ORIGINAL PAPER

Similarity-Based Interference in Sentence Comprehension in Aphasia: a Computational Evaluation of Two Models of Cue-Based Retrieval

Paula Lissón[1] · Dario Paape[1] · Dorothea Pregla[1] · Frank Burchert[1] · Nicole Stadie[1] · Shravan Vasishth[1]

Abstract

Sentence comprehension requires the listener to link incoming words with short-term memory representations in order to build linguistic dependencies. The cue-based retrieval theory of sentence processing predicts that the retrieval of these memory representations is affected by similarity-based interference. We present the first large-scale computational evaluation of interference effects in two models of sentence processing — the activation-based model and a modification of the direct-access model — in individuals with aphasia (IWA) and control participants in German. The parameters of the models are linked to prominent theories of processing deficits in aphasia, and the models are tested against two linguistic constructions in German: pronoun resolution and relative clauses. The data come from a visual-world eye-tracking experiment combined with a sentence-picture matching task. The results show that both control participants and IWA are susceptible to retrieval interference, and that a combination of theoretical explanations (intermittent deficiencies, slow syntax, and resource reduction) can explain IWA's deficits in sentence processing. Model comparisons reveal that both models have a similar predictive performance in pronoun resolution, but the activation-based model outperforms the direct-access model in relative clauses.

Introduction

When hearing a sentence, the listener has to link incoming words together and build up syntactic and semantic relations in real time. For instance, verbs have to be linked with their dependent arguments, which is commonly assumed to require retrieval from memory (Lewis, 1999; Van Dyke & Lewis, 2003; McElree, 2006). In a sentence like (1), focusing on the relative clause (RC) *The boy [who greeted the girl]*, when the verb *greeted* inside the relative clause is reached, the comprehender must retrieve the subject *boy*

from memory in order to understand who greeted whom.[1] (e.g., Chomsky, 1977). However, for the sake of simplicity, and assuming that both the noun phrase and the relative pronoun inherit the relevant retrieval features from the head noun *boy*, we will refer directly to the retrieval of *boy*. In linguistics, resolving who did what to whom is known as thematic role assignment. The doer of the action expressed by the relative clause verb *greeted* is the agent (*boy*) and the recipient of the action is the theme (*girl*).

(1)   The boy who greeted the girl plays with the dog.

Cue-based retrieval theory (McElree, 2000; McElree, Foraker, & Dyer, 2003; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Vasishth, Nicenboim, Engelmann, & Burchert, 2019) posits that items are retrieved from

Paula Lissón and Dario Paape contributed equally to this work.

✉ Dario Paape
paape@uni-potsdam.de

1   Department of Linguistics, University of Potsdam, Potsdam, Germany

---

[1]In formal analyses, it is usually assumed that the syntactic dependency within the relative clause is between the verb and the relative pronoun *who*, which enters into a semantic dependency with the noun phrase *the boy* in the matrix clause, to which it refers back.

memory based on their syntactic and semantic features,[2] we refer to all the different types of cue-based retrieval accounts as "the" cue-based retrieval theory, even though there are important differences in the underlying latent process assumed and the quantitative predictions of the different variants (e.g., Nicenboim and Vasishth, 2018; Lissón et al., 2021a). According to cue-based retrieval as implemented computationally in Lewis and Vasishth (2005), items are stored in memory as a bundle of feature-value pairs. In a subject relative clause such as in (1), the embedded verb *greeted* triggers the retrieval of an item in memory whose features match the retrieval cues [+animate] and [+RC subject], which identify the agent of the relative clause. When the relative clause verb *greeted* is read, the only item in memory that matches these retrieval cues is *boy*. Consider now sentence (2), which is an object relative clause.

(2)   The girl who the boy greeted plays with the dog.

In this case, when the comprehender reaches the verb *greeted*, there is one item in memory, *boy*, that matches all the retrieval cues set by the verb, but there is also another item, *girl*, that matches some of the retrieval cues ([+animate, −RC subject]). Following Jäger, Engelmann, & Vasishth (2017), we will refer to the fully matching item (*boy*) as the retrieval target, and to items with partial feature match (*girl*) as distractors.

A core assumption of the cue-based retrieval theory is that memory retrieval is subject to interference: When a retrieval is triggered, processing difficulty is predicted if multiple items in memory match the same retrieval cues. Therefore, the verb *greeted* should be more difficult to process in (2) than in (1). This effect is known as similarity-based interference and is indexed by a slowdown at the retrieval site and/or by occasional misretrievals of a distractor item, which results in misinterpretation, such as the girl being interpreted as the agent in (2). Similarity-based interference has been attested in multiple linguistic constructions across different languages (e.g., Van Dyke and Lewis, 2003; Van Dyke & McElree, 2006, 2011; Van Dyke, 2007; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Van Dyke & McElree, 2011; Martin, Nieuwland, & Carreiras, 2012; Jäger et al., 2017; Engelmann, Jäger, & Vasishth, 2019; Jäger, Mertzen, Van Dyke, & Vasishth, 2020; Vasishth & Engelmann, 2022).

Within the cue-based retrieval framework, two different models of retrieval processes have been proposed: the activation-based model (Lewis & Vasishth, 2005) and the direct-access model (McElree, 2000). The two models share the assumption that retrieval cues mediate access to items in memory. However, they make different assumptions regarding the underlying latent processes that unfold in memory retrieval.

The activation-based model was originally implemented in the cognitive architecture ACT-R (Anderson et al., 2004). Because the full ACT-R based model is implemented in the programming language Lisp, it is not easily accessible to the wider community in psycholinguistics. Partly in response to this problem, Engelmann et al. (2019) developed an implementation in R (R Core Team, 2020) that represents a simplified version of the full model.[3] The match between simulated data from the Lewis and Vasishth (2005) model and human experimental data has been studied in subject-verb and reflexive-antecedent dependencies, negative polarity items, and other linguistic constructions (Vasishth & Lewis, 2006; Wagers, Lau, & Phillips, 2009; Jäger et al., 2020; Dillon, Mishler, Sloggett, & Phillips, 2013; Jäger, Engelmann, & Vasishth, 2015; Nicenboim & Vasishth, 2016; Patil, Vasishth, & Lewis, 2016; Parker & Phillips, 2017; Vasishth et al., 2019).

In the activation-based model implemented in Lewis and Vasishth (2005), each memory item has a fluctuating activation value that determines both the probability and the latency of retrieval. When a retrieval is triggered, the retrieval cues spread activation to all matching items available in memory. Items with more matches accrue more activation, making them more likely to be retrieved, and decreasing retrieval latency. However, if the cued feature is present on multiple items in memory, the cue's activation is shared across all items, so that each item receives comparatively less activation. The reduced activation of the target item in memory and the increased activation of competing items are the source of similarity-based interference.

The direct-access model, developed by McElree and colleagues (McElree, 2000, 2006; McElree et al., 2003; Martin & McElree, 2011), also predicts similarity-based interference. It assumes that the availability of items in memory — that is, the probability of successful retrieval — decreases as a function of interference, but that retrieval times remain unaffected. However, low availability can lead to misretrievals and/or parsing failure: if a retrieval fails completely, that is, if no appropriate chunk can be retrieved to perform a syntactic attachment, words will be left "stranded," that is, fail to be integrated into the syntactic structure (Lewis & Vasishth, 2005; Bartek, Lewis, Vasishth, & Smith, 2011), and the parse will crash. If retrieval fails or if an incorrect chunk is retrieved, in a certain proportion of trials, a backtracking process is initiated that requires some extra processing time (Martin

---

[2]Following Parker et al. (2017).

[3]A Shiny app provides easy access to this simplified model's quantitative predictions under parametric variation: https://vasishth.github.io/RetrievalModels/.

& McElree, 2008). Backtracking, also known as reanalysis, is implicitly assumed to lead to the retrieval of the target (McElree, 1993).

Nicenboim and Vasishth (2018) compared the activation-based model and the direct-access model using self-paced reading data from unimpaired adult readers (Nicenboim, Vasishth, Engelmann, & Suckow, 2018). Nicenboim and Vasishth (2018) showed that the predictive performance of the default activation-based model implementation was worse compared to that of the direct-access model. However, the models had similar quantitative performance when the activation-based model was implemented with different variances for target and distractors. Lissón et al. (2021a) tested the two competing models against self-paced listening data from individuals with aphasia and control participants (Caplan, Michaud, & Hufford, 2015). The authors modeled the comprehension of English subject and object relative clauses as in (1) and (2), in self-paced listening and a sentence-picture matching task. Model comparisons showed similar quantitative performance, but major qualitative differences. Lissón et al. concluded that in order to disentangle the differences between the models, more studies using different linguistic constructions and different experimental paradigms were needed. This is the empirical gap that the current study aims to fill.

In the context of aphasic sentence processing, one assumption of the direct-access model is potentially overly constraining, namely the assumption that backtracking, if initiated, always leads to correct retrieval of the target. Because of this assumption, due to the added backtracking time in some of the correct trials, the direct-access model assumes that correct retrievals are, on average, slower than misretrievals. However, it is known from different cognitive tasks that "slow errors" can occur in addition to "fast errors" (e.g., Van Maanen, Katsimpokis, & Van Campen, 2019). The direct-access model's assumption that correct retrievals are on average slower than incorrect retrievals leads to incorrect predictions when modeling data from individuals with aphasia with the direct-access model (Lissón et al., 2021a): Individuals with aphasia (IWA) often have slower latencies in incorrect trials relative to correct trials (see Hanne, Burchert, de Bleser, & Vasishth, 2015; Adelt, Stadie, Lassotta, Adani, & Burchert, 2017, Pregla, Lissón, Vasishth, Burchert, & Stadie, 2021).

Based on the high prevalence of "slow errors" in IWA, Lissón et al. (2021b) implemented a modified version of the direct-access model. In this model, there is not only a distinction between trials in which backtracking is initiated and trials in which it is not, but also a distinction between trials in which backtracking is successful and trials in which it fails. Trials with failed backtracking are slower than trials without backtracking, but the parser is stuck with the result of the original misretrieval. Lissón et al. (2021b) tested

the modified direct-access model and the original direct-access model against self-paced listening data from IWA and control participants in German (Pregla et al., 2021). The models were compared using Bayes factors, and the result was inconclusive. In the present study, we compare the modified direct-access model against the activation-based model using a larger dataset.

We model interference effects in IWA and control participants using a subset of the data from Pregla, Vasishth, Lissón, Stadie, & Burchert (2022). We focus on two linguistic constructions in German: pronoun resolution and relative clauses. These constructions are well-suited for our modeling goals because IWA have difficulty processing them (Burchert, de Bleser, & Sonntag, 2003; Choy & Thompson, 2010; Caplan et al., 2015; Adelt et al., 2017; Pregla et al., 2021). Furthermore, given that cue-based retrieval is intended as a general model of sentence processing, it is necessary to investigate different constructions and test whether the proposed implementations are able to account for the entire range of data.

We also aim to establish links between model parameters and prominent theories of processing deficits in aphasia. Linking these verbally stated theories to model parameters is crucially important, because it enables us to derive constrained, testable predictions for each theory and to evaluate them against the data. Finally, our study is, to our knowledge, the first to compare two different models of cue-based retrieval using online eye-tracking data from the visual-world paradigm in combination with an offline sentence-picture matching task.[4]

We seek to answer the following questions:

1. Which model of cue-based retrieval offers a better account of interference effects in IWA and control participants across pronoun resolution and relative clauses?
2. How do the parameters of each model map onto theories of processing deficits in IWA?

