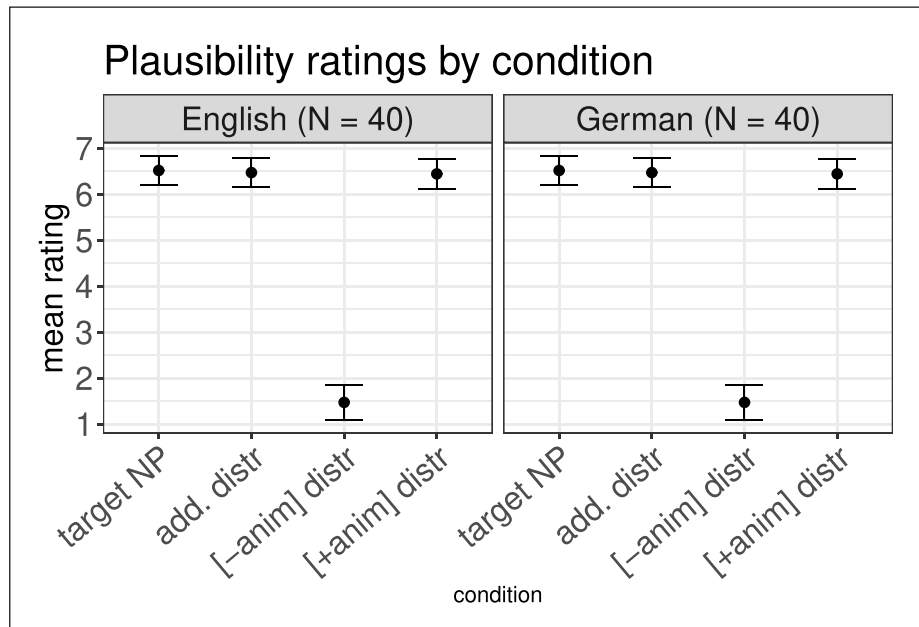

Figure 2: Mean by-condition plausibility ratings with 95% confidence intervals for the English and the German stimuli. 1 = very implausible, 7 = very plausible. Target NP (e.g., *the attorney* in Table 1). Add. distr = additional distractor that was present across all items (e.g., *the secretary*). [–anim] = inanimate distractor condition (e.g., *the meeting*), [+anim] = animate distractor condition (e.g., *the visitor*).

2.2 Participants

Our English study tested 61 participants.[5] These were mostly undergraduate students from the University of Massachusetts Amherst, MA, USA, who were reimbursed with 15 USD. The mean age was 19 years (range 18 to 28); 75% of participants reported female gender and 25% reported male gender.

For the German experiment, 121 participants were tested. The participants were undergraduate students from the University of Potsdam, Germany, who were reimbursed either with 15 Euro or course credit for their participation. The mean age was 24 (range 18 to 50), and 76% reported female gender and 24% male gender. All participants had normal or corrected-to-normal vision and no known history of language disorders.

---

[5] We aimed to test up to 120 participants but testing stopped due to COVID-19-related lab closures.

2.3 Procedure

All participants gave informed consent to take part in the study. The participants were seated in front of a presentation monitor (1440×900 resolution). Head movements were restricted using a head- and chinrest.

An EyeLink 1000 eye-tracker with a tower mount was used to record eye-movements. After a calibration procedure, six practice trials familiarized participants with the task. For each trial, participants first read a sentence and then answered a comprehension question. Sentences were presented in one line on the computer screen in a monospaced font (Consolas) of size 16. The eye-to-screen distance was 64 cm such that 4.5 characters were within one degree of visual angle.

The items were presented according to a Latin Square design such that each participant saw only one condition of each item, and the order of the items was randomized for each participant. The response choices for comprehension questions were displayed in the center top, left, right and bottom of the screen. The 'I don't know' choice was always presented in the same location at the bottom of the screen. The presentation location of the other three response choices was randomized.

As the German experiment was tested in a different lab, the setup was not identical: an EyeLink 1000 Plus[6] was used to conduct monocular tracking of the right eye. The monitor resolution was 1920×1080. The German sentences were presented in font size 14, and the eye-to-screen distance was 56 cm resulting in 2.4 characters within one degree of visual angle.

For both the English and German study, a break was offered halfway through the experiment to avoid fatigue effects. Participants were invited to take additional breaks whenever needed. After each break, a re-calibration was performed. Each experiment session lasted approximately one hour.

3. Predictions

Cue-based retrieval theory predicts a reading time slowdown for conditions with a [+subject] distractor compared to conditions with a [–subject] distractor. Such a main effect of SUBJECTHOOD would indicate syntactic interference. A reading time slowdown is also predicted for [+animate] distractor conditions compared to [–animate] distractor conditions. This main effect of ANIMACY would suggest semantic interference. Crucially, cue-based theories predict these reading time slowdowns at the point of retrieval, that is, at the critical verb. If both syntactic and semantic interference occur at the critical verb in the same reading measures, this would be consistent with the simultaneous use of retrieval cues. By contrast, if only syntactic interference is observed

---

[6] https://www.sr-research.com/eyelink-1000-plus/.

at the critical verb, but semantic interference is only observed post-critically, this would favor syntax-first accounts of sentence processing.

We present simulations from the cue-based retrieval model showing the predictions based on the default assumption that syntactic and semantic cues are used simultaneously.[7] We computed the quantitative predictions from an R implementation (R Core Team, 2019) of the Lewis and Vasishth (2005) model of cue-based retrieval for the Van Dyke (2007) design. We defined prior distributions on the free parameters, and then generated prior predictive data from the model (Vasishth, 2020); these are the predictions from the model before the data are taken into account (Gelman et al., 2014). Following the approach taken in Vasishth (2020), we only defined a prior distribution on the latency factor parameter, holding other parameters constant at the values reported in Engelmann et al., (2020) and Jäger et al. (2020). The latency factor is a scaling parameter that maps activations to retrieval time in milliseconds; it is usually a free parameter in ACT-R modeling (Anderson et al., 2004). For modeling reading times, the prior on the latency factor was *Beta*(4,6); this was the value used in Engelmann et al. (2020), Jäger et al. (2020) and Vasishth (2020).

In order to generate prior predictive data, the model was set up to use three equally-weighted cues for the retrieval at the verb: [$\pm$*subject*], [$\pm$*animate*], and [$\pm$*same_clause*]. The addition of the [$\pm$*same_clause*] cue is necessary to identify the correct subject.[8] The cue serves as a stand-in for the additional information—besides the structural position of the NP and its animacy—the parser uses to achieve correct retrieval. With only two cues, in the [+*subject*, +*animate*] conditions, the model would otherwise predict an equally-matched race between the target NP and the distractor NP (e.g., Jäger et al., 2020): In the [+*subject*, +*animate*] condition, the subject NP *the attorney* is animate and occupies a syntactic subject position, but this is also true for the distractor NP *the visitor*. Without a third cue, the model would retrieve each NP 50% of the time, thus predicting 50% misinterpretations, because it would have no way of identifying *the attorney* as the correct subject.

Figure 3 shows the prior predicted syntactic and semantic interference effects, along with the reported mean differences and their 95% confidence intervals at the critical region in Van

---

[7] However, note that there is some evidence suggesting that non-syntactic cues affect later processing, which may point towards a multi-stage processing architecture (Cunnings & Sturt, 2014; Lago et al., 2015; Sturt, 2003; Van Dyke, 2007; Wagers et al., 2009). To our knowledge, a multi-stage retrieval model has not been computationally implemented, so that pinning down and empirically evaluating the precise predictions of such a model with regard to the time course of processing will be a major project.

[8] It remains an open question how cue-based retrieval models can make use of relational information such as the [$\pm$*same_clause*] cue used here, since relational information of this sort does not generally characterize the features that should hold of any given chunk in memory. See Kush (2013) and Franck and Wagers (2020) for discussions of the difficulties inherent in encoding relational information in a cue-based framework, and proposals about how to address these theoretical challenges.

Dyke (2007) for the two interference types. We derived the estimates and confidence intervals for first-pass reading times (FPRT), regression-path durations (RPD), and total fixation times (TFT) from the published estimates and statistics. We chose these three measures because these are the reading time measures reported in Van Dyke (2007). Van Dyke (2007) also reported the proportion of first-pass regressions. We report only the reading time measures here, as they can be more straightforwardly mapped onto the retrieval times predicted by the Lewis & Vasishth (2005) model. In general, however, establishing a direct mapping between the latent cognitive processes assumed by the Lewis & Vasishth (2005) model and eye movements requires a more sophisticated modeling environment; such a model is presented in Rabe et al. (2021) and Rabe et al. (2023).
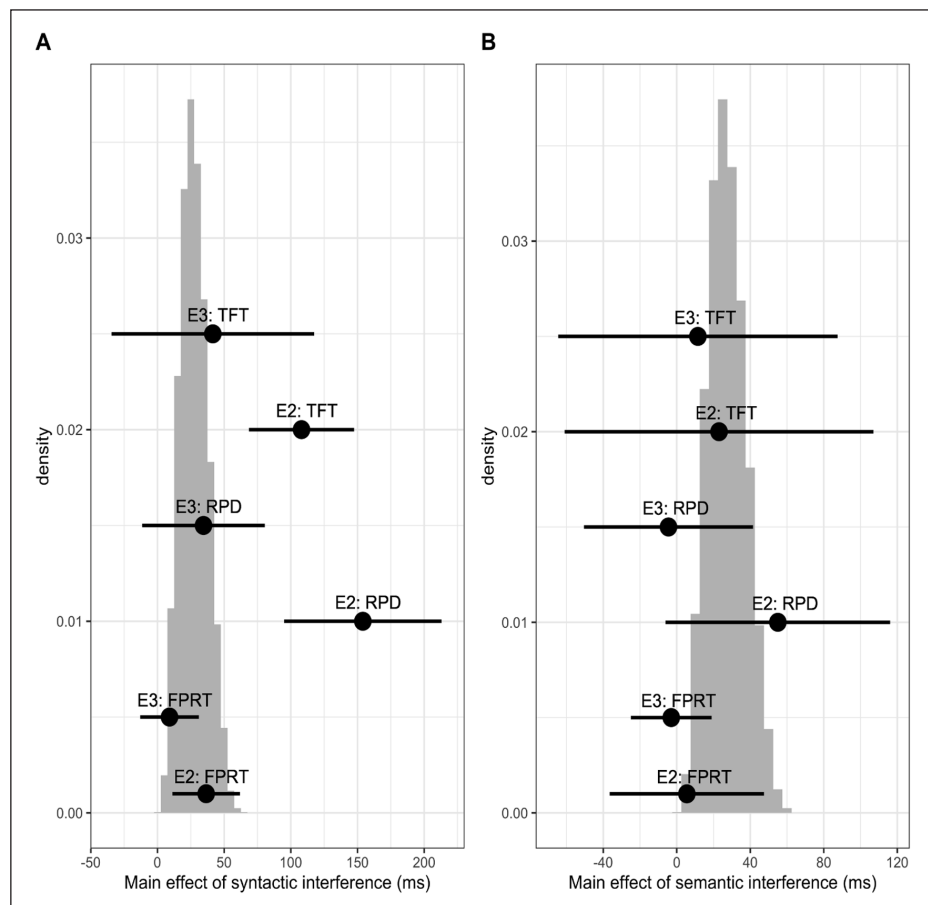

Figure 3: Prior predicted reading times for syntactic and semantic interference from an R implementation of the Lewis & Vasishth (2005) model as described in Engelmann et al. (2020); the implementation is available at https://github.com/felixengelmann/inter-act/. Also shown are the mean estimates of the reading time effects in milliseconds (ms) in the Van Dyke (2007) data from Experiments 2 and 3 (critical region), along with 95% confidence intervals. All the estimates from the Van Dyke (2007) data are derived from published estimates and statistics. FPRT = first-pass reading times, RPD = regression-path durations, TFT = total fixation times.