The paper is structured as follows: We begin by summarizing the theories of processing deficits in aphasia that we will evaluate, as well as their proposed connection to the parameters of the activation-based and modified direct-access models. We then introduce the two linguistic constructions of interest, namely pronoun resolution and relative clauses. Next, on the basis of these constructions, we discuss the implementation of the competing models

---

[4]Patil, Hanne, Burchert, de Bleser, & Vasishth (2016) modeled visual-world eye-tracking data from 7 IWA and 8 controls with different versions of the Lewis and Vasishth (2005) model, but the (modified) direct-access model has never been tested against visual-world eye-tracking data, and never with such a relatively large-scale dataset from IWA and controls.

in greater detail, and link the assumed parameters to influential theories from the aphasia literature. Finally, we fit the models to the data and evaluate whether the theoretical predictions are borne out. We also assess each model's predictive fit by repeatedly fitting the models to subsets of the data, simulating new data and comparing the simulated data to a different subset of the original data. To anticipate our results, across IWA and control participants, and across both linguistic constructions, neither of the two implementations of cue-based retrieval performs decisively better than the other in terms of predictive fit. However, the parameter estimates from each of the two models are informative with regard to the underlying deficits in IWA: There is support in the data for slow syntax, intermittent deficiencies, and resource reduction (see discussion below), but less support for delayed lexical access.

Theories of Processing Deficits in Aphasia

Individuals with aphasia have difficulties processing non-canonical sentences (e.g., Caramazza & Zurif, 1976; McAllister et al., 2009; Schumacher et al., 2015) such as object relative clauses and passives, especially when the thematic roles are semantically reversible. That is, IWA experience difficulties identifying the agent and theme of the verb (who did what to whom) based on morpho-syntactic cues alone. Similarly, IWA also experience difficulties comprehending binding relations, that is, pronouns and reflexives (e.g., *Justin told [Thomas$_i$ to shave himself$_i$]*; Edwards & Varlokosta, 2007; Choy & Thompson, 2010).

Caplan et al. (2015) discuss the different theories that aim to explain why these constructions are challenging for IWA[5]. For instance, Burkhardt, Piñango, & Wong (2003) and Burkhardt, Avrutin, Piango, & Ruigendijk (2008) argue that IWA compute syntactic dependencies at a slower-than-normal pace, which can lead to comprehension failure. According to this theory, known as *slow syntax*, the processing deficit in IWA is specific to syntactic structure building. By contrast, Ferrill, Love, Walenski, & Shapiro (2012), and Love, Swinney, Walenski, & Zurif (2008) posit that *delayed lexical access* causes the slowdown in the formation of syntactic dependencies. Love, Swinney, Walenski, & Zurif (2008) claim that when the sentence requires the reactivation of a lexical item in order to complete a dependency, the lexical reactivation is too slow, and an extragrammatical heuristic may be used instead, which can lead to comprehension errors. These two theories — slow syntax and delayed

lexical access — have been tested using the cross-modal lexical decision paradigm. However, studies using the visual-world paradigm in aphasia do not support a delay in lexical access or in syntactic structure building as the main source of comprehension deficits in IWA (e.g., Dickey, Choy, & Thompson, 2007). The data modeled in the present work is well-suited to test the predictions of these two theories, because we consider visual-world data in combination with reaction times and accuracies from a picture-selection task, which require similar motor responses as in the cross-modal lexical decision task.

Another theory, proposed by Caplan and colleagues, argues that IWA may have an impairment in the resources needed for parsing, such as memory capacity (Caplan, Waters, DeDe, Michaud, & Reddy, 2007; Caplan, 2012). Complex sentences create greater processing demands, and therefore IWA have more difficulties in complex sentences. This account is known as *resource reduction*. In addition, Caplan et al. (2013) claim that IWA may also exhibit *intermittent deficiencies* in the parsing system that block access to parsing operations such as relating the surface and base positions of words in the structure. The intermittent nature of these breakdowns would explain why IWA are able to understand complex sentences on some but not all trials.

All of these theoretical proposals can be incorporated into computational models of retrieval. For instance, Patil, Hanne, Burchert, de Bleser, & Vasishth (2016) modeled the comprehension of active vs. passive sentences in a visual-world eye-tracking experiment in German (Hanne, Sekerina, Vasishth, Burchert, & de Bleser, 2011), using different implementations of the Lewis and Vasishth (2005) model. The best-fitting model for IWA was one that assumed generally slowed processing (understood as a combination of delayed lexical access and slow syntax), as well as intermittent deficiencies. In another modeling study, Mätzig, Vasishth Engelmann, Caplan, and Burchert mapped parameters of the Lewis and Vasishth (2005) model to slowed processing, intermittent deficiencies, and resource reduction. The authors modeled accuracies in English subject and object relative clauses using the data from Caplan et al. (2015). They concluded that IWA's performance can be explained by a combination of these three deficits, and that slowed processing, intermittent deficiencies, and resource reduction may affect each individual with aphasia to a different degree. Similarly, in English relative clauses, using self-paced listening data from Caplan et al. (2015), Lissón et al. (2021a) also found that intermittent deficiencies, delayed lexical access, and slow syntax can explain the main processing deficits in IWA.

In the current work, we focus on the role of the proposed processing deficits in the context of the activation-based and the direct-access models of cue-based retrieval.

---

[5]Although the slow syntax and the delayed lexical access theories were originally proposed for Broca's aphasia, the studies by Caplan and colleagues (Caplan, Michaud, & Hufford, 2013, 2015) show that these deficits could also be playing an important role in impaired sentence comprehension in patients with other types of aphasia.

In our modeling, we follow Lissón et al. (2021a) and Lissón et al. (2021b) and implement intermittent deficiencies as increased stochastic noise in memory activations/availabilities. A higher noise value in IWA would mean more mistretrievals and presumably more parsing failures due to failed retrievals compared to unimpaired individuals. Delayed lexical access or slow syntax is assumed to delay the retrieval of the target item from memory, leading to a slowdown at the retrieval site in the activation-based model, and/or to misretrieval in both the activation-based and the direct-access model. In the DA model, which assumes backtracking as a key resource in parsing, resource reductions could disrupt this mechanism and lead to comprehension deficits.

We fit the models to data from a picture-selection task and a visual-world experiment that tested the comprehension of pronouns and relative clauses in German (Pregla et al., 2022). We now introduce each linguistic construction in turn.

Experiment 1: Pronoun Resolution

Consider sentence (3a). When the pronoun *er* ("he") is encountered, cue-based retrieval predicts that a search for its antecedent is triggered, using the cues [+animate, +masculine, +singular].[6] The experiment makes use of the fact that for some verbs, the implicit subject of a sentential complement is coreferential with the main clause subject (subject control) while for others it is coreferential with the main clause object (object control, e.g., Chomsky, 1981; Comrie, 1985). The verb *versprechen* ("promise") is lexically specified as a subject control verb (Müller, 2002). Even though we do not investigate control structures, we assume that the additional retrieval cue [+subj] is set at the pronoun, and that *Peter* is the retrieval target: Because *versprechen* ("promise") enforces subject control, replacing *er* ("he") with *sie* ("she") to refer to *Lisa* in (3a) results in unacceptability, and there is no alternative antecedent in the discourse to which the pronoun could refer.

(3) a. **Mismatch.**

$\text{Peter}^{+\text{subj}}_{+\text{masc}}$ verspricht nun $\text{Lisa}^{-\text{subj}}_{-\text{masc}}$, dass $\text{er}^{\{\text{subj}\}}_{\{\text{masc}\}}$ das kleine Lamm streichelt und krault.

$\text{Peter}^{+\text{subj}}_{+\text{masc}}$ promises now $\text{Lisa}^{-\text{subj}}_{-\text{masc}}$, that $\text{he}^{\{\text{subj}\}}_{\{\text{masc}\}}$ the small lamb pets and ruffles.

'Peter now promises Lisa that he will pet and ruffle the little lamb.'

b. **Match.**

$\text{Peter}^{+\text{subj}}_{+\text{masc}}$ verspricht nun $\text{Thomas}^{-\text{subj}}_{+\text{masc}}$, dass $\text{er}^{\{\text{subj}\}}_{\{\text{masc}\}}$ das kleine Lamm streichelt und krault.

$\text{Peter}^{+\text{subj}}_{+\text{masc}}$ promises now $\text{Thomas}^{-\text{subj}}_{+\text{masc}}$, that $\text{he}^{\{\text{subj}\}}_{\{\text{masc}\}}$ the small lamb pets and ruffles.

'Peter now promises Thomas that he will pet and ruffle the little lamb.'

Across the two sentences, the main clause object nouns, *Lisa* in (3a) and *Thomas* in (3b), partially match the retrieval cues from the pronoun. Both mismatch the [+subj] cue that the pronoun inherits from the verb, but *Thomas* matches the gender cue from the pronoun, which should lead to increased similarity-based interference. We will refer to sentences like (3a) as mismatch conditions, because the target noun (*Peter*) and the distractor noun (*Lisa*) do not share the same gender, and sentences like (3b), as match conditions.

In unimpaired populations, a processing advantage in gender mismatch conditions has been observed in English by Badecker and Straub (2002) and Runner and Head (2014), but not by Chow et al. (2014). Laurinavichyute, Jäger, Akinina, Roß, & Dragoy (2017) reported mixed results for German. In the aphasia literature, Choy and Thompson (2010) and Engel, Shapiro, & Love (2018) found that IWA had difficulties in pronoun resolution, but these studies did not target the gender mismatch configurations that Pregla et al. (2022) tested, and that we model in the present work.

In Experiment 1, we model interference as a function of the gender cue at the pronoun in the Pregla et al. (2022) data. Based on cue-based retrieval theory, we predict a processing advantage in gender mismatch conditions relative to the gender match conditions. We aim to (a) compare how the activation-based and the modified direct-access model fit these data, and (b) evaluate the theoretical accounts of processing deficits in aphasia by mapping them onto model parameters.

---

[6]The parser presumably also honors structural constraints during retrieval, such as adherence to Binding Principle B (Chomsky, 1981), as opposed to first retrieving and then eliminating structurally illicit antecedents (e.g., Chow, Lewis, & Phillips, 2014). In the current context, however, both possible antecedents are structurally available, so that we abstract away from structural retrieval cues.

Experiment 2: Relative Clauses

Relative clauses have been extensively studied in psycholinguistics. Subject relatives (SR) have been found to be easier to process than object relatives (OR) in multiple languages for both unimpaired controls (e.g., Grodner and Gibson, 2005; Fedorenko, Gibson, & Rohde, 2006; Gordon, Hendrick, Johnson, & Lee, 2006; Staub, 2010, Staub, Dillon, & Clifton Jr, 2017) and for IWA (e.g., Caramazza & Zurif, 1976; Burchert et al., 2003; Caplan et al., 2007; Dickey & Thompson, 2009; Caplan et al., 2013, 2015; Pregla et al., 2021). The subject-object asymmetry in IWA and controls has been computationally modeled in the cue-based retrieval framework (Mätzig et al., 2018; Vasishth et al., 2019; Lissón et al., 2021a) using self-paced listening data and offline measures in English. The present study is the first to model number interference in relative clauses in German, in IWA and unimpaired controls.

Our study mainly focuses on the effect of number interference within subject and object relative clauses rather than on the well-studied SR/OR asymmetry. This is because similarity-based interference alone cannot account for the asymmetry in German relative clauses, as the retrieval point for both SRs and ORs is the clause-final verb (see discussion below). However, it is still informative to check if the asymmetry can be captured by assuming changes in the parameters of the activation-based model and the modified direct-access model.