To compare the quantitative predictions of the cue-based retrieval model to the Van Dyke (2007) estimates, we use the region of practical equivalence (ROPE) approach (Freedman et al., 1984; Kruschke, 2015; Spiegelhalter et al., 1994). In essence, we compare the predicted range of an effect size with the observed 95% confidence intervals from the data. If the uncertainty intervals partly overlap, one can conclude that there is some degree of consistency between the predictions and data. If the regions do not overlap at all, we can conclude that the model predictions are not consistent with the data. A perfect overlap between the prediction and data is considered to indicate strong consistency. It is also possible that the data's uncertainty interval is so much larger than the range predicted by the model that it subsumes the model prediction. This would suggest weak consistency between the model and data; the consistency is weak because the uncertainty intervals from the data allow too broad a range of values, which would be considered uninformative (Roberts & Pashler, 2000).

Figure 3A reveals that for Van Dyke (2007)'s Experiment 2, the FPRT estimate for syntactic interference overlaps with the model estimates, but there is no overlap between the RPD and TFT estimates and the model predictions. The model's estimates for syntactic interference align more closely with the data from Van Dyke's Experiment 3 than with those from Experiment 2. The empirical estimates from Experiment 3 largely overlap with the model predictions. In **Figure 3B**, for the semantic effect, the FPRT estimate of Experiment 3 only partially matches the model predictions. All other empirical estimates from Experiments 2 and 3 largely overlap with the model predictions. However, the RPD and TFT estimates from both Experiments 2 and 3 are quite wide. In **Figure 3A**, for the syntactic effect, the model estimate is contained within the RPD and TFT estimates from Experiment 3. In **Figure 3B**, for the semantic effect, the RPD and TFT estimates are also so wide that they subsume the model predictions, allowing for too broad a range of values to be informative (Freedman et al., 1984; Kruschke, 2015; Kruschke & Liddell, 2018; Spiegelhalter et al., 1994).

The cue-based retrieval model's predictions are only for the critical region (the verb), where the retrieval is assumed to occur. We analyzed both the critical and post-critical region because effects that originate in the critical region can spill over to the post-critical region (Mitchell, 1984; Vasishth & Lewis, 2006), and because previous work has shown patterns in the post-critical regions that are consistent with semantic interference. We also analyzed the pre-critical region because Van Dyke (2007) found effects in the pre-critical region. Additionally, although the model predicts main effects of syntactic and semantic interference but not an interaction, we test for the SUBJECTHOOD × ANIMACY interaction. This is to evaluate the "gating proposal" in Van Dyke and McElree (2011), that is, the claim that semantic interference only occurs when the distractor is also a syntactic match. This implies that a reading time slowdown consistent with semantic interference would be observed at the verb in [+subject] but not in [−subject] conditions.

4. Statistical analyses

In the present paper, we move away from the null hypothesis significance testing approach and adopt a (Bayesian) estimation approach. This is because our goal is to quantify uncertainty of the effect estimates, so that future meta-analyses can incorporate these estimates, and replication attempts can use these to establish the consistency of the effect (Freedman et al., 1984; Spiegelhalter et al., 1994). Following e.g., Gelman et al. (2014), Jäger et al. (2020), Kruschke (2015), Kruschke and Liddell (2018), Nicenboim et al. (2023), Vasishth et al. (2018) and Vasishth and Gelman (2021), we report Bayesian 95% credible intervals (CrIs) of the posterior distributions. CrIs demarcate the range within which an unobserved parameter's value falls with 95% probability, given the data and the model. CrIs should not be used to make binary decisions like "effect present/absent" (Cumming, 2014; Kruschke & Liddell, 2018; Royall, 1997).

Following Van Dyke (2007), we report the three reading time measures first-pass reading times, regression-path durations, and total fixation times. In addition, we report the proportion of first-pass regressions. First-pass reading times include the sum of all fixations on a region n before a forward or backward saccade is launched. Regression-path durations consist of the sum of all first-pass fixation durations on region n, including any fixation durations that result from regressions out of region n, until n is left to the right. Total fixation time is defined as the sum of all fixations that occurred during the first pass and during re-reading of a region n. All the three reading time measures excluded 0 ms values (Nicenboim et al., 2023). The proportion of first-pass regressions measure is defined as the proportion of regressive saccades out of region n during the first-pass (Rayner, 1998). The measures were computed from the eye-tracking record using the `em2` package (Logacev & Vasishth, 2013).

All statistical modeling was carried out in the programming environment `R` (R Core Team, 2019). Bayesian hierarchical models were fit to the reading time data (e.g., Gelman et al., 2014; Kruschke, 2015), using the probabilistic programming language `Stan` (Carpenter et al., 2016) and the front-end to `Stan, brms` (Bürkner, 2017). A log-normal likelihood was assumed for the reading time data (Nicenboim et al., 2023). The models included the factors SUBJECTHOOD (–subject, +subject), ANIMACY (–animate, +animate), and the interaction as fixed effects. As shown in **Table 3**, these contrasts were sum-coded Schad et al., 2020). Participants and items were specified in the models as random effects, with full variance-covariance matrices.

Regularizing prior distributions were specified for all the model parameters except the intercept (Nicenboim et al., 2023). In the prior specifications below, we always parameterize the normal distribution with the standard deviation, following the practice in the R and Stan programming environments.

In all the models, the intercept had a $\mathcal{N}(0,10)$ prior; this is an uninformative prior that aids stable computation. The slopes had a regularizing prior, $\mathcal{N}(0,0.1)$, and all variance

components had a $\mathcal{N}(0,0.5)$ prior. For the correlation matrix of the random effects variance-covariance matrix, a regularizing LKJ prior was specified (Lewandowski et al., 2009). The shape parameter ν (nu) of the LKJ prior was set to 2. This ensures that extreme correlations like ± 1 are downweighted.

Table 3: Sum contrast coding for effects of Subjecthood, Animacy and their interaction. For the reading time measures, a Subjecthood or Animacy effect with a positive sign indicates slower reading times for [+subject/+animate] conditions, compared to [−subject/−animate] conditions. For the proportion of first-pass regressions and comprehension accuracy measures, a positive sign indicates a higher proportion of regressions, or a higher accuracy, for the [+subject/+animate] conditions, compared to the [−subject/−animate] conditions.

Condition	Subject	Animacy	Interaction
−subject, −animate	−0.5	−0.5	+0.25
−subject, +animate	−0.5	+0.5	−0.25
+subject, −animate	+0.5	−0.5	−0.25
+subject, +animate	+0.5	+0.5	+0.25

Each model was run with four chains and 4000 iterations. The first 2000 of these served as warm-up iterations; that is, they were not used for inference. The $\hat{R}$-statistic and trace plots were inspected to check model convergence (Gelman et al., 2014). All estimates of reading time effects are reported on the millisecond scale; these were back-transformed from the log-scale (Crow & Shimizu, 2018; Nicenboim et al., 2023).

For proportions of first-pass regressions and question response accuracies, hierarchical logistic models were fit with regularizing priors, and full variance-covariance matrices for subject and item random effects. The prior distribution on the intercept was specified as $\mathcal{N}(0,2)$; the slopes had a $\mathcal{N}(0,0.5)$ prior, and the variance components had a $\mathcal{N}(0,1)$ prior. Extreme correlations were also downweighted by setting the ν parameter of the LKJ prior to 2. In all models, four chains with 4000 iterations were specified. For the first-pass regression and response accuracy results, we also report the 95% credible intervals, back-transformed to the proportion scale from the log-odds scale.

5. Results

5.1 Comprehension question accuracy

Figure 4 shows the by-condition accuracies for both our English and German experiments. Displayed next to the accuracies from our study are the comprehension question accuracies reported in Van Dyke (2007) (Experiments 2 and 3).
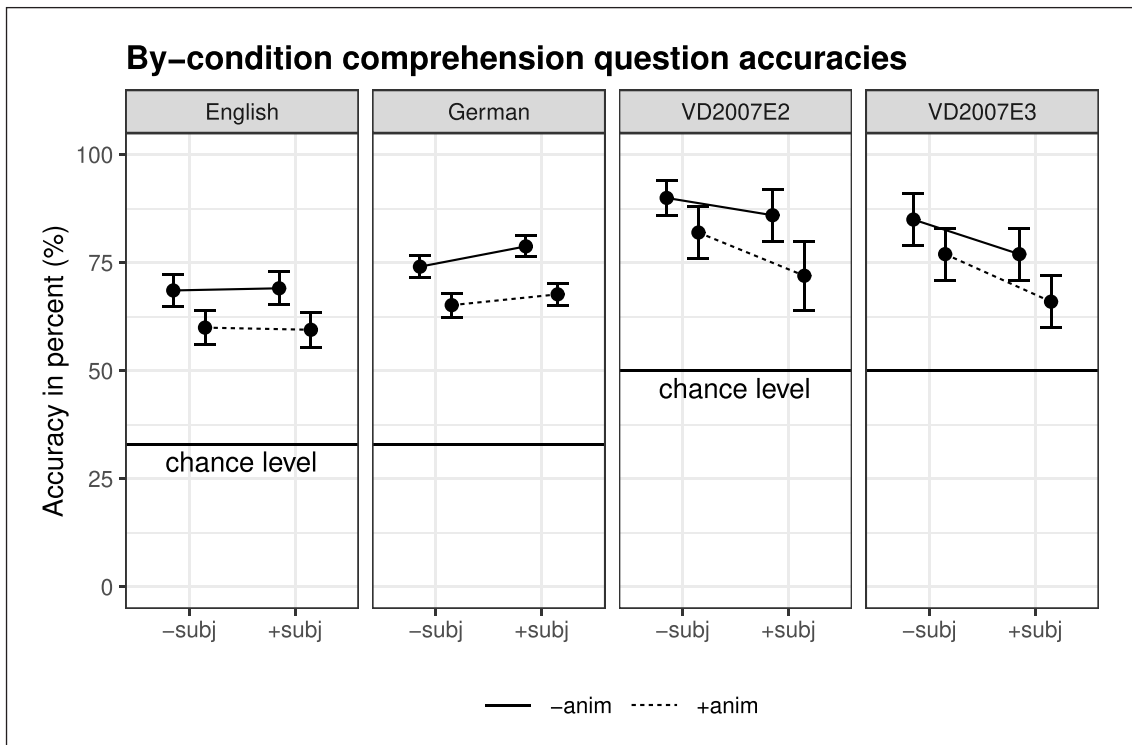
By–condition comprehension question accuracies

Figure displaying four panels labeled English, German, VD2007E2, and VD2007E3, showing accuracy in percent (%) versus –subj/+subj conditions, with –anim (solid line) and +anim (dotted line).

Figure 4: The two left-hand side plots display the data aggregated by condition: By-condition means and 95% confidence intervals for question response accuracy in percentages (%) in the English and the German experiment, respectively. The two right-hand side plots display estimates and 95% confidence intervals as reported in Van Dyke (2007) for comprehension question accuracy results (Experiment 2, labeled VD2007E2, and Experiment 3, labeled VD2007E3). ±subj: distractor is (not) a subject, ±anim: distractor is (not) animate. The difference in accuracy between the two sets of studies most likely arises from the fact that in our study participants had four response choices, whereas in the Van Dyke (2007) study participants had only two response choices.

Our study used a different question type than the Van Dyke (2007) study. In the Van Dyke (2007) experiments, the questions had a cloze format with two response choices (the target and one of the animate distractor NPs); chance-level accuracy in the Van Dyke data would be 50%. By contrast, our questions had four response choices: three NPs and the 'I don't know' option. If people picked one of the four options completely at random, then the chance level would be 25%. However, we would assume that if participants guessed, participants would pick one of the three NPs (NP1, NP2, or NP3), but not 'I don't know'. The chance-level accuracy in our data would therefore be around 33%. In Appendix A, we display the given responses by condition from our experiments.

Table 4 shows the English and German results for the statistical analysis of the accuracy data. In English, the most plausible values for the effect of SUBJECTHOOD are centered around

zero. In German, the 95% CrI for the SUBJECTHOOD effect ranges from 1% to 8%, suggesting, surprisingly, a somewhat higher comprehension accuracy when the distractor is also a subject.

Table 4: Comprehension question accuracies: Means of the posterior distributions with 95% credible intervals (CrIs) for the main effects of SUBJECTHOOD, ANIMACY and their interaction in English and German.

	Comprehension question accuracy	
	English	**German**
Fixed effects	**posterior mean [95% CrI]**	**posterior mean [95% CrI]**
SUBJECTHOOD	0% [–6, 6]%	4% [1, 8]%
ANIMACY	–10% [–15, –6]%	–11% [–16, –7]%
INTERACTION	–1% [–11, 8]%	–4% [–10, 2]%

In English, the main effect of ANIMACY has 95% CrI [–15, –6]%, suggesting that comprehension question accuracy is lower when the manipulated distractor is animate compared to when it is inanimate. Similarly, in German, the estimate ranges from –16% to –7%, consistent with a lower accuracy for animate distractor conditions, that is, semantic interference. This pattern for semantic interference was also reported in Van Dyke (2007).

5.2 Reading measure results

Figures 5 and **6** display the by-region and by-condition mean reading times with their 95% confidence intervals in first-pass reading times, regression-path durations, and total fixation times.

As discussed in Section 3, we analyze reading times at the pre-critical, critical, and post-critical regions. **Figure 7** shows the English and German effect estimates at the three regions, for first-pass reading times (FPRT), regression-path durations (RPD), and total fixation times (TFT). **Figure 8** shows the results for first-pass regressions out (FPR).

5.2.1 Pre-critical region (adverb)

5.2.1.1 English

For the English experiment, the most plausible values of the main effect of SUBJECTHOOD largely have a positive sign (FPRT 95% CrI [–4, 17] ms, RPD [15, 64] ms, TFT [4, 52] ms). In FPR, the main effect of SUBJECTHOOD has 95% CrI [3, 11]%. The positive sign of the effect suggests a reading time slowdown, and a higher proportion of first-pass regressions for [+subject] conditions.
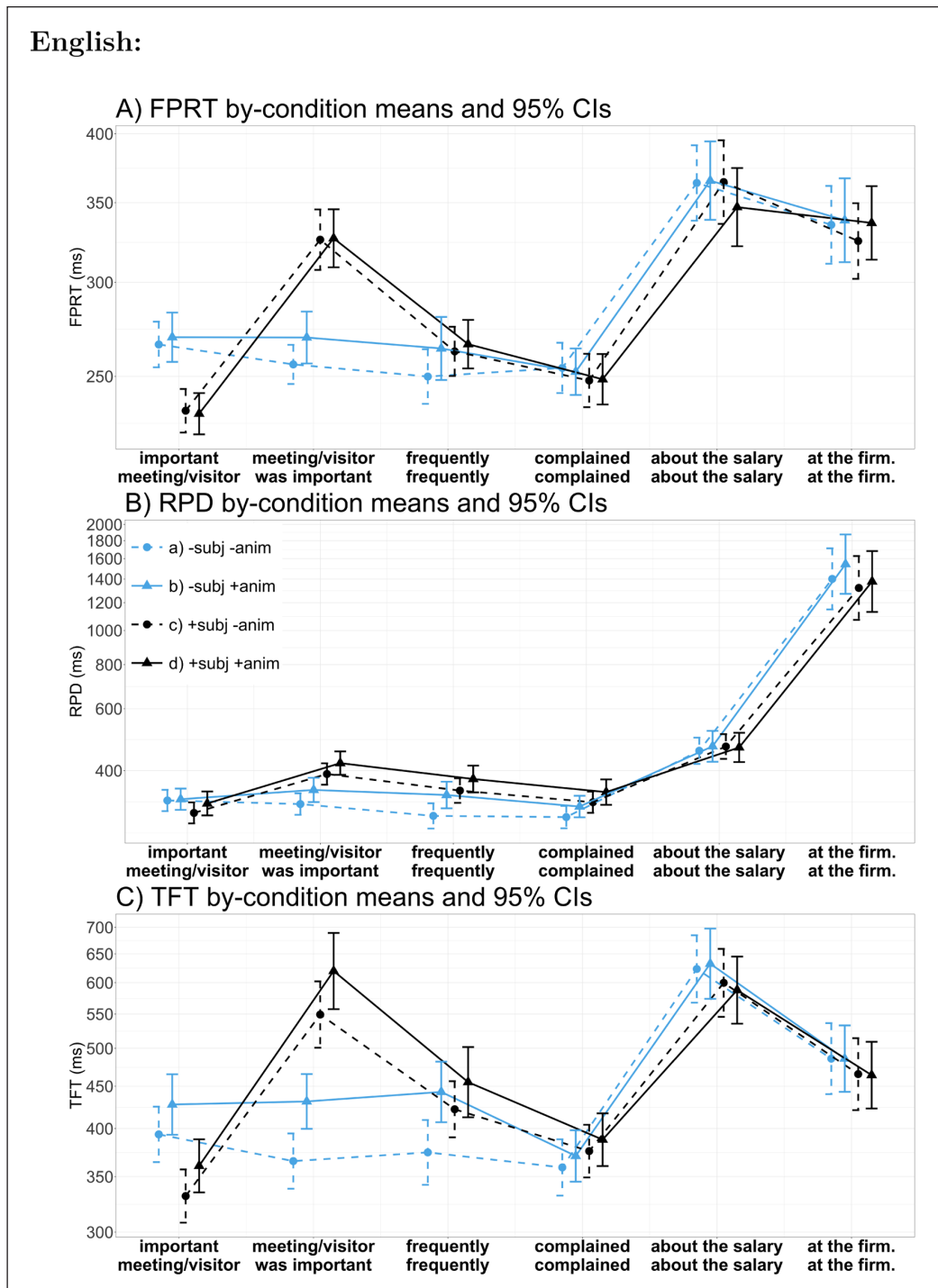
Figure 5: English: By-region plots: By-condition means with 95% confidence intervals (CIs) in A) first-pass reading times (FPRT), B) regression-path durations (RPD), and C) total fixation times (TFT). Shown are the reading times at the manipulated distractor and all following sentence regions. ±subj: distractor is (not) a subject, ±anim: distractor is (in)animate.

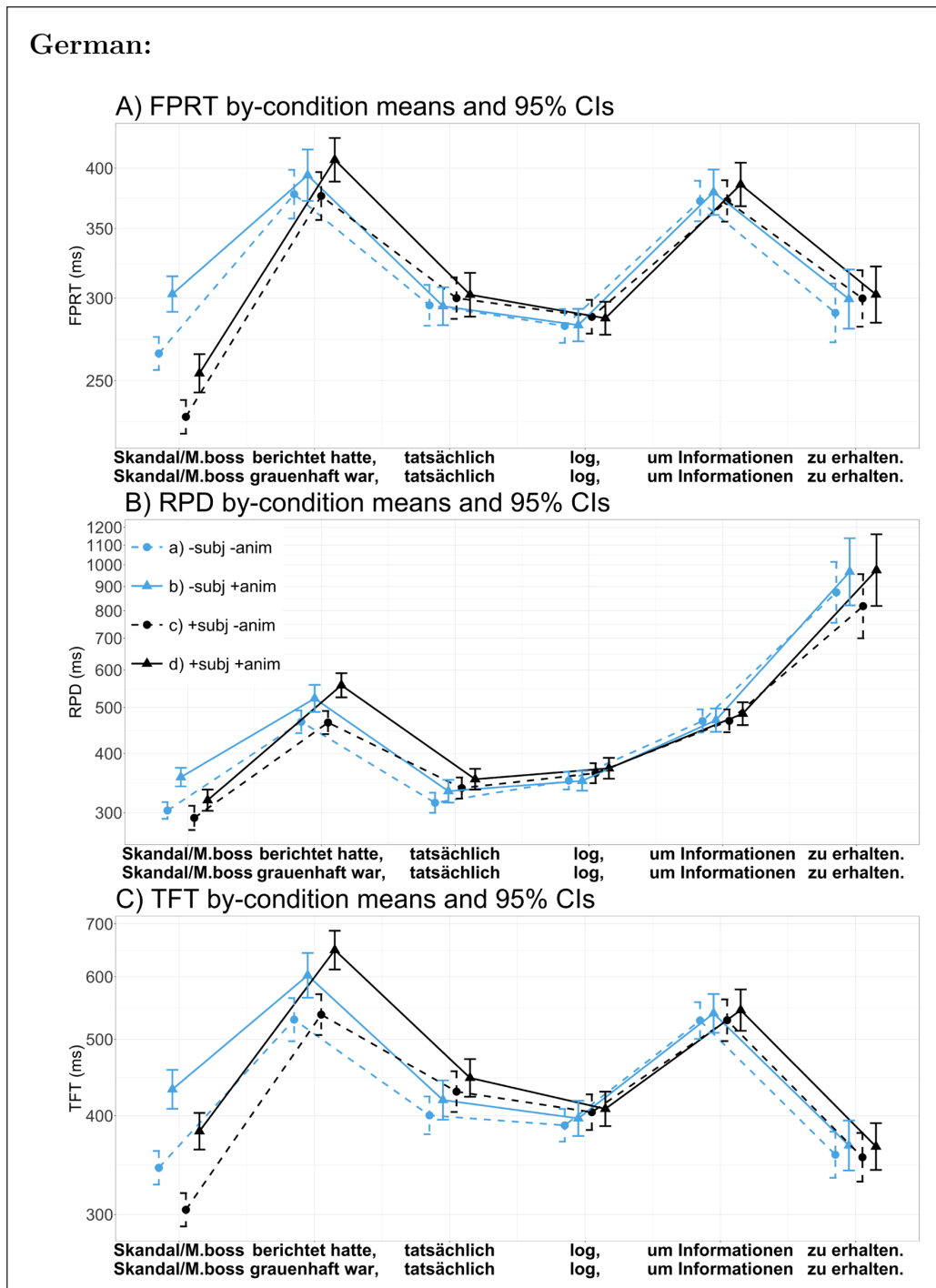**Figure 6:** German: By-region plots: By-condition means with 95% confidence intervals (CIs) in A) first-pass reading times (FPRT), B) regression-path durations (RPD), and C) total fixation times (TFT). Shown are the reading times at the manipulated distractor and all following sentence regions. ±subj: distractor is (not) a subject, ±anim: distractor is (in)animate.
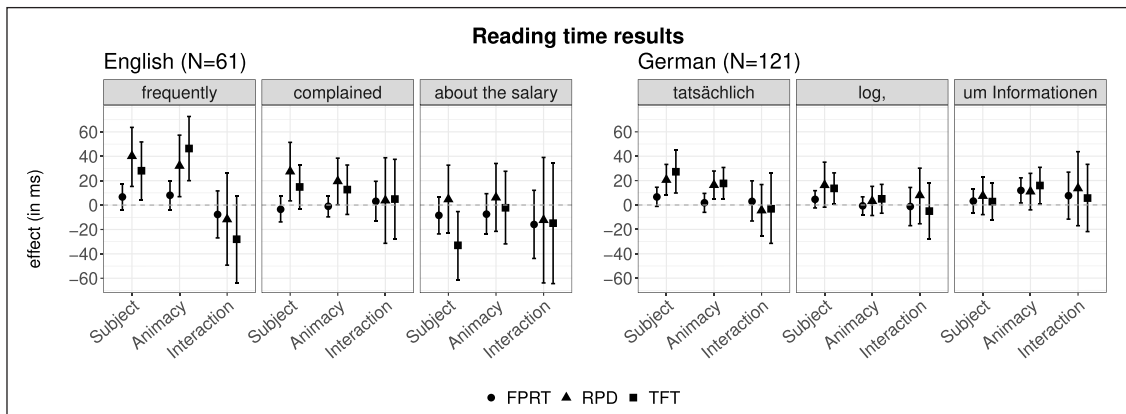
Figure 7: English (left panels): Posterior means with 95% credible intervals (CrIs) for the effects of SUBJECTHOOD, ANIMACY and their interaction at the pre-critical adverb (*frequently*), the critical verb (*complained*), and the post-critical region (*about the salary*). In German (right panels), the effect estimates are also shown at the pre-critical adverb (*tatsächlich*), the critical verb (*log,*), and the post-critical region (*um Informationen*). All values were back-transformed from the log-scale to the millisecond scale. FPRT = first-pass reading times, RPD = regression-path duration, TFT = total fixation times. Recall that a positive sign for the main effects of SUBJECTHOOD or ANIMACY indicates a reading time slowdown for [+subject/+animate] conditions, compared to [–subject/–animate] conditions.
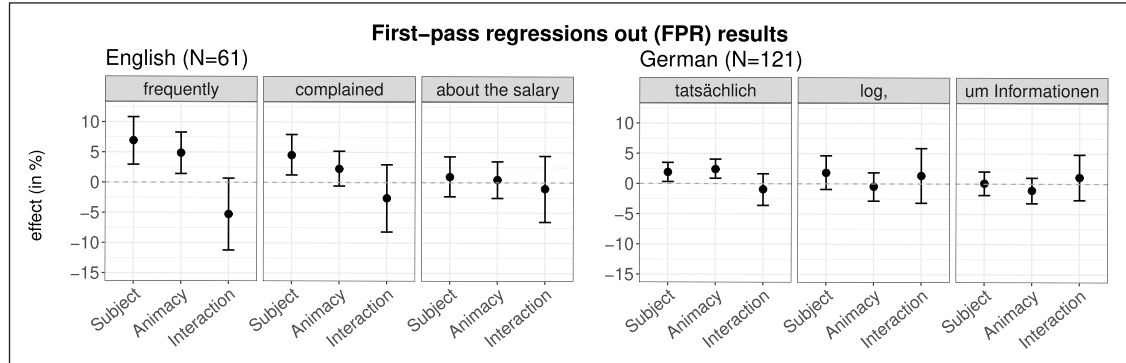