Consider the sentences in (4). When the verb *badet/baden* (bathes/bathe) is encountered at the end of the sentence, two retrievals are triggered, because the agent and the theme of the action expressed by the verb need to be identified. Our modeling focuses on the retrieval of the agent because that is the theoretically interesting event (this is explained below).

(4)   a. **SR, match**

Hier ist [der Esel]$_{+\text{singular}}$, der [den Tiger]$_{+\text{singular}}$ gerade badet$_{\{\text{singular}\}}$.

Here is the$_{+\text{singular}}$ donkey who the$_{+\text{singular}}$ tiger now bathes$_{\text{singular}}$.

'Here is the donkey who bathes the tiger.'

b. **OR, match**

Hier ist [der Esel]$_{+\text{singular}}$, den [der Tiger]$_{+\text{singular}}$ gerade badet$_{\{\text{singular}\}}$.

Here is the$_{+\text{singular}}$ donkey who the$_{+\text{singular}}$ tiger now bathes$_{\text{singular}}$.

'Here is the donkey who the tiger bathes.'

c. **SR, mismatch**

Hier ist [der Esel]$_{+\text{singular}}$, der [die Tiger]$_{+plural}$ gerade badet$_{\{\text{singular}\}}$.

Here is the$_{+\text{singular}}$ donkey who the$_{+plural}$ tigers now bathes$_{\text{singular}}$.

'Here is the donkey who bathes the tigers.'

d. **OR, mismatch**

Hier ist [der Esel]$_{+singular}$, den [die Tiger]$_{+\text{plural}}$ gerade baden$_{\{\text{plural}\}}$.

Here is the$_{+singular}$ donkey who the$_{+\text{plural}}$ tigers now bathe$_{\text{plural}}$.

'Here is the donkey who the tigers bathe.'

In (4a) and (4b), both noun phrases in the sentence are singular (*der Esel, der/den Tiger*). This is expected to cause similarity-based interference during the retrieval of the subject. By contrast, in (4c) and (4d), the second noun phrase is plural (*die Tiger*), which should result in easier identification of the subject, based on the retrieval cue from the verb ([singular] or [plural]). We will refer to sentences like (4c) and (4d) as mismatch conditions, because target and distractor do not share the same number, and sentences like (4a) and (4b), as match conditions. Both types of relative clauses should be easier to process in mismatch configurations compared to match conditions. Thus, (4c) and (4d) should be easier to process than (4a) and (4b), respectively.

In German, when the head noun is masculine (such as in our items), the morphological form of the relativizer (*der* for nominative, *den* for accusative) provides disambiguating information. Therefore, in our items, by the time the comprehender reaches the relativizer, they should be able to identify the agent or the theme of the relative clause due to the overt case marking. Retrieval should occur at the verb, and the number cue should facilitate processing in (4c) vs. (4a) and (4d) vs. (4b), because in (4c) and (4d) only the subject noun phrase matches the number cue at the verb.

Studies addressing the comprehension of subject vs. object relatives in German with case-unambiguous relativizers (such as our items) are scarce and have mainly addressed the SR/OR asymmetry (Friederici, Steinhauer, Mecklinger, & Meyer, 1998; Burchert et al., 2003; Adelt et al., 2017). By contrast, the main goals in our Experiment 2 are to compare the performance of the activation-based and the modified direct-access model by modeling number interference in both types of relative clauses, and to evaluate the different theories of processing deficits in aphasia based on the model estimates.

Methods

The data that we model come from the experiments carried out by Pregla et al. (2022). The participants, procedure, and materials described here summarize the methods in Pregla et al. (2022). By contrast, the dependent variables and contrast coding described here are specific to the present paper. Pregla et al. (2022) analyze the visual-world eye-tracking data, whereas we model the reaction times from the picture-selection task that followed the visual-world paradigm task.

Participants and Procedure Twenty-one IWA (9 females, mean age = 60.2 years, SD = 11.4; mean education = 15.2 years, SD = 3.2) and fifty control participants (32 females, mean age = 47.7 years, SD = 19.6; mean education = 18.1 years, SD = 4.0), all native speakers of German, took part in an auditory sentence-picture matching task combined with visual-world eye-tracking. Individuals with aphasia were in the chronic phase (at least 1 year after onset of the aphasia). More information about the patients is given in Appendix A. The procedure was as follows: At the beginning of the trial, a preview phase of 4000 ms was used to introduce two pictures to the participants. One of the pictures (target) corresponded to the correct meaning of the sentence, whereas the other picture (foil) depicted the opposite thematic interpretation. After the preview phase, an auditory recording of the sentence started playing. Sentences were presented at a normal speech rate, and participants were instructed to select the picture that matched the meaning of the sentence. The pictures were displayed until participants made a choice, or for a maximum time of 30 s. Once the participant pressed the choice button, the trial ended. During the trial, eye movements were recorded using a SensoMotoric Instruments eye-tracker (SMI RED250mobile; binocular tracking, Experiment Center version 3.7, sampling rate 250 Hz). The proportion of looks to the target picture against looks to the foil (or no picture) was calculated. The response time and the accuracy of the picture selection were also recorded. Each participant completed the experiments twice, in two sessions (test and retest), with a gap of approximately 2 months.[7] Participants also performed a battery of tests in order to assess auditory and visual comprehension, morphological discrimination, and lexical decision latency.

Materials In Experiment 1, two conditions (mach and mismatch, as in example 3), with 10 items each were included. Example pictures accompanying the pronoun sentences are shown in Fig. 1. The pronoun items always used subject-control verbs, so that the target of the retrieval was always the first noun phrase.[8] Control verbs were selected from the ZAS Database of Clause-Embedding Predicates (Stiebels et al., 2018).

In experiment 2, 20 items per relative clause type (subject/object relative) were included. The noun phrase of the matrix clause (henceforth NP1) was always masculine and singular. Out of the 20 items, 10 had a singular embedded noun phrase (henceforth NP2), and 10 had a plural embedded noun phrase. The items were constructed using 10 bisyllabic transitive action verbs, and the noun

---

[7]The test vs. retest main effect is not included in our modeling (see the section on dependent variables) because the model structure had to be simplified due to convergence issues. The test-retest reliability for these data has been investigated in Pregla et al. (2021).

[8]The pronoun items that we use here were extracted from a larger experiment that also contained object-control verbs, and fillers.

Fig. 1 Example pictures used in the picture-selection task in Experiment 1. For the sentences in example (3) the left picture is the target and the right picture is the foil
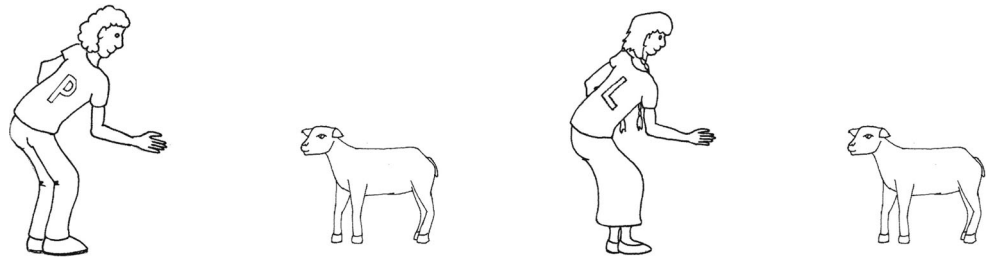
phrases were always bisyllabic animal names. Example pictures for the relative clause conditions are given in Fig. 2.

Dependent Variables and Contrast Coding To assess participants' lexical access speed, which is important for evaluating the delayed lexical access hypothesis of Love et al. (2008) and Ferrill et al. (2012), we computed their average reaction times in a lexical decision task, based on LEMO 2.0 (Stadie, Cholewa, & de Bleser, 2013). In this task, participants have to decide whether an auditorily presented item is a word or a non-word. Participants responded by pressing one of two buttons on a computer keyboard, and the accuracy and response times were recorded. We computed the average reaction times in correct trials, which yielded a single measure (lexical decision time, LDT) for each participant that we use as a continuous predictor in the models.

Response times in lexical decision tasks have been previously used as a measure of lexical access speed in

IWA. For instance, Caplan et al. (2015) correlated the lexical decision times for correct responses with the accuracy and reading times in a self-paced listening task. We centered and scaled the LDT predictor within groups. An LDT × group interaction would thus tell us whether an increase in LDT leads to a larger increase in RT for IWA compared to controls.

Another predictor in the models is the proportion of fixations on the target picture (centered and scaled within groups) at the critical sentence region, where retrieval is assumed to take place. We use the proportions of looks to the target at the critical region as a proxy for retrieval (see the next section for more details). The remaining predictors used in both models were sum-coded, with the following contrasts: group was coded with IWA as +1 and controls as −1; the high interference conditions (gender match in pronouns, number match in relative clauses) were coded as +1, and the low interference conditions (gender mismatch, number mismatch) as −1. In the relative clauses

Fig. 2 Example pictures used in the picture-selection task in Experiment 2. The pictures correspond to the correct interpretations of the example sentences in (4). Within both the match and the mismatch conditions, participants had to select the correct interpretation (SR interpretation vs. OR interpretation)

(a) SR, match.

(b) OR, match.

(c) SR, mismatch.

(d) OR, mismatch.
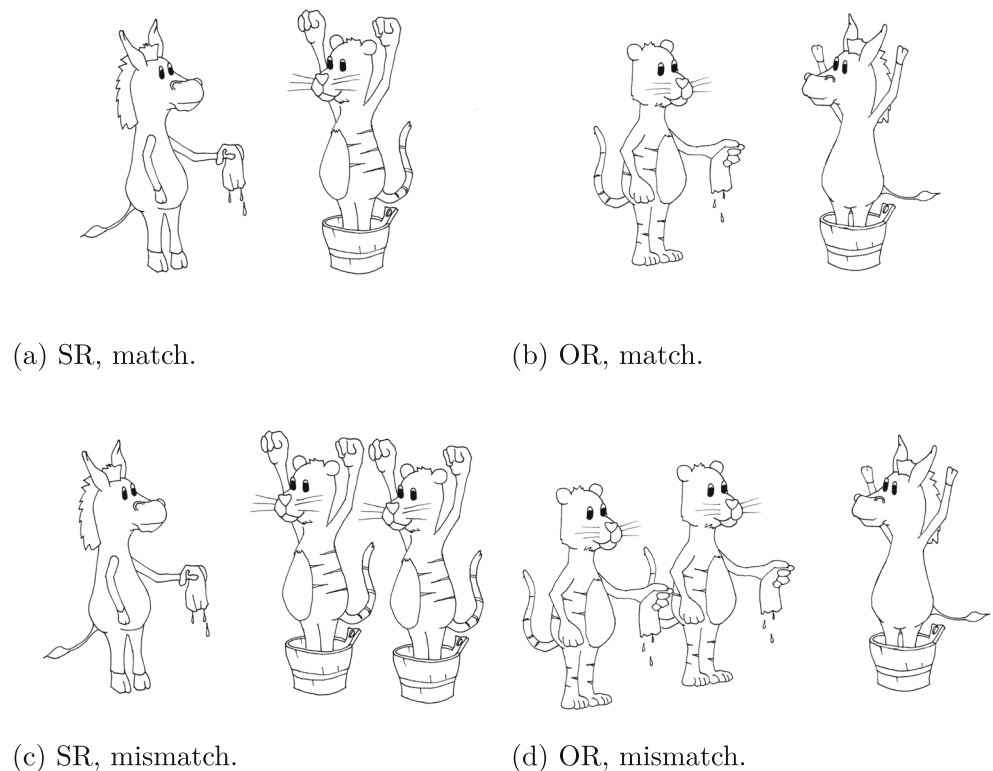
sub-experiment, the relative clauses were coded as OR +1, and SR −1.

We base our statistical inferences on the posterior distribution of the parameters, which we summarize with the mean and 95% credible interval (CrI). This is the convention used to report summaries of parameter values for which there is support in the data. When interpreting the estimates, the width of the CrI should be taken into account, as it shows the range of plausible parameter values that lie with 95% probability given our model and data.

For both the activation-based model and the modified direct-access model, the data from the pronoun experiment and the relative clause experiment were fitted separately.[9] The four models were implemented in Stan (Carpenter et al., 2017) and fitted in R (R Core Team, 2020) via the rstan package (Stan Development Team, 2020). The packages brms (Bürkner, 2017) and bayesplot (Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019) were used for examining and plotting the posterior distributions of the parameters. For each model, three chains each, with at least 6000 iterations each were run. Each chain included at least 3000 warm-up iterations. Convergence was assessed by checking that $\widehat{R}$ was below 1.01 and by visually inspecting the convergence of the chains (Gelman et al., 2013). We also verified that the models could recover simulated parameter values. For both models, mildly informative priors were used (Sorensen, Hohenstein, & Vasishth, 2016; Nicenboim et al., 2021; Schad, Betancourt, & Vasishth, 2021). Details about the implementation and the priors are available in Appendix C.

Modeling Assumptions

Neither the activation-based model nor the direct-acccess model have a linking function that maps proportions of looks to a picture to retrieval times and/or retrieval probabilities of memory chunks. Therefore, we need to specify linking assumptions between fixations on the target picture, the assumed retrieval processes, and the reaction times and accuracies in the picture-selection task.

For the two sentence types, we assume two retrieval events. The first retrieval takes place in the middle of the sentence, at the *critical region*. In pronoun resolution, the critical region is the pronoun, and in relative clauses, it is the relativizer. Our linking assumption is that proportions of looks to the target at the critical region can be used as a proxy for retrieval. This assumption is based on the fact that the critical region provides the necessary cues to retrieve the

target. Therefore, we predict that more looks to the target picture at the retrieval site correspond to a higher probability that the target has been retrieved at this point.

The second retrieval event happens at the verb region, that is, at the end of the sentence. In both experiments, the retrieval target must be re-accessed at this point, as a dependency with the verb needs to be established. We assume that the second retrieval is linked to the first, so that more looks to the target at the critical region (the pronoun/relativizer) go along with higher activation/availability, resulting in faster and/or more accurate retrieval at the verb region. As the picture-selection task takes place immediately after hearing the verb region, we assume that accuracies and RT in this task should show the interference effects predicted by the cue-based retrieval theory.

We do not model retrieval failures, that is, trials in which neither the target nor the distractor can be retrieved. This simplification is necessary because participants in the Pregla et al. (2021) study had to select one of the two pictures and were not given the option to respond "I don't know," which could be interpreted to index retrieval failure (Nicenboim & Vasishth, 2018). Furthermore, even in the presence of such an option, participants may resort to guessing rather than responding "I don't know," although it has been suggested that IWA only use such compensatory processes rarely (Hanne et al., 2011; Burchert, Hanne, & Vasishth, 2013; Arantzeta, Webster, Laka, Martínez-Zabaleta, & Howard, 2018). Parameters related to guessing can, in principle, be implemented in probabilistic cognitive models (e.g., Oberauer, 2006; Logačev & Vasishth, 2016), including the direct-access model, but we do not attempt such an implementation here, given that our models are already relatively complex.

In what follows, we will present the implementation and the fits of the activation-based model and the modified direct-access model to the data in turn. We also present quantitative model comparisons, which allow us to assess the relative goodness of fit of each model to the data.

Activation-Based Model

The activation-based model can be implemented as a lognormal race of evidence accumulators (Nicenboim & Vasishth, 2018): For the two experiments, we assume that there are two accumulators of noisy evidence that correspond to the retrieval candidates in memory, namely the first and the second noun phrase (NP1 or NP2, target or distractor). For each trial i, the finishing times FT for NP1 and NP2 are each sampled from a lognormal distribution with location μ_{NP1} or μ_{NP2} respectively, and scale σ. The accumulator with the faster FT_i determines both the

---

[9]Initially, we intended to model the pooled data from both experiments. However, the complex structure of fixed and random effects lead to convergence issues in the modified direct-access model. Therefore, we ended up implementing two separate versions of each model, one for each experiment.

selected picture (target or foil) and the reaction time for trial i. This implementation maps straightforwardly onto the notion of memory chunks with fluctuating activation values, with the chunk with the highest activation being retrieved on a particular trial.

The hierarchical structure of the models is implemented in the μ of both accumulators, which include fixed and random effects. The fixed effects added to μ_{NP1} and μ_{NP2} in the model for pronoun resolution are as follows: group (IWA vs. control), condition (match vs. mismatch), and the group $\times$ condition interaction. We also added the average reaction time from the lexical decision task (LDT), and the group $\times$ LDT interaction. Furthermore, we added the proportion of looks to the target at the critical region (fixations), the fixations $\times$ group interaction, and the three-way interaction fixations $\times$ condition $\times$ group. In addition, both μ included by-subject and by-item varying intercepts; the fixed effect of group included an adjustment by item, and the fixed effect of condition included an adjustment by subject. The parameter σ included a fixed effect of group. In addition, the model for relative clause conditions also included a fixed effect for RC type, an RC type $\times$ group and an RC $\times$ condition interaction, and the RC $\times$ group $\times$ condition three-way interaction. The predictions of the activation-based model for the two experiments are as follows:

1. An increase in fixations to the target picture at the critical region should lead to a decrease in RT for the target accumulator in the picture-selection task, as we assume that the first retrieval influences the second retrieval. If participants retrieved the target at the critical region, re-accessing it at the verb should be easier, meaning that there should be more correct retrievals as well as faster retrieval times.
2. The mean finishing time of the target accumulator should be faster for the mismatch conditions relative to the match conditions, as similarity-based interference slows retrieval. The mean of the distractor accumulator should be similar or slower in mismatch conditions relative to match conditions.
3. IWA should have slower RT relative to controls, so IWA's μ should be higher. This would be in line with the *slow syntax* theory. Similarly, IWA should have a higher σ, that is, more noisy accrual of evidence corresponding to more variable activation values, which would be in line with *intermittent deficiencies*.
4. If a delay in lexical access is causing processing difficulties in IWA, we would expect the effect of LDT to lead to a bigger increase in RTs for the target accumulator for IWA relative to controls, as higher RT for the target accumulator would indicate more

difficulty in retrieving the target. This would be in line with *delayed lexical access*.

In addition, given earlier results (Burchert et al., 2003; Adelt et al., 2017), in Experiment 2, IWA should have longer mean finishing times for the target accumulator in OR compared to SR.

Modified Direct-Access Model

We implement the modified direct-access model as a hierarchical mixture model in the Bayesian framework, following Lissón et al. (2021b). Mixture models integrate multiple generative processes in one model (see Nicenboim et al., 2021 chapter 20, for a tutorial on these models in Stan). The implementation of the modified direct-access model as a mixture model allows us to take into account the probability and cost of backtracking as a latent variable. We assume that both correct and incorrect responses are generated from one of two distributions: Responses without backtracking follow a distribution with parameters μ and σ, while responses with backtracking follow another distribution with parameters $\mu\star = \mu + \delta$ and σ, where δ is the time needed for backtracking.

The direct-access model assumes that the availability of items in memory determines their probability of retrieval. In our implementation, we map availability to the parameter θ, which is the probability of retrieval of the target. Given that interference is expected to affect availability, we add a main effect of condition to θ. Because we expect IWA to have lower base availability compared to control participants, we also add a main effect of group to θ. We also add a main effect of fixations, following the same logic as for the activation-based model: More fixations on the target at the critical region should lead to a higher probability of retrieval of the target at the verb. In order to evaluate the delayed lexical access theory, we include LDT as a fixed effect to θ, and the interaction LDT $\times$ group. This interaction tests the *delayed lexical access theory* in IWA: If longer LDT leads to a larger decrease in θ for IWA, this would suggest that delayed lexical access lowers the probability of retrieval of the target, causing difficulties in the retrieval process.

The original direct-access model assumes that when the initial retrieval fails, a costly process of backtracking (or reanalysis) can be triggered, which leads to correct retrieval of the target (McElree et al., 2003; Martin & McElree, 2008). Our modified direct-access model adds the assumption that backtracking can fail. This is reflected in the added parameter θ_b, which represents the probability of correct retrieval after backtracking. The additional parameter makes the modified direct-access model more

suitable for modeling data from individuals with aphasia, as it allows for slow, incorrect responses. If IWA show a lower θ_b, relative to controls, this could point towards a disruption in the process of backtracking as a main source of comprehension difficulties in IWA. The parameter P_b estimates the proportion of trials for which backtracking is performed after an initial misretrieval. The parameter δ estimates the amount of time (in log ms) that backtracking takes. Main effects of group are added to the parameters θ_b, P_b and δ. P_b and θ_b additionally have by-subject random intercepts.[10]

The mixture process for a given trial i works as follows:

(a) if the retrieval of the target succeeds, with probability θ, RT_i is drawn from LogNormal(μ, σ).

(b) if the retrieval of the target fails $(1 - \theta)$, backtracking is initiated with probability P_b. RT_i is sampled from LogNormal($\mu + \delta, \sigma$). After backtracking, the target is retrieved with probability θ_b, and the distractor with probability $1 - \theta_b$.

(c) if the retrieval of the target fails and there is no backtracking, a misretrieval is predicted with probability $(1 - \theta) \cdot (1 - P_b)$, and RT_i is sampled from LogNormal(μ, σ).

Notice that the probability of successful retrieval of the target, θ, and the probability of backtracking, P_b are assumed to be independent. Interference can only indirectly affect response times through lower θ and the added cost of backtracking δ. Therefore, in the μ parameter, which estimates the mean average RT, we do not include an adjustment by condition, but we do include an adjustment by group, since retrieval may generally be slower in IWA compared to controls. The noise parameter, σ, also has an adjustment by group, as IWA may have more variable retrieval times. The priors used, as well as the full hierarchical model, are shown in Appendix C.

Due to the cost of backtracking δ, correct and incorrect responses following backtracking (b) are expected to be slower, on average, than correct retrievals (a) and misretrievals without backtracking (c). The RTs corresponding to an initial successful retrieval of the target (a) and to misretrievals without backtracking (c) are sampled from the same distribution. The predictions of the modified direct-access model for the two experiments are explained below.

1. Fixations to the target picture at the critical region should lead to an increase in the probability of retrieval of the target, as we assume that the first retrieval influences the second retrieval. Therefore, the estimate

of the main effect of fixations to the target on the probability of successful retrieval θ should be positive.

2. The probability of successful retrieval θ should be higher for non-interference conditions relative to interference conditions, that is, higher in mismatch vs. match conditions.

3. IWA should have slower RTs relative to controls, so IWA's μ should be higher. This would be in line with the *slow syntax* theory. Similarly, IWA should have a higher σ, which would be in line with *intermittent deficiencies*.

4. If IWA's slower access to items from memory leads to difficulties in the retrieval, we would expect LDT to lead to a bigger decrease in θ for IWA relative to controls. This would be in line with *delayed lexical access*.

5. We expect IWA to have a lower probability of backtracking, P_b, and a lower probability of retrieval of the target after backtracking, θ_b. This would be in line with the *resource reduction* theory, assuming that backtracking is a parsing resource that is impaired in IWA.