Figure 8: First-pass regressions out (FPR) results in English (left panels) and German (right panels): Posterior means and 95% credible intervals (CrIs) for the effects of SUBJECTHOOD, ANIMACY and their interaction at the pre-critical adverb, the critical verb, and the post-critical region. All values were back-transformed from the log-odds scale to percentages. Recall that a positive sign for the main effects of SUBJECTHOOD or ANIMACY indicates a higher proportion of first-pass regressions out for [+subject/+animate] conditions, compared to [–subject/–animate] conditions.

The main effect of ANIMACY indicates a reading time slowdown and more regressions in [+animate] distractor conditions (95% CrIs FPRT [–4, 20] ms, RPD [7, 57] ms, TFT [20, 73] ms, FPR [1, 8]%). Nested comparisons in RPD show that in [+subject] conditions, the 95% CrI

is [4, 66] ms, and in [–subject] conditions, it is [–7, 53] ms. The SUBJECTHOOD × ANIMACY interaction has the following 95% CrIs: FPRT [–27, 12] ms, RPD [–49, 26] ms, TFT [–64, 8] ms, FPR [–11, 1]%.

5.2.1.2 German

In German, we see a pattern similar to English. The main effect of SUBJECTHOOD mainly has a positive sign, consistent with a reading time slowdown and more regressions in [+subject] conditions (FPRT 95% CrI [–1, 15] ms, RPD [8, 33] ms, TFT [10, 45] ms, FPR [0, 3]%).

The effect of ANIMACY also suggests a reading time slowdown and a higher proportion of first-pass regressions in [+animate] conditions (RPD [5, 28] ms, TFT [5, 31] ms, FPR [1, 4]%). Nested comparisons for RPD show that in [+subject] conditions, the 95% CrI is [3, 34] ms, and in [–subject] conditions, it is [–3, 29] ms. The FPRT estimate is centered around zero.

In German, the interaction has the following 95% CrIs: FPRT [–13, 20] ms, RPD [–26, 17] ms, TFT [–31, 26] ms, FPR [–4, 2]%.

5.2.2 Critical region (verb)

5.2.2.1 English

The effect of SUBJECTHOOD has the following 95% CrIs: RPD [3, 51] ms, TFT [–3, 33] ms, and FPR [1, 8]%. This suggests a slowdown in reading times and a higher proportion of regressions in the [+subject] distractor conditions, consistent with syntactic interference.

For the main effect of ANIMACY, the RPD estimate has 95% CrI [0, 38] ms. Nested comparisons in RPD show that in [+subject] conditions, the 95% CrI is [–9, 40] ms, and in [–subject] conditions, the 95% CrI is [–10, 48] ms. The effect of ANIMACY in TFT ranges from –8 ms to 33 ms. In FPR, the interval ranges from –1% to 5%, suggesting more regressions out from the critical region when the distractor is animate. These results are consistent with a semantic interference effect. The SUBJECTHOOD × ANIMACY interaction shows 95% CrIs centered on zero (FPRT [–13, 19] ms, RPD [–31, 39] ms, TFT [–28, 37] ms, FPR [–8, 3]%).

5.2.2.2 German

In German, similar to English, the most plausible values of the effect of SUBJECTHOOD have a largely positive sign (95% CrI FPRT [–2, 12] ms, RPD [–2, 35] ms, TFT [1, 26] ms). This slowdown for [+subject] distractor conditions is consistent with a syntactic interference effect. There is also a somewhat higher proportion of regressions for [+subject] compared to [–subject] distractor conditions (95% CrI FPR [–1, 5])%). However, unlike English, the most plausible values of the effect of ANIMACY are centered around zero in all reading measures (FPRT [–8, 7] ms, RPD [–9, 15] ms, TFT [–7, 17] ms, FPR [–3, 2]%). For the SUBJECTHOOD × ANIMACY

interaction, the 95% CrIs are centered around zero (FPRT [–17, 14] ms, RPD [–15, 30] ms, TFT [–28, 18] ms, FPR [–3, 6]%).

5.2.3 Post-critical region

5.2.3.1 English

The English experiment shows an effect of SUBJECTHOOD in TFT, but with a negative sign (95% CrI [–61, –5] ms). For the other measures, the most plausible values of the effect of SUBJECTHOOD are centered around zero (FPRT [–24, 7] ms, RPD [–23, 33] ms, FPR [–2, 4]%). Similarly, the effect estimates for the main effect of ANIMACY are centered on zero (FPRT [–24, 9] ms, RPD [–22, 34] ms, TFT [–32, 28] ms, FPR [–3, 3]%). That is, there is no indication of a reading time slowdown for [+animate] distractor conditions at this region. The 95% CrIs for the SUBJECTHOOD × ANIMACY interaction are also centered on zero (FPRT [–44, 12] ms, RPD [–64, 39] ms, TFT [–64, 34] ms, FPR [–7, 4]%).

5.2.3.2 German

In the German experiment, the reading time patterns differ from those in the English experiment. In all reading measures, the estimated effect of SUBJECTHOOD is centered on zero (95% CrI FPRT [–7, 13] ms, RPD [–8, 23] ms, TFT [–12, 18] ms, FPR [–2, 2]%). However, the main effect of ANIMACY has a mainly positive sign (95% CrI FPRT [2, 22] ms, RPD [–4, 26] ms, TFT [1, 31] ms). This suggests that there is a reading time slowdown for [+animate] conditions at the post-critical region in German. Nested comparisons in RPD show that in [+subject] conditions, the 95% CrI is [–18, 25] ms, and in [–subject] conditions, it is [–4, 41] ms. For FPR, the estimate of the ANIMACY effect ranges from [–3 to 1]%. The interaction shows the following 95% CrIs: FPRT [–12, 27] ms, RPD [–17, 44] ms, TFT [–22, 33] ms, FPR [–3, 5]%.

6. Discussion

The offline accuracy data from both English and German showed reduced comprehension accuracy when the distractor is animate, suggesting a lasting effect of semantic interference, consistent with the findings in Van Dyke (2007). In German, a small increase in accuracy (1–8%) was seen when the distractor was a subject. Descriptively, this positive effect of SUBJECTHOOD on response accuracy is driven mainly by comprehension questions about NP2 and NP3 (see the means in Appendix A). Speculatively, the presence of an additional clause boundary within the relative clause in the [+subject] conditions may have made these two noun phrases more distinct and/or salient.

Of central interest to the present study are the online reading data. In the English experiment, the pre–critical region showed reading time slowdowns and more regressions for the [+subject]

and [+animate] conditions. The critical region also showed reading time slowdowns and more regressions for the [+subject] and [+animate] conditions, consistent with syntactic and semantic interference. The post–critical region showed only an unexpected syntactic effect, a reading time speed–up for [+subject] distractor conditions. This may reflect a recovery from the processing difficulty at the previous regions. Previous work has reported such speed–ups, attributing them to "… readers trying to make up for lost time after having been slowed down." (Paape et al., 2018, p. 39).

Similar to the English data, in German the pre-critical region showed reading time slowdowns and a higher proportion of regressions in the [+subject] and [+animate] conditions. The critical region exhibited a pattern that is consistent only with a syntactic interference effect. In contrast to English, the critical region did not show any indication of a semantic interference effect. In German, the post-critical region exhibited a reading time slowdown in [+animate] distractor conditions that is consistent with semantic interference.

Next, we discuss the effects in the pre-critical, and critical and post-critical regions separately.

6.1 Pre-critical region effects

Pre-critical effects are surprising given the assumptions of the Lewis and Vasishth (2005) model. **Figure 7** shows reading time patterns in both languages that could be indicative of syntactic and semantic effects arising simultaneously prior to the critical region. There are four possible explanations for the effects in the pre-critical region.

6.1.1 Explanation 1: Potential confounds

Recall that the pre-critical adverb was added to "absorb" potential clause boundary effects. Therefore, one possible explanation is that the pre-critical syntactic effects are a consequence of processing two versus one clause boundaries, such that the processing of two clause boundaries led to the reading time slowdown (Arnett & Wagers, 2017; Wagers, 2008). However, this would only explain the syntactic, but not the semantic effects at the pre-critical adverb.

It is also possible that interference effects from the previous regions spilled over to the pre-critical adverb. Speculatively, interference effects likely also occur on the verbs in the embedded sentences, and may contribute to the syntactic and semantic effects on the pre-critical adverb. Because there are lexical and structural differences prior to the pre-critical adverb, we do not statistically analyze these regions, but the reading time measures show large differences between conditions (see **Figures 5** and **6**).

Although it can't be ruled out that wrap-up effects or spillover effects contribute to the increased reading times on the adverb, it seems more likely that other factors, such as encoding interference or predictive processing (discussed below), influence the reading time patterns

observed for both the syntactic and the semantic manipulation across both languages in this region.

6.1.2 Explanation 2: Parafoveal-on-foveal effects

It is theoretically possible that the effects on the pre-critical adverb are a consequence of parafoveal processing. Specifically, the observed reading time slowdowns could be a consequence of parafoveal-on-foveal (POF) effects: fixation durations on a word n (here, the adverb) can be affected by the properties of the next word $n+1$ (here, the critical verb) (Rayner et al., 2003). POF effects have predominantly been shown for low–level—orthographic and lexical—properties of the word $n+1$ (Inhoff & Rayner, 1986; Kennedy & Pynte, 2005; Kennedy et al., 2002; Rayner, 1975; Vitu et al., 2004), although the evidence for POF effects is often interpreted as conflicting (e.g., Angele et al., 2008; Angele et al., 2013; Henderson & Ferreira, 1993; Hyönä & Bertram, 2004; Risse & Seelig, 2019). It remains widely debated under what experimental conditions POF effects arise, and what causes them (Angele et al., 2008; Drieghe, 2011; Risse & Kliegl, 2012, 2014; Schotter et al., 2012). The POF explanation seems unlikely for the syntactic and semantic effects we observe in our study, because there is no support in the eye–tracking literature for syntactic or semantic information in the parafovea influencing the processing of a fixated word n (e.g., Hyönä, 2012; Inhoff & Rayner, 1980; Inhoff, 1982; Staub et al., 2007; but see the co-registration study by López-Peréz et al., 2016). For this reason, we propose that there are more plausible explanations for the pre–critical syntactic and semantic effects in our data, which we present next.