6. Similarly, we also expect IWA to have a higher cost of backtracking, δ, which would be in line with *slow syntax*.

In addition, given earlier results showing that OR are more difficult to process than SR (Burchert et al., 2003; Adelt et al., 2017), for the relative clause construction, θ should be lower in OR compared to SR.

We now move on to the modeling results, which will be presented separately for pronoun resolution and relative clauses.

Modeling Results

Experiment 1 — Pronoun Resolution

In the pronoun resolution items, NP1 is always the target of the dependency. Therefore, μ_{NP1} accumulates evidence for the retrieval of the target, and μ_{NP2} for the retrieval of the distractor. Figure 3 shows the distribution of estimated finishing times for the two accumulators (NP1 and NP2) across conditions and groups. The results confirm our predictions: IWA have longer finishing times relative to controls in both conditions. In controls, the means of the NP1 accumulator in the mismatch and match conditions are quite similar (1391 ms vs. 1465 ms), although responses are faster on average in the mismatch condition, as expected. IWA show a larger effect of interference: The mean of the NP1 accumulator in the mismatch condition is 4532 ms, compared to 5735 ms in the match condition. The interference effect can also be seen in the overlap of the

---

[10]Ideally, δ should also have a by-subject adjustment. However, this is a complex hierarchical model, and a by-subject intercept on δ led to convergence issues.

Fig. 3 Distribution of the accumulators of evidence across groups and conditions for the Experiment 1 (pronoun resolution). The dashed lines represent the means of the distributions
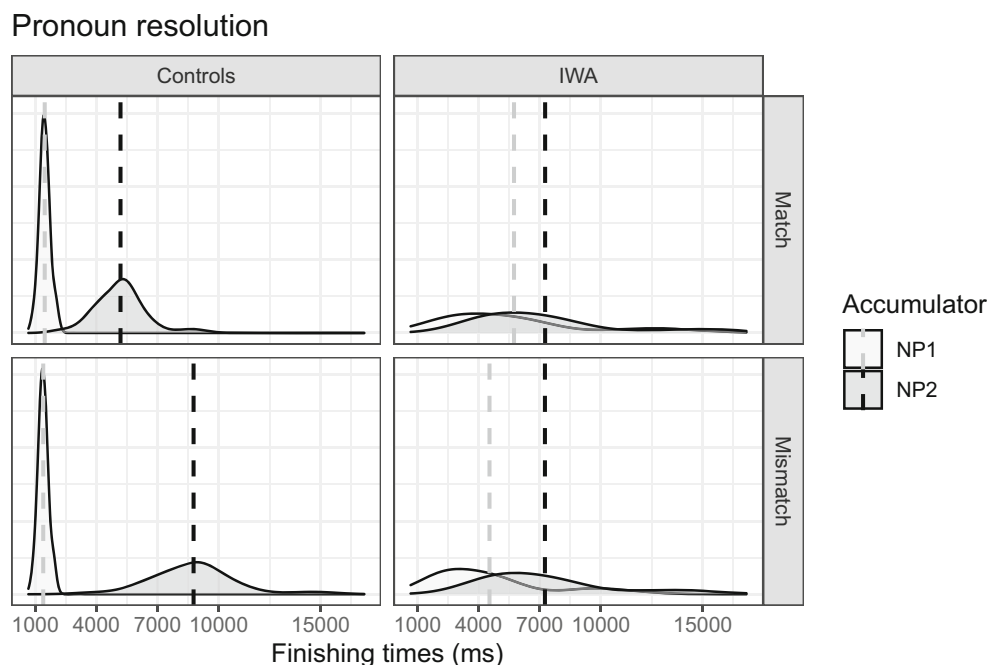
distributions within each plot. Whereas the distributions lie far apart from each other in controls, in IWA, the distributions overlap. This indicates that IWA are more likely to retrieve the distractor than controls, especially in the interference condition (match). In general, the plots show that IWA experience a bigger interference effect. This is in line with the estimates for the group × condition interaction (μ_{NP1}: 199 ms, CrI: [81, 322] ms and μ_{NP2}: 1424 ms, CrI: [400, 2720] ms).

With regard to the fixed effects on μ_{NP1} and μ_{NP2}, due to space limitations, we will only comment on the estimates that are relevant to the processing theories of aphasia that we are evaluating. The estimates for all parameters in this model and their credible intervals are shown in Appendix D.

The NP1 accumulator showed an indication of an LDT × group interaction (836 ms, CrI: [539, 1152] ms), but no indication of such an interaction was observed for the NP2 accumulator. This suggests that additional time needed for lexical access leads to a larger slowdown in IWA in the target accumulator, as predicted by the delayed lexical access theory. The estimates for fixations and the fixations × group interaction do not point in the predicted direction: Rather than facilitating correct retrieval of the target, an increase in fixations on the target picture leads to an increase in RTs in both accumulators (NP1: 41 ms, CrI: [−77, 160] ms; NP2: 518 ms, CrI: [−141, 1212] ms). However, due to large uncertainty around the estimates, the results are also compatible with no effect of fixations.

Finally, as predicted under the intermittent deficiencies theory, IWA have higher noise than controls (σ_{IWA}: 0.65 log ms, CrI: [0.62,0.69] log ms, $\sigma_{controls}$: 0.28 log ms, CrI: [0.27,0.29] log ms).

Modified Direct-Access Model

We begin by assessing the posterior distribution of θ, which is the probability of retrieving the target during the first retrieval attempt. Figure 4 shows that the probability of retrieval of the target is very high for controls: The mean of the distribution lies above 95% in both conditions (CrI mismatch: [98, 99]%, CrI match: [96, 98]%). By contrast, IWA show lower retrieval probabilities overall. This can be also seen in Fig. 4, where IWA's mean estimate for mismatch is 72% CrI: [66, 77]%, whereas the estimate for match is 55% CrI: [47, 62]%. The group × condition interaction is inconclusive (2% CrI: [−1, 5]%).

A unit increase in LDT leads to −5% CrI: [−8, −1]% in θ, and a negative LDT × group interaction (−9% CrI: [−12, −7]%) is consistent with the assumption that IWA are more affected by increased LDT. There was neither an indication of an effect of fixations (2% [−1, 5]%), nor of a fixation × group interaction (−2% CrI: [−5, 2]%). This means that for both groups, there is no indication that an increase in fixations to the target picture led to an increase in the probability of successful retrieval of the target.

The estimated probability of backtracking for IWA is 22% CrI: [13, 31]% compared to 66% CrI: [51, 79]% for controls. The distribution of the cost of backtracking, δ, is centered around 5592 ms, CrI: [3924, 7738] ms for IWA, and around 2827 ms, CrI: [2277, 3551] ms for controls. The

Fig. 4 Posterior distribution of the probability of initial retrieval of the target, θ for the two groups, in the pronoun conditions. The vertical lines stand for the means of the distributions, and the shaded areas indicate the 95% credible interval

Pronoun conditions, posterior distribution of θ
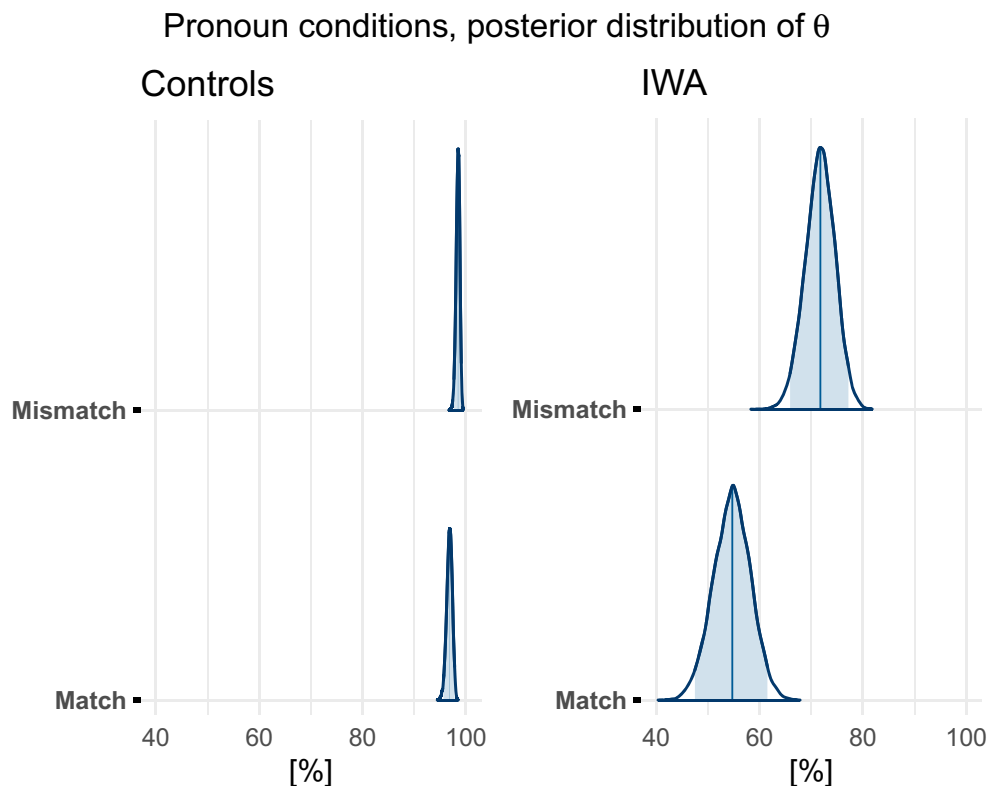
probability of retrieval of the target after backtracking, θ_b, is shown in Fig. 5. Backtracking leads to the retrieval of the target around 84% of the time for controls (CrI: [70, 94]%), and 58% of the time for IWA (CrI: [42, 73%]). Slower and less successful backtracking is consistent with slow syntax and resource reduction in IWA.

Finally, IWA's μ (2376 ms, CrI: [2079, 2701] ms) is higher than controls' μ (1320 ms, CrI: [1202, 1447] ms); and IWA also have a higher noise estimate (0.46 log ms, CrI: [0.43,0.5] log ms) relative to controls (0.24 log ms, CrI: [0.23,0.25] log ms), as predicted under the slow syntax and intermittent deficiencies theories.

Experiment 2 — Relative Clauses

Activation-Based Model

In the relative-clause items, the accumulator mean μ_{NP1} stands for the retrieval of NP1 as the agent of the action, whereas μ_{NP2} stands for the retrieval of NP2 as the agent. Depending on the trial, NP1 (in subject relatives) or NP2 (in object relatives) will be the target of the retrieval, as we model the retrieval of the agent.

Figure 6 shows the distribution of finishing times of the two accumulators in subject relative clauses. As expected, IWA have higher finishing times than controls across conditions. The mean of the NP1 accumulator (target) is roughly the same across conditions, whereas the mean of the NP2 accumulator (distractor) is higher in the mismatch condition than in the match condition. In general, controls show almost no overlap between the distributions, which indicates that controls retrieve the target (NP1) most of the time. By contrast, in IWA, the two distributions partially overlap, meaning that IWA often retrieve the distractor (NP2).

Figure 7 shows the distribution of finishing times of the two accumulators in object relative clauses. IWA have higher finishing times than controls across conditions, and
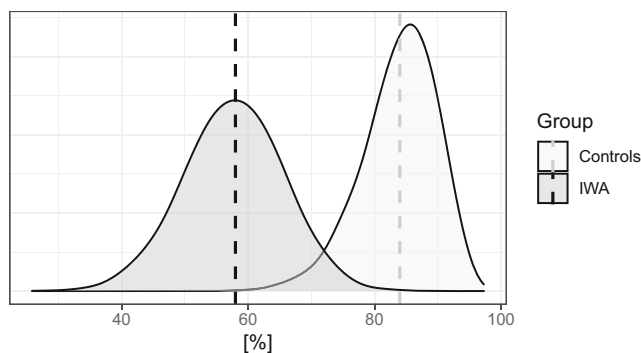

Posterior distribution of θ_b

Fig. 5 Posterior distribution of the probability of retrieval of the target after backtracking, for the two groups, in the pronoun conditions. The dashed lines stand for the means of the distributions

Fig. 6 Distribution of the accumulators of evidence across groups and conditions for subject relative clauses. The dashed lines indicate the means of the distributions
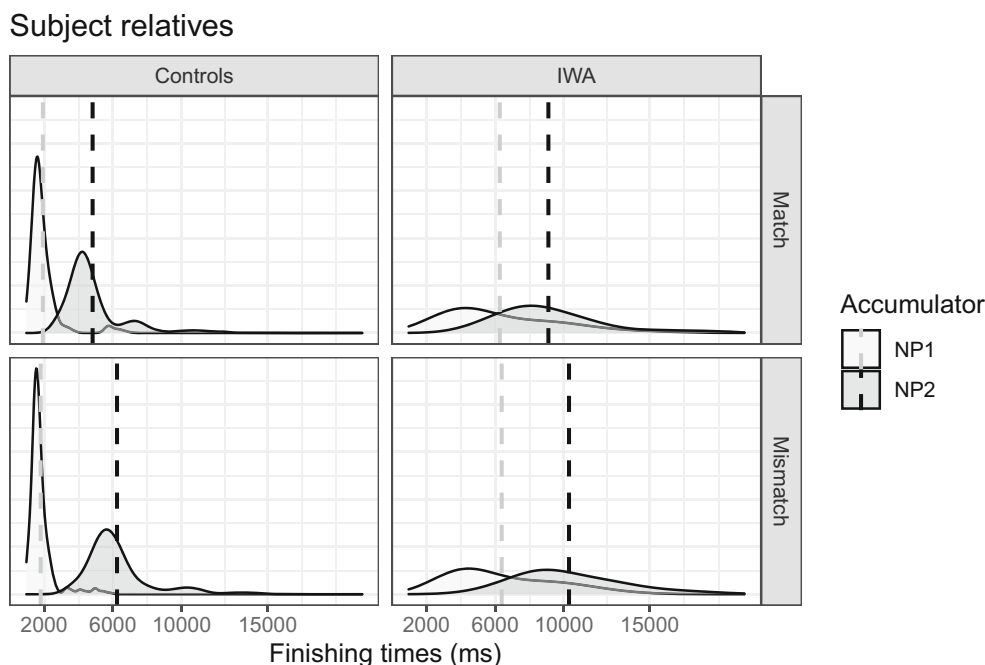

both groups have slightly lower finishing times in the NP2 accumulator (target) in mismatch vs. match conditions. Crucially, in the match condition, for IWA (right upper panel in Fig. 7, light dashed line), the mean of the NP1 accumulator is lower than the mean of the NP2 accumulator. Since NP2 is the retrieval target in object relatives, the pattern indicates that in the match condition, IWA retrieve the distractor more often than the target. That is, in the match condition, IWA are more likely to misinterpret the sentence than to interpret it correctly. However, in the

mismatch condition, the mean of the two accumulators overlap, which indicates that IWA are equally likely to retrieve NP1 or NP2 on average.

Comparisons between Figs. 6 and 7 show that controls perform similarly in subject and object relatives, whereas IWA display a subject-object asymmetry: IWA are estimated to correctly interpret subject relatives most of the time. By contrast, IWA are estimated to misinterpret object relatives more often, especially in the match condition.

Fig. 7 Distribution of the accumulators of evidence across groups and conditions for object relative clauses. The dashed lines indicate the means of the distributions. Note that in the mismatch condition, for IWA, the means of the two distributions overlap
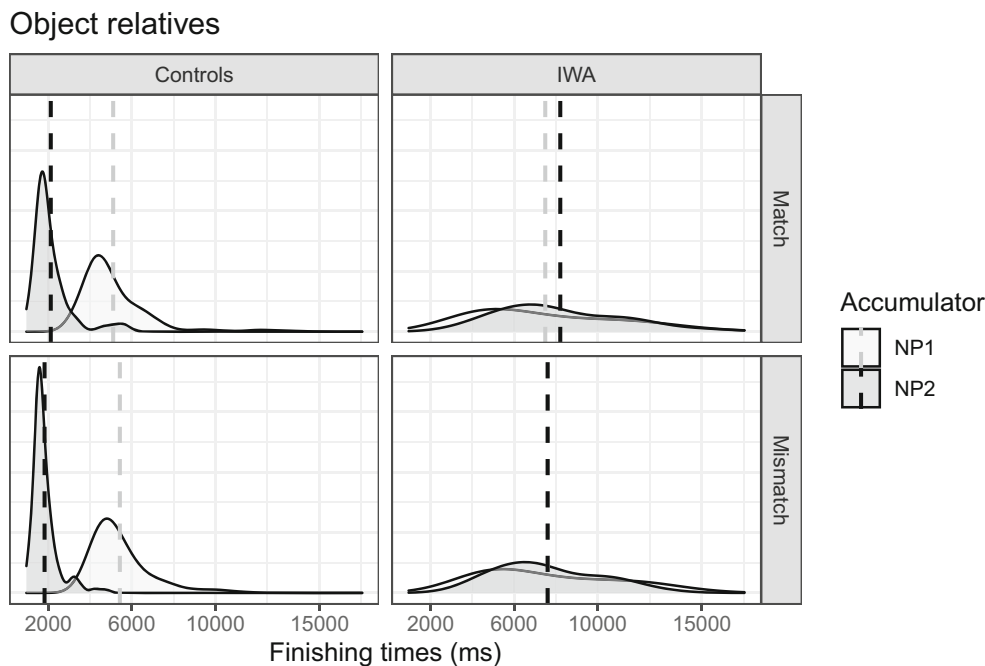
The model estimates for the fixed effects and interactions on μ_{NP1} and μ_{NP2} are shown in Appendix D. No indication of an effect was found for condition or the condition × group interaction, but there was a RC type × condition interaction on μ_{NP2} (631 ms, CrI: [385, 884] ms): Interference (match) in OR lead to higher finishing times for μ_{NP2} relative to no-interference (mismatch). The three-way interaction RC type × group × condition for μ_{NP2} (−249 ms, CrI: [−491, −10] ms) indicates that the effect of condition within RC is different for the two groups in the μ_{NP2} accumulator: In the SR trials, the difference between match and mismatch conditions is bigger for controls. By contrast, in OR trials, the difference between match and mismatch conditions is bigger for IWA.

There was no indication of an LDT × group interaction, a LDT × condition interaction, or a LDT × condition × group interaction. There was an effect of fixations on μ_{NP2} (289 ms, CrI: [61, 521] ms). This main effect is uninformative, given that NP2 was the retrieval target in OR but not in SR. There was no indication of a fixations × condition interaction or a fixations × group × condition interaction, so that the role of fixations remains inconclusive. Finally, as predicted, IWA have higher noise than controls (σ_{IWA} 0.55 log ms, CrI: [0.53, 0.57 log ms], $\sigma_{controls}$ 0.31 log ms, CrI: [0.3, 0.32 log ms]).

Modified Direct-Access Model

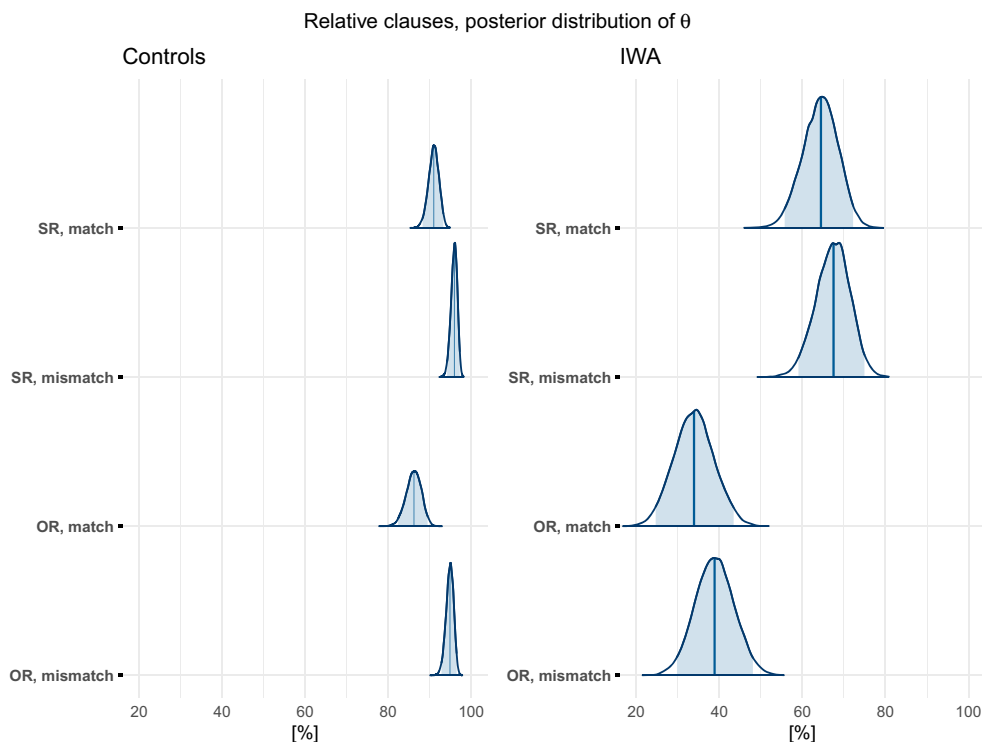The posterior distributions of θ, the probability of initial retrieval of the target, by group and condition are displayed in Fig. 8. While controls have a slightly lower θ in OR relative to SR in the match conditions, SR and OR have a similar θ in mismatch conditions, around 95%. This indicates that, in line with the model predictions, mismatch facilitates the retrieval of the target, especially in OR. The number mismatch also benefits IWA on average, but IWA exhibit a stronger subject-object asymmetry, irrespective of the number manipulation, with higher θ in SR relative to OR for both match and mismatch conditions.

The estimates of the model confirm the patterns shown in Fig. 8. The effect of condition (−10% CrI: [−13, −6]%) suggests that match conditions elicit a lower θ across the board, but a condition × group interaction (7% CrI: [3, 10]%) suggests that the effect of condition is stronger for controls than for IWA. There was no indication of a RC type × condition interaction (−1% CrI: [−5, 2]%), nor of a RC type × condition × group interaction (1% CrI: [−3, 4]%). The effect of RC type (−13%, CrI: [−18, −9]%) and the RC type × group interaction (−7%, CrI: [−12, −3]%) indicate that object relatives are more difficult to process than subject relatives, and more so for IWA than for controls.

There was no indication of an effect of LDT on θ (−3%, CrI: [−8, 2]%), nor of a group × LDT interaction (3%, CrI: [−2, 8]%). A unit increase in fixations led to −2%, CrI: [−5, 1]% on θ, and the interaction group × fixations 4%, CrI: [1, 7]% suggests that the effect of fixations is different for IWA and control participants: In IWA, an increase in fixations of looks to the target leads to a higher θ.