6.1.3 Explanation 3: Predictive processing

One possible explanation for the simultaneous syntactic and semantic effects in the pre–critical region in both languages is predictive processing (Jäger et al., 2015b; Levy, 2008; Levy & Keller, 2013). The cue-based retrieval model as implemented by Lewis and Vasishth (2005) specifies a left-corner parsing algorithm that follows X-bar rules (Chomsky, 1986) for incremental syntactic structure building. The left-corner parser operates according to bottom–up and top-down principles (Aho & Ullman, 1972). Assuming a context-sensitive phrase structure grammar, once the left corner of the right–hand side of a phrase structure rule is identified, the upcoming structure is predicted (Brasoveanu & Dotlacil, 2020, Chapter 4). For our critical verb phrase (VP) *frequently complained*, once the adverb has been identified, the *VP* is already predicted, possibly allowing a subject retrieval to be triggered at *frequently*.

However, the predictive processing proposal would require additional assumptions to explain the pre-critical semantic effects. The adverb would need to set semantic retrieval cues when the identity of the verb is still unknown. One possibility is that the processor has a probabilistic expectation for an upcoming verb with an animacy cue, given that subjects are frequently

animate (Bock & Warren, 1985; Clark & Begun, 1971). Thus, the semantic cue might be put to use to retrieve an encoding from memory that is the likely subject even before the exact identity of the verb is known.

Although predictive processing might be a plausible explanation for the pre-critical effects, there is an alternative explanation for the reading time patterns observed in our data, namely, encoding interference.

6.1.4 Explanation 4: Encoding interference

The observed slowdowns caused by matching distractors could be a consequence of encoding interference, that is, the faulty encoding of one (or more) noun phrases in memory (Oberauer & Kliegl, 2006). The memory model of Oberauer and Kliegl (2006) assumes that items in memory are represented by feature bundles. These memory items – such as the target and distractor nouns in our study – will compete for shared features. This can have detrimental effects on the quality of the representations as features on representations can be lost (feature overwriting; Lange & Oberauer, 2005; Nairne, 1990; Neath, 2000). The degraded representations reduce the items' overall activation, increasing the processing time it takes to activate a target item among competing items. Villata et al. (2018) call this the leveling effect. However, leveling on its own does not account for the interference effects observed at the pre-critical region because reduced activation should only affect retrieval, which is assumed to occur at the verb (see Yadav et al., 2023). One would need to add the assumption that competition between noun phrases for one or more features arises at the competitor and proceeds continuously, possibly slowing down processing at the pre-critical region (Lago et al., 2021; Villata et al., 2018). This would still not fully explain why the effects should be detected specifically at the pre-critical region, as opposed to the embedded distractor noun phrase (the point of encoding) or the verb (the point of retrieval). One possibility is that the encoding difficulty is, in fact, triggered at the point of encoding the distractor noun phrase into memory, and the reading slowdowns at the pre-critical region reflect spillover processing from this difficulty. It is difficult to put this hypothesis to a convincing test with our current data, because of the substantial lexical and structural differences across the region of encoding. It is also possible that encoding interference simply impacts reading times at a delay relative to the point of encoding; evaluating this possibility would require developing an explicit linking hypothesis between encoding difficulty and the time course of processing difficulty. Nevertheless, to our minds, encoding interference is a plausible explanation for the effects in the pre-critical region. Because the memory representations can compete for syntactic as well as semantic features, encoding interference can explain both the syntactic and semantic effects at the pre-critical adverb.

Both predictive processing and encoding interference seem plausible to explain the reading time patterns at the pre-critical adverb. There are some findings in the literature that

are compatible with the proposal that the effects may be due to memory retrieval driven by predictive processing. For example, pre-verbal structure building has been shown in a number of verb-final languages (e.g., Aoshima et al., 2004; Bader & Lasser, 1994; Kamide & Mitchell, 1999; Konieczny, 2000; Vasishth & Lewis, 2006), as well as English, a verb-medial language Omaki et al. (2015). For instance, Omaki et al., (2015) found that in filler-gap dependency resolution, the parsing mechanism actively makes a prediction of the gap position prior to accessing the verb properties. These findings are compatible with a view that a subject retrieval may occur at a pre-verbal modifier.

Another study that seems to be compatible with the hypothesis that subject retrieval may be initiated pre-critically is Wagers and McElree (2009). In two speed-accuracy tradeoff studies, the authors showed that in sentences such as *The officer was informed that the driver (abruptly) fainted*, the presence of VP-level adverbs resulted in a processing speed-up on the verb, compared to when no modifier was present (the speed-up was replicated in a follow-up experiment; Wagers, p.c., and as cited in Wagers and McElree, 2022). This was observed for conditions with VP-level adverbs, but no such difference was observed for S-level adverbs (evaluative or epistemic modality adverbials). Wagers and McElree (2009) proposed that this speed-up may demonstrate that verb-processing is given a "head start" in adverbial conditions.

There are also findings in the literature that are compatible with the encoding interference explanation. As mentioned earlier, Van Dyke (2007) (Experiment 3) reported a semantic interference effect at the pre-critical region. In addition, a recent study by Lago et al. (2021) observed that interference effects emerged pre-critically in subject-verb number agreement dependencies.

It is possible that both memory retrieval driven by predictive processing as well as encoding interference are driving the pre-critical effects. Recent computational modeling work has independently accounted for encoding interference and cue-based retrieval processes during sentence processing (Yadav et al., 2023). Yadav et al., (2023) reported, for subject-verb number agreement dependencies, that their encoding-plus-retrieval model can better capture observed empirical interference effects than the assumptions in the Lewis and Vasishth (2005) cue-based retrieval model. Due to the exploratory nature of the findings in the present study, it will be important to try to replicate this pattern in a future study, and attempt to determine the source of the effect on the pre-verbal region.

6.2 Critical and post-critical region effects

As discussed above, a pattern common to the English and the German data is syntactic and semantic interference effects appearing in the pre-critical region. In the critical verb region, both languages also show reading time slowdowns consistent with syntactic interference. However, the semantic interference patterns on the critical region differ in English compared to German:

English shows a reading time slowdown consistent with semantic interference, but German does not. In German, the post-critical region shows a slowdown consistent with semantic interference, but English does not.

Can we interpret the differences as systematic? This question can only be investigated by computing the estimates of the interaction between semantic interference and language (Nieuwenhuis et al., 2011). In order to check whether there is any indication of an interaction, we combined the English and German data and then looked at the posterior distributions (and 95% credible intervals) for the coefficient representing the interaction between semantic interference and language.

The results of this analysis for the critical and post-critical regions are shown in **Table 5**. In the critical region, where only English shows a semantic effect, we observe estimates that largely have a negative sign. In the post-critical region, the sign of the interaction reverses, because the semantic interference effect appears in German but not in English. The estimates at the critical region are small, and only the RPD estimate might be suggestive of a small difference between English and German in the time course of semantic interference effects. At the post-critical region, the TFT estimate has a largely positive sign, indicating a small difference between English and German with regard to the semantic interference effect: a post-critical slowdown compatible with semantic interference in German, but not in English.

Table 5: The means of the posterior distributions (in ms) of the interaction between the semantic interference effect and the language (English vs. German), along with 95% credible intervals. Shown are the estimates for the critical and post-critical regions.

region	DV	estimate	lower	upper
crit	RPD	–9.39	–19.78	0.98
	TFT	–4.52	–15.18	5.93
postcrit	RPD	1.84	–11.96	15.57
	TFT	9.55	–4.20	23.47

If the differences between English and German are systematic, one plausible explanation for the delayed effect of semantic interference in German could be the presence of case marking in the German items: overt case marking cues are absent in English, but in German the determiner in a noun phrase carries case marking. Because German noun phrases contain overt case morphology, it is likely that syntactic cues could dominate in determining retrieval at the critical region in German. This speculation could be tested by investigating interference in other languages that have overt case marking. Our German results highlight the importance of cross-linguistic investigations of interference effects in psycholinguistics.

If case cues are driving the delay in semantic interference effects in German, an interesting implication is that, at the critical region (the verb), most individual German readers should show semantic interference effects centered around zero in the reading time measures, whereas most English speakers should show positive effects. To explore this possibility, we extracted posterior distributions of the individual-level parameter estimates for semantic interference for regression-path duration and total fixation time, in both English and German.[9] The individual-level estimates (along with 95% credible intervals) are shown in **Figure 9**.
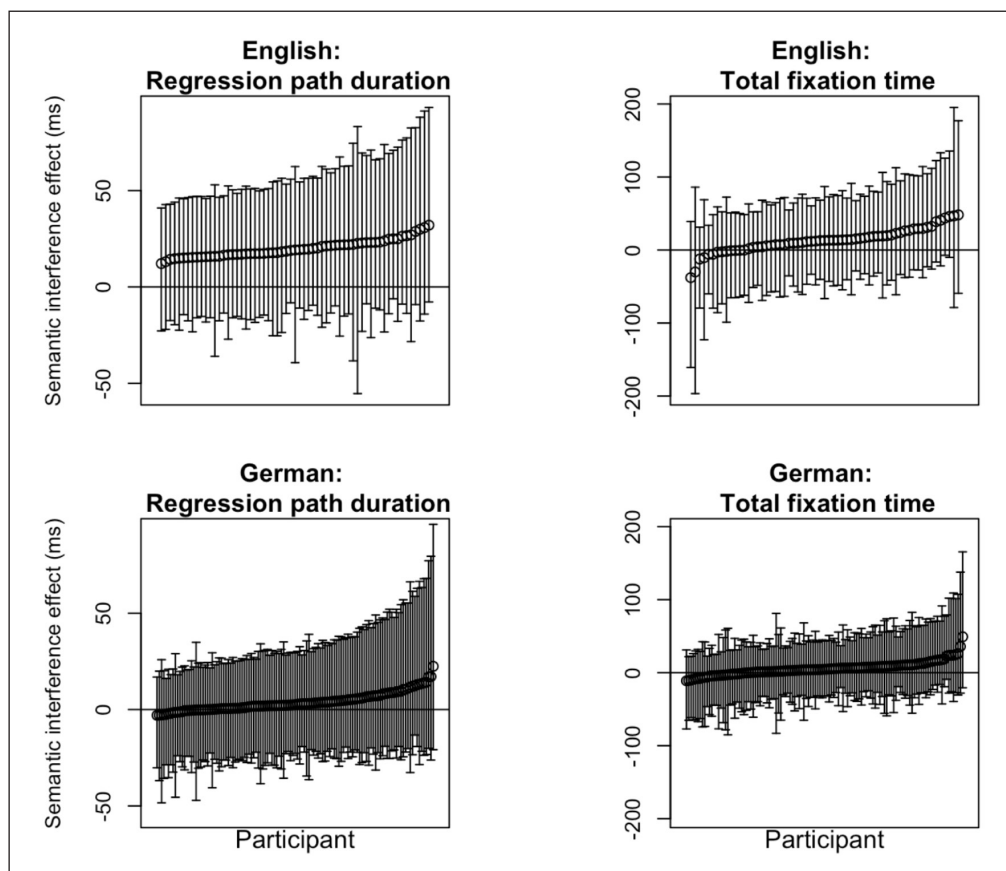

Figure 9: Individual-level estimates (with 95% credible intervals) for English and German regression-path duration and total fixation time in the critical region. Each dot represents the (shrunken) estimate of the mean semantic interference effect of each individual participant, and the 95% credible intervals show the uncertainty of the estimate.