The estimated probability of backtracking given an initial incorrect retrieval is 35% CrI: [20, 50]% for IWA, whereas

Fig. 8 Posterior distribution of the probability of retrieval of the target (θ) across groups and conditions in relative clauses. The vertical lines stand for the means of the distributions, and the shaded areas represent the 95% credible interval
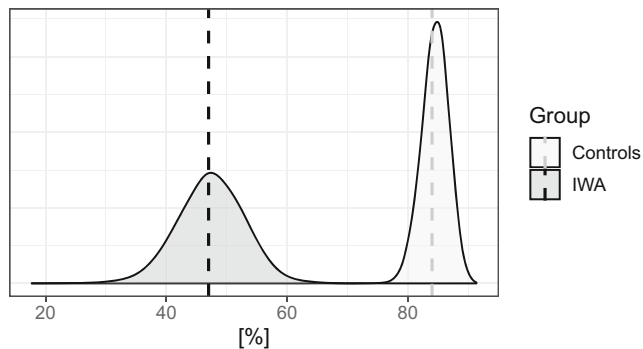

Relative clauses, posterior distribution of θ

Posterior distribution of θ_b

Fig. 9 Posterior distribution of the probability of retrieval of the target after backtracking (θ_b) across groups in the relative clauses experiment

for controls, it is 87% CrI: [83, 91]%. The posterior distribution of θ_b is shown in Fig. 9: After backtracking, IWA retrieve the target about half of the time. By contrast, controls retrieve the target more than 80% of the time. In addition, IWA are estimated to need 3457 ms, CrI: [2425, 4659] ms for backtracking, whereas controls need 1829 ms, CrI: [1637, 2032] ms.

Finally, as predicted, μ is higher for IWA (3744 ms, CrI: [3304, 4234] ms) than for controls (1613 ms, CrI: [1492, 1736] ms), and σ is also higher for IWA (0.41 log ms, CrI: [0.38 log ms, 0.44 log ms]) than for controls (0.24 log ms, CrI: [0.23 log ms, 0.24 log ms]).

Discussion

In the two models, the location and scale parameters (μ and σ) of the RT distribution were consistently higher for IWA than for controls. We linked these parameters to the slow syntax and intermittent deficiencies theories, respectively. Both models thus seem to be generally in line with these two theories of processing deficits in aphasia. We will now discuss the implications for the remaining theories within each model.

Activation-Based Model

We hypothesized that the accumulators in the activation-based model should reflect the interference effect predicted by cue-based retrieval theory, namely, lower mean finishing times for the target accumulator, and similar or higher mean finishing time for the distractor accumulator in the mismatch conditions compared to the match conditions. The accumulators show this pattern across the two experiments. In addition, in Experiment 2, the distribution of the accumulators across relative clause types shows that IWA experience a subject-object asymmetry, that is, IWA have

more difficulties processing object relatives, in line with previous findings.

The conclusions for the rest of our predictions are more complex, since the results differ across the two experiments. For instance, a group × LDT interaction was found for the target accumulator in pronoun resolution. This interaction indicates that slower lexical access leads to increased processing difficulty for IWA, as predicted by the delayed lexical access theory. However, there was no indication of such an interaction in relative clauses. We therefore conclude that more research is needed in order to establish the role of delayed lexical access in the activation-based model.

The effect of looks to the target at the critical region also remains inconclusive. No effect of fixations was found in pronoun resolution. In relative clauses, an effect of fixations was found for the NP2 accumulator, but no indication of an interaction between fixations and RC type was found. Given that NP2 was the retrieval target in OR but not in SR, the main effect of fixations is uninformative.

Modified Direct-Access Model

We expected similarity-based interference to result in a lower probability θ of successful retrieval for the target in the match conditions compared to the mismatch conditions. The data from both experiments are in line with this prediction. In addition, in Experiment 2, IWA show a large effect of relative clause type, irrespective of the condition: IWA have more difficulties understanding object relatives compared to subject relatives. This subject-object asymmetry is broadly in line with the accuracies in Adelt et al. (2017), although Adelt et al. (2017) found this pattern in both IWA and controls.

The probability of backtracking is consistently lower for IWA than for controls, as is the probability of retrieval of the target after backtracking (θ_b). This pattern is expected under the resource reduction theory, assuming that backtracking makes use of parsing resources. In addition, the average cost of backtracking, δ, is twice as high for IWA compared to controls in both experiments. This adds support for the slow syntax theory.

According to the delayed lexical access theory, IWA should be more affected by delays in lexical access, as measured by a lexical decision task. The observed group × LDT interaction lends some support to this theory in the pronoun resolution sub-experiment, but not in relative clauses. Therefore, the effect of delayed lexical access in the modified direct-access model remains inconclusive, as for the activation-based model. The effect of fixations is also inconclusive: Although in relative clauses there is some indication that fixations at the critical region may lead to a increase in the probability of retrieving the target

for IWA, no effect of fixations was found in pronoun resolution.

Model Comparisons

The activation-based model and the modified direct-access model make different assumptions about the retrieval mechanism, and thus the generative process behind the observed data. Within the framework of each model's assumptions, conclusions can be drawn about plausible underlying deficits. However, one crucial question remains open: Which model captures the generative process better? In order to answer this question, we performed 10-fold cross-validation (Vehtari, Gelman, & Gabry, 2017; see also Nicenboim et al., 2021, chapter 17, for a tutorial on carrying-out cross-validation for Bayesian models such as the ones discussed here). This is a standard procedure in machine learning for quantifying the relative predictive fit of two or more models. Importantly, cross-validation can also be applied when the models assume different generative processes, as is the case with the activation-based and the modified direct-access models.

The procedure for 10-fold cross-validation is as follows: The data are partitioned into 10 balanced subsets containing about the same amount of data per subject.[11] One of the 10 subsets is held out, and the model is fit to the remaining subsets. The posterior distributions from the resulting model are used to compute predictive accuracy on the held-out subset. This is repeated 10 times, so that all subsets are covered. The expected log pointwise predictive density, $\widehat{elpd}$, is then calculated as a measure of predictive accuracy. $\widehat{elpd}$ is the summed log-likelihood of all observed, held-out data points under each model. Models are compared by computing the difference in $\widehat{elpd}$, ($\widehat{\Delta elpd}$), with higher $\widehat{elpd}$ indicating better predictive fit. Because $\widehat{elpd}$ is an estimate, the difference in $\widehat{elpd}$ between two models has an associated standard error, which has the standard frequentist interpretation: $\widehat{\Delta elpd} \pm 2 \times$ SE gives a 95% confidence interval. If the difference in $\widehat{elpd}$ between the models is greater than $2 \times SE$, we conclude that there are grounds to assume that the model with the higher $\widehat{elpd}$ provides the better fit for the given data.

The results of the cross-validation are shown in Table 1. In pronoun resolution, the modified direct-access model has a predictive advantage, but since the SE of $\widehat{\Delta elpd}$ is large, the result is not conclusive. In relative clauses,

---

[11]Given that our dataset contains data from two experimental sessions (test and retest), an alternative way to perform cross-validation would be to train the models on the test data and to use the retest data to compute predictive accuracy. However, we chose to use the pooled data from each of the two experiments, in order to maximize the amount of data, especially for IWA.

Table 1 Differences in $\widehat{elpd}$ between the two models and their corresponding SE

Model	$\widehat{\Delta elpd}$	SE
Pronoun resolution	−109	133
Relative clauses	403	167

Positive differences indicate an advantage for the activation-based model, whereas negative differences indicate an advantage for the modified direct-access model

the activation-based model has a clear advantage over the modified direct-access model, but the advantage is mostly driven by the control participants, as shown in Appendix E.

We also evaluated the predictive performance of the original direct-access model, that is, a model in which backtracking can only lead to the retrieval of the target. The results are shown in Appendix F. In relative clauses, the activation-based model outperforms both the original and the modified-direct access model, while the result for pronoun resolution is inconclusive.

General Discussion

This is the first-ever computational investigation of competing models of similarity-based interference in German language comprehension in IWA and unimpaired controls. We investigated interference in two linguistic constructions, namely pronoun resolution and relative clauses. Two models of cue-based retrieval were implemented in a Bayesian framework: The activation-based model of Lewis and Vasishth (2005) and a modified version of the direct-access model of McElree (2000) as implemented by Nicenboim and Vasishth (2018). The activation-based model assumes a direct connection between retrieval latency and retrieval probability for memory items, whereas the modified direct-access model assumes a constant retrieval latency, along with a costly backtracking mechanism that triggers when retrieval fails. In the original direct-access model, backtracking leads to the correct retrieval of the target item from memory (McElree, 1993). In our modified direct-access model, backtracking can fail, leading to a costly misretrieval. We argue that this is a more suitable model for individuals with aphasia, as it can account for slow incorrect responses, a pattern that is frequently found in the aphasia literature (Hanne et al., 2015; Adelt et al., 2017; Lissón et al., 2021b; Pregla et al., 2021).

The predictive performance of the two models was compared against data from a visual-world experiment (Pregla et al., 2021), using the reaction time and accuracy in the picture selection task as dependent variables. Looks to the target at the critical sentence region, where retrieval is

assumed to occur, were used as a predictor, along with the mean reaction times from a lexical decision task. We linked the parameters of each computational model to prominent theories of processing deficits in aphasia, aiming to answer two main questions: (a) Which model is better able to fit the data from IWA and control participants across the two experiments? and (b) What do the parameters in each model tell about the processing deficits and about interference in IWA? We will now discuss the answers to these questions, as well as the relation of our results to prior work in computational modeling of processing deficits in aphasia.

First, both models of retrieval perform well across the two linguistic constructions tested, in the sense that the relevant parameters are affected in the expected direction by group differences and by similarity-based interference. The activation-based model outperforms the modified direct-access model in the relative clauses experiment, mainly because it provides a better predictive fit for the data from control participants. However, both models perform similarly at fitting data from IWA. In pronoun resolution, the two models show similar predictive fit across groups and across conditions.

Second, with regard to the underlying processing deficits in aphasia, both models are in line with slow syntax (Burkhardt et al., 2003; Burkhardt et al., 2008) and intermittent deficiencies (Caplan et al., 2013). Resource reduction (Caplan et al., 2007; Caplan, 2012), as implemented here, can only be evaluated with respect to the modified direct-access model, and the results show that the model is in line with this deficit. There was no strong indication in our data, across the two experiments and for both models, that delayed lexical access (Love et al., 2008; Ferrill et al., 2012) is a source of processing deficits in IWA: The predicted relationship between individual lexical decision latency and participant group was only found in some conditions. More experiments are needed in order to explore the role of this deficit.

Regarding the effect of similarity-based interference, based on the results for the activation-based model, we can conclude that in pronoun resolution, IWA are more sensitive to gender interference than control participants. The interaction group × condition was inconclusive in the modified direct-access model. Both models estimate that in relative clauses, the effect of number interference is rather small for both groups. The models suggest that IWA experience a subject-object asymmetry, whereas control participants do not. Below, we discuss the comparatively small effect of number mismatch in relative clauses and some possible explanations of the subject-object asymmetry in IWA.

Number Mismatch Versus Subject-Object Asymmetry

The results of the models show that for IWA, the presence of two candidate NPs with distinctive number features is of limited use in both subject and object relatives with regard to successful comprehension (see also the descriptive statistics for relative clauses, split by condition, in Appendix B). The modified direct-access model estimates no effect of number mismatch for IWA, while pronouns did show some indication of a gender mismatch effect. The activation-based model estimates that number mismatch between the NPs in object relative clauses slightly increases the probability of retrieving the target; however, the target is only retrieved about half the time in these conditions. For IWA, the main difference is between subject and object relative clauses, not between high- and low-interference conditions, as shown in Fig. 10. The subject-object asymmetry in relative clauses in IWA is in line with the canonicity effects reported in several German studies with IWA (e.g., Burchert et al., 2003; Burchert & de Bleser, 2004; Hanne et al., 2011; Adelt et al., 2017; Pregla et al., 2021). Canonicity effects refer to the fact that sentences with non-canonical word order (e.g., object-subject-verb in German) are more difficult to process than sentences with canonical word order.