---

[9] A unique advantage of the Bayesian linear mixed modeling framework is that the participant-level adjustments to the fixed effects are parameters in the model; this makes it possible to derive the posterior distribution of each participant's effect. The approach is to extract the posterior distributions of the fixed and random effects from the model, and then back-transform them to reading times in milliseconds for individual participants (see Nicenboim et al., 2023, for details).

Figure 9 does in fact suggest that, in regression-path duration, most German speakers are showing effects centered around zero; a few participants show non-zero effects with positive means, consistent with semantic interference effects. By contrast, in English, in regression-path duration, all speakers show positive mean effects. In total fixation time, the individual-level estimates are qualitatively and quantitatively similar in English and German. The semantic effect that was present in English regression-path durations seems to disappear in total fixation times. It is possible that re-reading drives the effect (total fixation times are the sum of first-pass reading times and re-reading times), and that during later reading stages, and at subsequent regions, the online processing difficulty is attenuated. The uncertainties around individual participants' estimates are large; this is because we have relatively few data points from each participant in each condition (10 data points per condition in both English and German). The individual-level patterns in regression-path duration are suggestive of qualitative differences in the behavior of English vs. German speakers. Independent support for this idea would come from a replication attempt of our experiment: if the regression-path duration pattern in **Figure 9** can be replicated (ideally with a much larger number of data points per participant), that would be a convincing validation of our speculation that case marking in German may be driving the absence of a semantic interference effect in that language at the critical region.

7. General discussion

To our knowledge, our study is the first cross-linguistic investigation of whether syntactic and semantic interference effects arise simultaneously during online dependency formation. To establish whether there is cross-linguistic support for syntactic and semantic interference during retrieval, two eye-tracking experiments tested similarity-based interference in English and German. The German study is also the largest-sample study to date on retroactive interference in sentence comprehension. Both languages were tested with the same experimental method and design, as well as similar syntactic constructions, namely, subject-verb dependencies that have been widely investigated in the similarity-based interference literature.

In both languages, we saw indications of semantic interference in the offline accuracies and in reading times. Both languages showed online support from reading that is consistent with syntactic and semantic interference effects. Our data thus contribute to the large body of evidence on syntactic and semantic interference effects during online dependency resolution (e.g., Arnett & Wagers, 2017; Cunnings & Sturt, 2018; Dillon et al., 2013; Laurinavichyute & von der Malsburg, 2022; Lowder & Gordon, 2014; Tabor et al., 2004; Van Dyke & Lewis,.2003; Van Dyke, 2007; Van Dyke & McElree, 2011).

At the pre-critical region, syntactic and semantic effects emerged simultaneously in both English and German. At the critical region, the two languages diverged: English continued to show syntactic and semantic interference effects, but German only showed patterns consistent

with syntactic interference. In German, the post-critical region showed indications of a delayed effect of semantic interference.

In Section 6, we proposed that the pre-critical syntactic and semantic effects could be driven by encoding interference and/or predictive processing which would be compatible with previous interference work (e.g., Lago et al., 2021; Smith et al., 2021; Van Dyke, 2007; Yadav et al., 2023).

At the critical verb, where retrieval is assumed to occur, the English data suggest that both types of retrieval cues can in principle be used simultaneously. This is compatible with the default assumption of the cue-based retrieval model, and in line with constraint-based accounts. By contrast, the timing lag in the German data might indicate that the assumption of simultaneous use of cues does not occur in all contexts: it might vary across languages and/or constructions. The timing lag in German might suggest a dominant effect of syntactic cues during retrieval that is present in German. An exploratory analysis of the ANIMACY × REGION (critical vs. post-critical) interaction might be suggestive of a semantic effect at the post-critical but not the critical region in first-pass reading times, although the estimates of two out of three reading measures include negative values (95% CrIs FPRT: [0, 23] ms, RPD [–13, 25] ms, TFT [–10, 27] ms). If, in German, syntactic cues can have a dominant effect, this would be in line with syntax-first accounts, which assume that syntactic information takes priority over semantic information during processing. This includes the proposals that syntactic cues may "gate" semantic cues during retrieval, or that in some configurations, they may be weighted more highly, or take complete precedence over non-structural cues (Cunnings & Sturt, 2014; Dillon et al., 2013; Kush, 2013; Sturt, 2003; Van Dyke & McElree, 2011; Yadav et al., 2022). The proposal that overt case marking might lead to syntactic cues taking precedence in our experiment setup should be investigated systematically in future work.

How do our data compare to the findings on syntactic and semantic interference in Van Dyke (2007)? Our findings partially align with the findings in the Van Dyke (2007) study: Both our English and German experiments, as well as Van Dyke's Experiment 3, observed pre-critical reading time slowdowns. Van Dyke ascribed the slowdowns in the pre-critical region to a difference in between-condition plausibility observed in a pretest. However, these effects may also have been a consequence of encoding interference. At the critical region, our English findings are partially compatible with the findings in the Van Dyke (2007) study. Similar to our English experiment, the summary of the Van Dyke estimates in **Figure 1** shows that Experiment 2 had slowdowns compatible with syntactic and semantic interference. However, Experiment 3 is more compatible with our German data: the critical region showed only a slowdown compatible with syntactic interference while post-critically, a slowdown compatible with semantic interference was observed. Given these later effects of semantic interference in addition to the semantic interference effects in offline accuracy in both of our and Van Dyke's experiments, it is possible that semantically similar distractors can continue to affect processing and have a lasting detrimental effect on overall comprehension.

Our findings at the critical region on semantic interference also partially corroborate the effects observed in Van Dyke and McElree (2011). Both our English study and Van Dyke and McElree (2011)'s Experiment 1B observed slowdowns consistent with semantic interference at the retrieval point. Van Dyke and McElree (2011)'s Experiment 2B showed no indication of semantic interference when distractors were in object position. While the findings of Experiments 1B and 2B may be suggestive of semantic interference only occurring when the distractor additionally matches the syntactic cue, this question remains unresolved given that there was no indication of an interference × experiment interaction (see Section 1).

Our results can contribute only in a limited way to this debate: the data indicate that there may be gating in specific linguistic contexts. Our English data at the critical region are not compatible with the gating proposal. The interaction estimates at the critical region are centered on zero. These data suggested that the reading time slowdowns consistent with semantic interference are comparable in [+subject] and [–subject] conditions (RPD 95% CrIs [–9, 40] ms, and [–10, 48] ms, respectively). This could be an indication that semantic interference can arise for distractors that do not additionally match the syntactic cue, at least for some languages or syntactic structures. In our German data, at the critical region, we neither observed a reading time slowdown compatible with semantic interference nor an interaction estimate that would suggest semantic interference only from distractors that additionally match the syntactic cue. However, the timing lag in German (discussed above) might suggest that syntactic cues are evaluated before semantic cues. Further investigation is needed to address this question.

Do our results align with the quantitative predictions of the cue-based retrieval model as shown in Section 3? We compare the model predictions (95% credible intervals) for syntactic and semantic interference with the estimates from our English and German experiments (95% credible intervals), and with the estimates from the original Van Dyke (2007) study (95% confidence intervals). We focus only on the critical region because, strictly speaking, the model's predictions are for this region only: the retrieval of the subject NP should be triggered when the verb is read. In order to fully interpret interference effects that occur before or after the verb, a more sophisticated understanding of the eye-parser relationship is needed, which is beyond the current capabilities of the Lewis & Vasishth model (see Rabe et al., 2023).

As stated in Section 3, we use the region of practical equivalence (ROPE) approach for the comparison of the model predictions with the empirical estimates (Freedman et al., 1984; Kruschke, 2015; Spiegelhalter et al., 1994). Recall that a partial overlap of the estimates indicates that the data and the model predictions have some degree of consistency. A perfect overlap between the data and the predictions indicates strong consistency whereas no overlap suggests that the data are not consistent with the model predictions. The empirical estimates may also be

so large that they subsume the model predictions, which would be considered an uninformative outcome.

Figure 10 shows that for our English data, the empirical RPD and TFT estimates for the effects of SUBJECTHOOD (syntactic interference) and ANIMACY (semantic interference) largely overlap with the model predictions. For the German data, the empirical RPD and TFT estimates for syntactic interference effects are also consistent with the model predictions. However, the semantic effect estimate does not match the model predictions in any measure. In sum, most of our empirical findings from English and German lie within the range of plausible values predicted by the model, with the exception of the semantic effect in German.
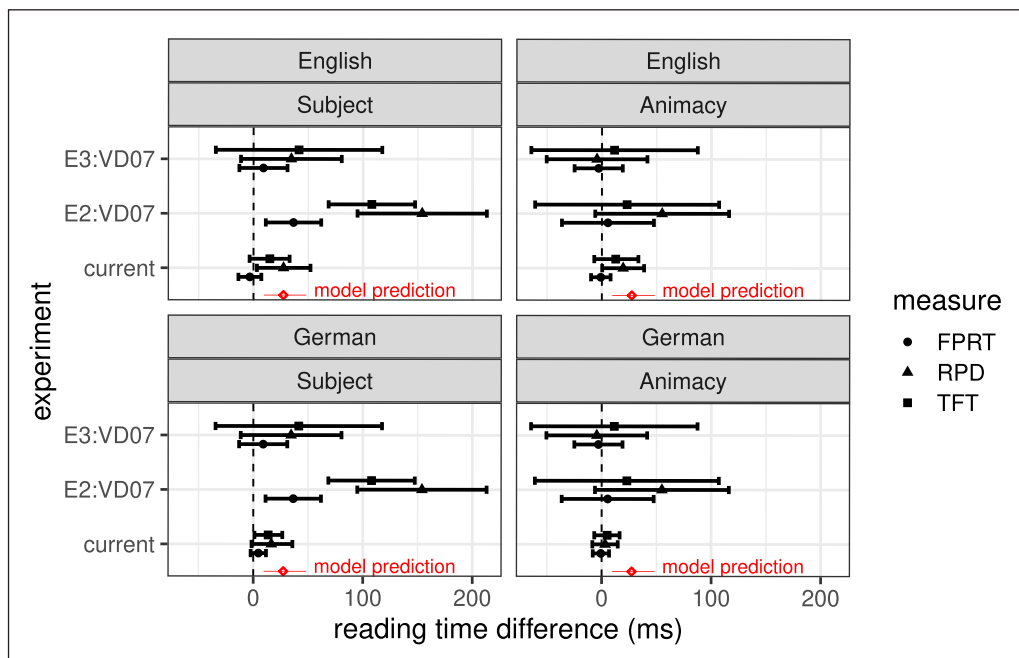

Figure 10: Shown are the prior predictions (95% credible intervals) from the cue-based retrieval model (in red) compared to the observed effect estimates from our study and the Van Dyke (2007) study. Shown are the observed effect estimates at the critical region in first-pass reading times (FPRT), regression-path durations (RPD), and total fixation times (TFT). The upper left panel shows the English estimates (posterior means with 95% credible intervals) of our study (= current) for the effect of SUBJECTHOOD (syntactic interference) as well as the Van Dyke (2007) means and 95% confidence intervals of Experiments 2 and 3 (E2:VD07 and E3:VD07, respectively). The upper right panel shows our English estimates and the Van Dyke (2007) estimates for the effect of ANIMACY (semantic interference). The lower left panel shows our German estimates (= current) and the Van Dyke estimates for syntactic interference. The lower right panel shows our German estimates and the Van Dyke (2007) estimates for semantic interference. The Van Dyke (2007) estimates shown in the upper panels are duplicated in the lower two panels next to our German data for easier comparability.

As discussed in detail in Section 3, **Figure 10** also shows that for Van Dyke (2007)'s Experiments 2 and 3, most of the empirical estimates match the model predictions. However, the intervals from the original study subsume the estimates from our experiments and the model estimates. The intervals from the original study are too wide to be informative. Overall, the graphical summary of the predictions and data highlights an important observation that was made in previous work (Jäger et al., 2015a; Jäger et al., 2020; Nicenboim et al., 2018; Vasishth, 2023; Vasishth et al., 2023; Vasishth & Engelmann, 2021; Vasishth et al., 2019; Vasishth & Gelman, 2021; Vasishth et al., 2018): there is an urgent need for higher-powered studies of the interference effect (as well as other phenomena) that yield more precise estimates of the effects.

Overall, our offline data from English and German show patterns consistent with semantic interference impeding comprehension. Our online data from two languages provide additional support for syntactic and semantic similarity-based interference in online processing. Surprisingly, both German and English show pre-critical syntactic and semantic effects that could be the result of encoding interference and/or predictive processing. Our study design does not allow us to distinguish between the two, and further investigation is necessary. At the critical region of interest and the post-critical regions, the reading time patterns differ across the two languages. Our English online data are compatible with the claim that both types of interference can arise simultaneously at the critical retrieval site, suggesting that syntactic and semantic retrieval cues can be used simultaneously. However, the German data show a different pattern, indicating that syntactic cues can precede semantic cues during online sentence comprehension. These cross-linguistic differences may arise from the amount of morphosyntactic information available in a particular language.

8. Conclusion

This is the first cross-linguistic investigation (English and German) that presents support for syntactic and semantic interference effects in sentence comprehension. Both languages reveal syntactic and semantic effects on a pre-verbal modifier that are compatible with encoding interference and/or predictive processing effects. The reading time patterns observed at the critical verb region suggest cross-linguistic differences: Our English data suggest that both types of interference can arise simultaneously during retrieval, in line with the cue-based theory's predictions. However, in German, a language with richer morphological marking than English, syntactic cues may take precedence over semantic cues. Additionally, in both languages, our offline comprehension question data suggest that semantic interference can adversely affect overall comprehension.

Appendix A

Question responses

Recall that a third of questions in our experiments targeted NP1 (the target, e.g., *the attorney*), one third targeted NP2 (the distractor, e.g., *the secretary*), and another third targeted NP3 (the manipulated distractor, e.g., *visitor/meeting*). Interestingly, when the target NP1 (e.g., *the attorney*) was the correct response, participants frequently erroneously chose NP2 (e.g., *the secretary*), and vice versa. This tendency seems to be stronger in English than in German.

Table 6: English: Given responses (in %) by condition for questions that targeted A) NP1 (*the attorney*), B) NP2 (*the secretary*), or C) (*the meeting/visitor*).

English:				
A) Responses in percent (%) to questions targeting NP1 (the attorney)				
	NP1	NP2	NP3	'I don't know'
	(attorney)	(secretary)	(visitor)	(?)
Condition	**correct**	incorrect	incorrect	incorrect
a. −subject, −animate	60	39	0	1
b. −subject, +animate	53	38	7	2
c. +subject, −animate	58	39	0	3
d. +subject, +animate	49	35	12	4
B) Responses in percent (%) to questions targeting NP2 (the secretary)				
	NP1	**NP2**	**NP3**	**'I don't know'**
	(attorney)	(secretary)	(visitor)	(?)
Condition	incorrect	**correct**	incorrect	incorrect
a. −subject, −animate	27	71	0	2
b. −subject, +animate	26	59	9	6
c. +subject, −animate	29	70	0	1
d. +subject, +animate	23	64	10	3
C) Responses in percent (%) to questions targeting NP3 (the visitor/meeting)				
	NP1	**NP2**	**NP3 (meeting/**	**'I don't know'**
	(attorney)	(secretary)	visitor)	(?)
Condition	incorrect	incorrect	**correct**	incorrect
a. −subject, −animate	10	8	76	6
b. −subject, +animate	12	10	68	10
c. +subject, −animate	10	7	80	3
d. +subject, +animate	17	13	66	4

Table 7: German: Given responses (in %) by condition for questions that targeted A) NP1 (*the journalist*), B) NP2 (*the colleague*), or C) (*the scandal/mafia boss*).

German:				
A) Responses in percent (%) to questions targeting NP1 (the journalist)				
	NP1	**NP2**	**NP3**	**'I don't know'**
	(journalist)	(colleague)	(mafia boss)	(?)
Condition	**correct**	incorrect	incorrect	incorrect
a. –subject, –animate	84	14	0	2
b. –subject, +animate	76	15	6	3
c. +subject, –animate	86	13	0	1
d. +subject, +animate	77	11	7	5
B) Responses in percent (%) to questions targeting NP2 (the colleague)				
	NP1	**NP2**	**NP3**	**'I don't know'**
	(journalist)	(colleague)	(mafia boss)	(?)
Condition	incorrect	**correct**	incorrect	incorrect
a. –subject, –animate	22	69	0	9
b. –subject, +animate	25	56	9	10
c. +subject, –animate	18	75	0	7
d. +subject, +animate	17	61	9	13
C) Responses in percent (%) to questions targeting NP3 (the scandal/mafia boss)				
	NP1	**NP2**	**NP3 (meeting/**	**'I don't know'**
	(journalist)	(colleague)	mafia boss)	(?)
Condition	incorrect	incorrect	**correct**	incorrect
a. –subject, –animate	13	9	69	9
b. –subject, +animate	17	9	61	13
c. +subject, –animate	12	4	75	9
d. +subject, +animate	21	5	63	11

Data accessibility statement

All materials, data and reproducible code are available from https://doi.org/10.17605/OSF.IO/A7CG2.

Ethics and consent

In line with the rules of the German Research Foundation (DFG), the German experiments were exempt from an ethics vote. As specified by the DFG, psycholinguistic experiments that use

non–invasive methods and that test healthy participants do not require a special ethics vote if the experiments do not pose a risk or physical/emotional burden to participants and as long as participants are debriefed about the study (see https://www.dfg.de/foerderung/faq/geistes_ sozialwissenschaften/index.html#anker13417818).

The English experiment was approved by the UMass Institutional Review Board as Protocol #1820 "Eye-tracking study on reading and memory."

Acknowledgements

Competing interests

One of the listed authors on this article, B.W. Dillon, is an Editor in Chief at *Glossa Psycholinguistics*. To avoid any potential conflict of interest, this article underwent double masked peer review at *Glossa: a journal of general linguistics*, and once accepted, was transferred to *Glossa Psycholinguistics* for publication.

Author contributions

Daniela Mertzen: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, writing – review & editing, visualization, project administration

Dario Paape: software, formal analysis, validation, writing – review & editing

Brian W. Dillon: conceptualization, methodology, resources, validation, writing – review & editing

Ralf Engbert: conceptualization, review, supervision, funding acquisition

Shravan Vasishth: conceptualization, methodology, software, formal analysis, review & editing, supervision, funding acquisition

References

Aho, A. V., & Ullman, J. D. (1972). *The theory of parsing, translation and compiling, Vol. I: Parsing.* Prentice Hall.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1060. DOI: https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought.* Lawrence Erlbaum Associates Publishers.

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n + 1 and n + 2 previews simultaneously. *Visual Cognition, 16*(6), 697–707. DOI: https://doi.org/10.1080/13506280802009704

Angele, B., Tran, R., & Rayner, K. (2013). Parafoveal-foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 526–538. DOI: https://doi.org/10.1037/a0029492

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. DOI: https://doi.org/10.1016/j.jml.2004.03.001

Arnett, N., & Wagers, M. (2017). Subject encodings and retrieval interference. *Journal of Memory and Language, 93*, 22–54. DOI: https://doi.org/10.1016/j.jml.2016.07.005

Avetisyan, S., Lago, S., & Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension? *Journal of Memory and Language, 112*, 104087. DOI: https://doi.org/10.1016/j.jml.2020.104087

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language, 56*(1), 65–85. DOI: https://doi.org/10.1016/j.jml.2006.08.004

Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (1st ed., pp. 225–242). Lawrence Erlbaum Associates.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*(1), 47–67. DOI: https://doi.org/10.1016/0010-0277(85)90023-X

Brasoveanu, A., & Dotlacil, J. (2020). *Computational cognitive modeling and linguistic theory* (1st ed.). Springer Cham. DOI: https://doi.org/10.1007/978-3-030-31846-8_4

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. DOI: https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software, 20*, 1–37. DOI: https://doi.org/10.18637/jss.v076.i01

Chomsky, N. (1986). *Barriers.* MIT Press.

Clark, H. H., & Begun, J. S. (1971). The semantics of sentence subjects. *Language and Speech*, *14*(1), 34–46. DOI: https://doi.org/10.1177/002383097101400105

Clifton, C., Juhasz, B., Ashby, J., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317–334. DOI: https://doi.org/10.1016/B978-008044980-7/50017-3

Crow, E. L., & Shimizu, K. (2018). *Lognormal distributions: Theory and applications*. Routledge. DOI: https://doi.org/10.1201/9780203748664

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. DOI: https://doi.org/10.1177/0956797613504966

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117–139. DOI: https://doi.org/10.1016/j.jml.2014.05.006

Cunnings, I., & Sturt, P. (2018). Retrieval interference and sentence interpretation. *Journal of Memory and Language*, *102*, 16–27. DOI: https://doi.org/10.1016/j.jml.2018.05.001

Dillon, B. W., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*, 85–103. DOI: https://doi.org/10.1016/j.jml.2013.04.003

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements*. Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199539789.013.0046

Engelmann, F., Jäger, L. A., & Vasishth, S. (2020). The effect of prominence and cue association in retrieval processes: A computational account. *Cognitive Science*, *43*, e12800. DOI: https://doi.org/10.1111/cogs.12800

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368. DOI: https://doi.org/10.1016/0749-596X(86)90006-9

Franck, J., & Wagers, M. (2020). Hierarchical structure and memory mechanisms in agreement attraction. *PLOS ONE*, *15*(5), 1–33. DOI: https://doi.org/10.1371/journal.pone.0232163

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies* (Doctoral dissertation). University of Massachusetts. Amherst, MA.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–586). Lawrence Erlbaum Associates.

Frazier, L., & Clifton, C. (1996). *Construal*. MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210. DOI: https://doi.org/10.1016/0010-0285(82)90008-1

Freedman, L. S., Lowe, D., & Macaskill, P. (1984). Stopping rules for clinical trials incorporating clinical opinion. *Biometrics*, *40*(3), 575–586. https://pubmed.ncbi.nlm.nih.gov/6518241/. DOI: https://doi.org/10.2307/2530902

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

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Third Edition). Chapman; Hall/CRC. DOI: https://doi.org/10.1201/b16018

Glaser, Y. G., Martin, R. C., Van Dyke, J. A., Hamilton, A. C., & Tan, Y. (2013). Neural basis of semantic and syntactic interference in sentence comprehension. *Brain and Language*, *126*(3), 314–326. DOI: https://doi.org/10.1016/j.bandl.2013.06.006

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114. DOI: https://doi.org/10.1016/j.jml.2004.02.003

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1304–1321. DOI: https://doi.org/10.1037/0278-7393.32.6.1304

Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 201–221. DOI: https://doi.org/10.1037/h0078814

Hyönä, J. (2012). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819–838). Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199539789.013.0045

Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology*, *16*(1–2), 104–127. DOI: https://doi.org/10.1080/09541440340000132

Inhoff, A. W. (1982). Parafoveal word perception: A further case against semantic preprocessing. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 137–145. DOI: https://doi.org/10.1037/0096-1523.8.1.137

Inhoff, A. W., & Rayner, K. (1980). Parafoveal word perception: A case against semantic preprocessing. *Perception & Psychophysics*, *27*(5), 457–464. DOI: https://doi.org/10.3758/BF03204463

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*, 431–439. DOI: https://doi.org/10.3758/BF03208203

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

Jäger, L. A., Engelmann, F., & Vasishth, S. (2015a). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*(617). DOI: https://doi.org/10.3389/fpsyg.2015.00617

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

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*. DOI: https://doi.org/10.1016/j.jml.2019.104063

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149. DOI: https://doi.org/10.1037/0033-295X.99.1.122

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. *Language and Cognitive Processes, 14*(5–6), 631–662. DOI: https://doi.org/10.1080/016909699386211

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision research, 45*(2), 153–168. DOI: https://doi.org/10.1016/j.visres.2004.07.037

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology Section A, 55*(4), 1307–1337. DOI: https://doi.org/10.1080/02724980244000071

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*, 627–645. DOI: https://doi.org/10.1023/A:1026528912821

Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press. DOI: https://doi.org/10.1016/B978-0-12-405888-0.00008-8

Kruschke, J., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178–206. DOI: https://doi.org/10.3758/s13423-016-1221-4

Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing* (Doctoral dissertation). University of Maryland. College Park, MD.

Lago, S., Acuña–Fariña, C., & Meseguer, E. (2021). The reading signatures of agreement attraction. *Open Mind, 5*, 132–153. DOI: https://doi.org/10.1162/opmi_a_00047

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language, 82*, 133–149. DOI: https://doi.org/10.1016/j.jml.2015.02.002

Lange, E. B., & Oberauer, K. (2005). Overwriting of phonemic features in serial recall. *Memory, 13*(3–4), 333–339. DOI: https://doi.org/10.1080/09658210344000378

Laurinavichyute, A., & von der Malsburg, T. (2022). Semantic attraction in sentence comprehension. *Cognitive Science, 46*(2), e13086. DOI: https://doi.org/10.1111/cogs.13086

Levy, R. P. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177. DOI: https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language, 68*(2), 199–222. DOI: https://doi.org/10.1016/j.jml.2012.02.005

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989–2001. DOI: https://doi.org/10.1016/j.jmva.2009.04.008

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*(1), 93–115. DOI: https://doi.org/10.1007/BF01708421

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

Logacev, P., & Vasishth, S. (2013). *Em2: A package for computing reading time measures for psycholinguistics* [R package version 0.9]. R package version 0.9. https://cran.rproject.org/src/contrib/Archive/em2/

López-Peréz, P., Dampuré, J., Hernández-Cabrera, J., & Barber, H. (2016). Semantic parafoveal-on-foveal effects and preview benefits in reading: Evidence from fixation related potentials. *Brain and Language, 162*, 29–34. DOI: https://doi.org/10.1016/j.bandl.2016.07.009

Lowder, M., & Gordon, P. C. (2014). Effects of animacy and noun-phrase relatedness on the processing of complex sentences. *Memory & Cognition, 42*, 794–805. DOI: https://doi.org/10.3758/s13421-013-0393-7

MacDonald, M. C., Perlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703. DOI: https://doi.org/10.1037/0033-295X.101.4.676

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111–123. DOI: https://doi.org/10.1023/A:1005184709695

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*, 283–312. DOI: https://doi.org/10.1006/jmla.1997.2543

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras & M. Just (Eds.), *New methods in reading comprehension research.* Erlbaum.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251–269. DOI: https://doi.org/10.3758/BF03213879

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin and Review, 7*, 403–423. DOI: https://doi.org/10.3758/BF03214356

Nicenboim, B., Schad, D. J., & Vasishth, S. (2023). *Introduction to Bayesian data analysis for cognitive science* [Under contract with Chapman and Hall/CRC Statistics in the Social and Behavioral Sciences Series]. https://vasishth.github.io/bayescogsci/

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. DOI: https://doi.org/10.1111/cogs.12589

Nicol, J., & Antón-Méndez, I. (2009). The effect of case marking on subject-verb agreement errors in English. In W. Lewis, S. Karimi, H. Harley, & S. Farrer (Eds.), *Time and again: Theoretical perspectives on formal linguistics in honor of D. Terence Langendoen* (1st ed., pp. 135–150). John Benjamins Publishing. DOI: https://doi.org/10.1075/la.135.10nic

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*(1), 5–19. DOI: https://doi.org/10.1007/BF01069043

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience, 14*(9), 1105–1107. DOI: https://doi.org/10.1038/nn.2886

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*(2), 601–626. DOI: https://doi.org/10.1016/j.jml.2006.08.009

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyperactive gap filling. *Frontiers in Psychology, 6*, 384. DOI: https://doi.org/10.3389/fpsyg.2015.00384

Paape, D., Hemforth, B., & Vasishth, S. (2018). Processing of ellipsis with garden-path antecedents in French and German: Evidence from eye tracking. *PLOS ONE, 13*(6), 1–46. DOI: https://doi.org/10.1371/journal.pone.0198620

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94*, 272–290. DOI: https://doi.org/10.1016/j.jml.2017.01.002

Parker, D., Shvartsman, M., & Dyke, J. A. V. (2018). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In L. Escobar, V. Torres, & T. Parodi (Eds.), *Language processing and disorders* (p. 121). Cambridge Scholars Publishing.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 940–961. DOI: https://doi.org/10.1037/0278-7393.24.4.940

Rabe, M. M., Chandra, J., Krügel, A., Seelig, S. A., Vasishth, S., & Engbert, R. (2021). A Bayesian approach to dynamical modeling of eye-movement control in reading of normal, mirrored, and scrambled texts. *Psychological Review, 28*, 803–823. DOI: https://doi.org/10.1037/rev0000268

Rabe, M. M., Paape, D., Mertzen, D., Vasishth, S., & Engbert, R. (2023). *SEAM: An integrated activation-coupled model of sentence processing and eye movements in reading.* https://osf.io/r39cx/

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*(1), 65–81. DOI: https://doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422. DOI: https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358–374. DOI: https://doi.org/10.1016/S0022-5371(83)90236-0

Rayner, K., White, S. J., Kambe, G., Miller, B., & Liversedge, S. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H.

Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 213–234). Elsevier. https://hdl.handle.net/2381/3902. DOI: https://doi.org/10.1016/B978-044451020-4/50013-X

Rich, S., & Wagers, M. (2020). Semantic similarity and temporal contiguity in subject-verb dependency processing, In *Proceedings of 33rd Annual CUNY Conference on Human Sentence Processing*.

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word $n+2$ in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 1026–1042. DOI: https://doi.org/10.1037/a0027735

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word $n+1$ during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 653–668. DOI: https://doi.org/10.1037/a0034997

Risse, S., & Seelig, S. (2019). Stable preview difficulty effects in reading with an improved variant of the boundary paradigm. *Quarterly Journal of Experimental Psychology*, *72*(7), 1632–1645. DOI: https://doi.org/10.1177/1747021818819990

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

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358–367. DOI: https://doi.org/10.1037/0033-295X.107.2.358

Royall, R. (1997). *Statistical evidence: A likelihood paradigm.* Chapman; Hall, CRC Press.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*, 104038. DOI: https://doi.org/10.1016/j.jml.2019.104038

Schotter, E., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*, 5–35. DOI: https://doi.org/10.3758/s13414-011-0219-2

Slioussar, N. (2018). Forms and features: The role of syncretism in number agreement attraction. *Journal of Memory and Language, 101*, 51–63. DOI: https://doi.org/10.1016/j.jml.2018.03.006

Smith, G., Franck, J., & Tabor, W. (2021). Encoding interference effects support self-organized sentence processing. *Cognitive Psychology*, *124*, 101356. DOI: https://doi.org/10.1016/j.cogpsych.2020.101356

Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 157*(3), 357–416. DOI: https://doi.org/10.2307/2983527

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1162–1169. DOI: https://doi.org/10.1037/0278-7393.33.6.1162

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

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50*(4), 355–370. DOI: https://doi.org/10.1016/j.jml.2004.01.001

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 431–450. DOI: https://doi.org/10.1037/0278-7393.30.2.431

Traxler, M. J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology Section A, 55*(1), 75–96. DOI: https://doi.org/10.1080/02724980143000172

Traxler, M. J. (2005). Plausibility and verb subcategorization in temporarily ambiguous sentences: Evidence from self-paced reading. *Journal of Psycholinguistic Research, 34*, 1–30. DOI: https://doi.org/10.1007/s10936-005-3629-2

Traxler, M. J., & Frazier, L. (2008). The role of pragmatic principles in resolving attachment ambiguities: Evidence from eye movements. *Memory & Cognition, 36*, 314–328. DOI: https://doi.org/10.3758/MC.36.2.314

Trueswell, J. C., Tanenhaus, M., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318. DOI: https://doi.org/10.1006/jmla.1994.1014

Trueswell, J. C., Tanenhaus, M., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528–553. DOI: https://doi.org/10.1037/0278-7393.19.3.528

Turk, U., & Logacev, P. (2022). *Agreement attraction in Turkish: The case of genitive attractors*. DOI: https://doi.org/10.31234/osf.io/5rmvu

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*(2), 407–430. DOI: https://doi.org/10.1037/0278-7393.33.2.407

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalysed ambiguities. *Journal of Memory and Language, 49,* 285–316. DOI: https://doi.org/10.1016/S0749-596X(03)00081-0

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*(3), 247–263. DOI: https://doi.org/10.1016/j.jml.2011.05.002

Vasishth, S. (2020). Using approximate Bayesian computation for estimating parameters in the cue-based retrieval model of sentence processing. *MethodsX, 7*, 100850. DOI: https://doi.org/10.1016/j.mex.2020.100850

Vasishth, S. (2023). Some right ways to analyze (psycho)linguistic data. *Annual Review of Linguistics, 9*(1), 273–291. DOI: https://doi.org/10.1146/annurev-linguistics-031220-010345

Vasishth, S., & Engelmann, F. (2021). *Sentence comprehension as a cognitive process: A computational approach*. Cambridge University Press. DOI: https://doi.org/10.1017/9781316459560

Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics, 59*, 1311–1342. DOI: https://doi.org/10.1515/ling-2019-0051

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language, 82*(4), 767–794. DOI: https://doi.org/10.1353/lan.2006.0236

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. DOI: https://doi.org/10.1016/j.jml.2018.07.004

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences, 23*, 968–982. DOI: https://doi.org/10.1016/j.tics.2019.09.003

Vasishth, S., Yadav, H., Schad, D. J., & Nicenboim, B. (2023). Sample size determination for Bayesian hierarchical models commonly used in psycholinguistics. *Computational Brain & Behavior, 6*, 102–126. DOI: https://doi.org/10.1007/s42113-021-00125-y

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology, 9*, 2. DOI: https://doi.org/10.3389/fpsyg.2018.00002

Vitu, F., Brysbaert, M., & Lancelin, D. (2004). A test of parafoveal-on-foveal effects with pairs of orthographically related words. *European Journal of Cognitive Psychology, 16*, 154–177. DOI: https://doi.org/10.1080/09541440340000178

Wagers, M. (2008). *The structure of memory meets memory for structure in linguistic cognition* (Doctoral dissertation). University of Maryland, College Park, MD, USA.

Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*, 206–237. DOI: https://doi.org/10.1016/j.jml.2009.04.002

Wagers, M., & McElree, B. (2009). Focal attention and the timing of memory retrieval in language comprehension, In *Proceedings of the Architectures and Mechanisms for Language Processing Conference*. Barcelona, Spain.

Wagers, M., & McElree, B. (2022). Memory for linguistic features and the focus of attention: Evidence from the dynamics of agreement inside DP. *Language, Cognition and Neuroscience*, 1–16. DOI: https://doi.org/10.1080/23273798.2022.2057559

Yadav, H., Paape, D., Smith, G., Dillon, B., & Vasishth, S. (2022). Individual differences in cue-weighting in sentence comprehension: An evaluation using Approximate Bayesian Computation. *Open Mind, 6*, 1–24. DOI: https://doi.org/10.1162/opmi_a_00052

Yadav, H., Smith, G., Reich, S., & Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *Journal of Memory and Language, 129*, 104400. DOI: https://doi.org/10.1016/j.jml.2022.104400