In contrast to English, where subject and object relative clauses are distinguished by the subject NP intervening or not intervening in the object-verb dependency, cue-based retrieval cannot explain the subject-object asymmetry in German: Both in subject and in object relative clauses, the verb is clause-final and two NPs have to be retrieved, one that is adjacent to the verb and one that is not. Consequently, cue-based retrieval would predict no processing difference between subject and and object relatives in German.

One possible explanation for the differential effects of number marking and RC type is that German relative clauses feature case marking, and that IWA may pay more attention to case than to number cues. Case marking may be more difficult to process in German object relative clauses compared to subject relative clauses, given that there is a case mismatch between the target noun phrase and the relative pronoun in the former. One influential proposal is that of *case attraction*. Case attraction is analogous to the well-studied phenomenon of number attraction, which has been explained in terms of feature percolation (e.g., Schlesewsky, 1996; Eberhard, 1997; Nicol, Forster, & Veres, 1997; Bader & Meng, 1999; Fanselow, Schlesewsky, Cavar, & Kliegl, 1999; Logačev & Vasishth, 2012; Czypionka, Dörre, & Bayer, 2018). If the head noun and the relative pronoun mismatch in case, as in object relatives, the [+nom-

Fig. 10 Descriptive statistics for the relative clauses experiment. The dots stand for the means, and the error bars show the standard error of the means
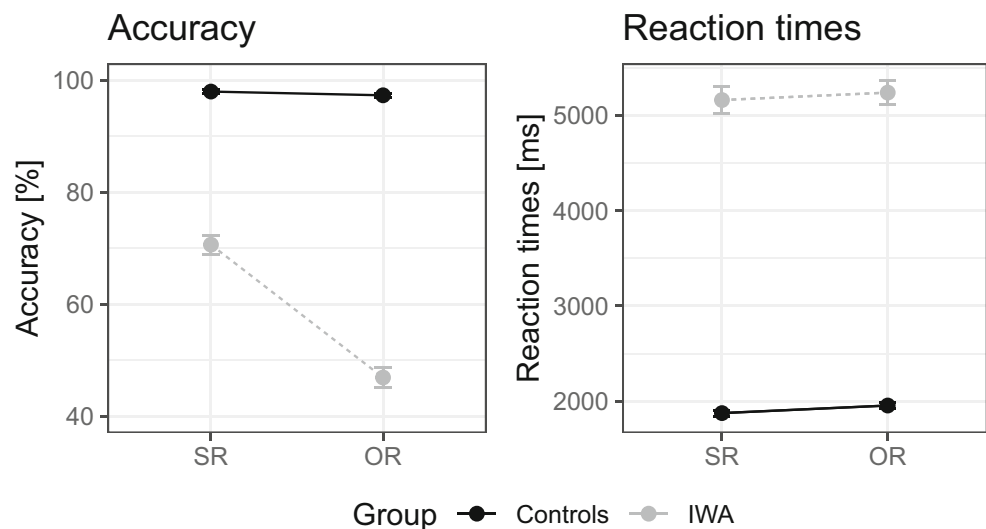
inative] case feature of the head noun could percolate down to the relative pronoun, overriding its original [+accusative] case feature. Because the case feature of the relative pronoun signals the syntactic role of the relativized noun phrase, the object relative could be misinterpreted as a subject relative. The case attraction theory thus predicts that subject relatives are easier to process than object relatives, because the head noun and the relativizer have an identical [+nominative] feature in subject relatives.

The proposal that IWA pay more attention to case than to number cues can be seen a differential weighting of retrieval cues. The cue-weighting proposal, implemented by Engelmann (2016) in the framework of the Lewis and Vasishth (2005) model, claims that depending on the linguistic structure, some retrieval cues may receive more weight, and therefore contribute more strongly to memory activation, than others (Dillon et al., 2013; Cunnings & Sturt, 2014; Parker & Phillips, 2017; Engelmann et al., 2019; Vasishth et al., 2019). Differences in weighting between case and number cues could be integrated in both the activation-based and the direct-access model, possibly at the individual participant level, as recently proposed by Yadav, Paape, Smith, Dillon, & Vasishth (2021).

Studies that have investigated processing of number and case in IWA in German provide mixed evidence with regard to differential cue weighting. For instance, Hanne et al. (2015) investigated IWA's use of case and number cues to interpret semantically reversible SVO vs. OVS sentences in German. Their data indicate that processing of case marking may be more impaired than processing of number marking. This contrasts with the results in Adelt, Burchert, Adani, & Stadie (2020), who tested case-unambiguous vs. case-ambiguous, number-disambiguated object relatives. The

authors found that IWA have a general processing advantage in the case-unambiguous conditions. The study of Adelt et al. (2020) supports the idea that IWA may rely more on case cues than on number cues in relative clauses. However, neither Adelt et al. (2020) nor Hanne et al. (2015) included both case and number cues within the same items. Our modeling shows that when both case and number cues are included in a sentence, IWA do not benefit from the extra number cue, suggesting that cue weighting may be a factor.

Comparison with Previous Work

Lissón et al. (2021a) investigated English relative clause processing in IWA vs. controls using a large-scale dataset from Caplan et al. (2015). Lissón et al. found an agent-first bias for control participants in English: In non-canonical clauses, such as object relatives, unimpaired controls tend to initially assign the agent role to the first noun phrase in the sentence, which is incorrect in object relatives. By contrast, IWA do not show an agent-first bias. The agent-first bias in unimpaired controls has been attested in visual-world studies in both English (Mack, Wei, Gutierrez, & Thompson, 2016, passives) and German (Hanne et al., 2015, OVS sentences; Hanne et al., 2015, object relatives). In these studies, control participants initially show increased looks to the foil picture in non-canonical sentences. The foil picture in non-canonical sentences depicts the canonical interpretation of the sentence. As soon as they hear the relevant morphological cues (e.g., the relativizer in unambiguous German relative clauses), control participants start looking at the target picture.

In the study of Lissón et al. (2021a), which modeled self-paced listening data, this processing bias was reflected

in the estimates for the direct-access model. Controls had a lower probability of initial correct retrieval than IWA in object relatives (controls 40%, IWA 50%). That is, controls were estimated to initially retrieve the distractor more often than the target in object relatives. The model could still account for the higher accuracy of controls by assuming a high probability of backtracking for controls (80%) relative to IWA (20%). This pattern supported the notion that controls initially processed the first noun phrase in object relatives as the agent, until they revised and corrected their interpretation by backtracking. Surprisingly, in the present modeling of the German data, we do not see such an agent-first bias. In the present study, controls did not have a lower θ relative to IWA, and controls' θ was always above 90% for all conditions. One possibility is that in our modeling, the RT at the end of the sentence do not reflect the agent-first bias, given that the initial interpretation has already been revised at this point.

With respect to model comparisons, Lissón et al. (2021a) found that the activation-based model had a predictive advantage over the original direct-access model, although the difference was not decisive, given the large standard error of $\widehat{\Delta elpd}$. In the present study, we show that in relative clauses in German, the activation-based model provides a better fit than both the original and the modified direct-access model. Our results contrast with Nicenboim and Vasishth (2018), who compared the predictive performance of the activation-based model and the original direct-access model using self-paced reading data from unimpaired controls in German (Nicenboim et al., 2018). Nicenboim and Vasishth (2018) found that the original direct-access model provided a better fit to their data. However, the data modeled in Nicenboim et al. (2018) differ from our data in one important aspect: In correct trials (that is, trials with correct responses to the comprehension question), RTs were on average higher than the RTs in incorrect trials. This pattern in Nicenboim et al.'s data is crucial, because higher RTs for correct responses is what the original direct-access model assumes. Consequently, the advantage in predictive performance of the original direct-access model vs. the activation-based model came from the slow, correct responses. By contrast, our data show the opposite pattern. Correct responses are, on average, faster than incorrect responses, especially for IWA. Although in the present paper we have implemented a modified direct-access model that can account for slow incorrect responses, the cross-validation shows that the activation-based model outperforms both the original and the modified-direct access model in relative clauses in German.

Overall, our results are in line with previous modeling work in sentence comprehension in aphasia using the activation-based model (Patil et al., 2016; Mätzig et al., 2018; Lissón et al., 2021a). If we assume that sentence comprehension is mediated by an activation-based model of cue-based retrieval, the performance of IWA can be explained by a combination of processing deficits, namely slow syntax and intermittent deficiencies.

Limitations and Future Directions

The data that we modeled in this paper (Pregla et al., 2021) is the largest-ever compilation of online measures for IWA in German. Nevertheless, the size of the IWA group (21 subjects) remains relatively small when compared to the number of subjects tested in typical eye-tracking experiments with unimpaired participants. Collecting online data from impaired populations is difficult, which is why most studies in the aphasia literature have a smaller number of participants. Usually, online experiments have 3 to 12 IWA and 10 to 20 control participants (e.g., Burkhardt et al., 2003; Love et al., 2008; Dickey & Thompson, 2009; Choy & Thompson, 2010; Hanne et al., 2011; Mack, Ji, & Thompson, 2013; Hanne et al., 2015; Mack et al., 2016; Adelt et al., 2017; Engel et al., 2018). An exception is the data presented in Caplan et al. (2013, 2015), with more than 50 IWA, although it does not include eye-tracking data.

The activation-based model outperforms the modified direct-access model in relative clauses, but both models perform similarly in pronoun resolution. We believe that the inability to find a clear answer in pronoun resolution has to do with the inherent limitations of the sample size, both in terms of subjects and items. Crucially, the relative clause experiment tested 40 items per subject (10 items per condition, 4 conditions), whereas the pronoun resolution tested 20 items per subject (10 items per condition, 2 conditions). Future work comparing these models will require much more data in order to distinguish between the models.

We have reported the most complex hierarchical structure that yielded converging fits, but in order to account for individual differences, an even more complex hierarchical structure would be necessary. This is especially true for the modified direct-access model, where an individual adjustment for the effect of δ would help in understanding the variability in the process of backtracking. The same holds for both models regarding the effect of eye fixations in the visual world paradigm. In addition, this work focuses on average, group-level effects. While the comparisons between groups yield an estimate of the average performance, it has been shown that IWA have large within and between-subject variability (Patil et al., 2016; Mätzig et al., 2018; Pregla et al., 2021).

A future direction is to develop an individual-level modeling approach, in which parameters are estimated for each subject, in a similar vein to the modeling work in Mätzig et al. (2018) for offline data. Recent modeling work shows that under the cue-based retrieval, even among unimpaired participants, individual differences can modulate interference effects (Yadav et al., 2021). Therefore, obtaining individual parameter estimates for IWA would be more informative regarding both interference effects, and the extent to which each processing deficit plays a role for each IWA. Furthermore, the possible role of feature percolation, that is, of target features being overwritten by features of distractor items, has been recently investigated by Yadav, Smith, & Vasishth (2021). Their results support the view that in unimpaired populations, feature percolation feeds retrieval processes, suggesting that a hybrid model may better explain similarity-based interference.

A second limitation of this work concerns the implementation of the models. In order to compare the two competing models in a common architecture, we implemented simplified versions that focus on a single retrieval event. However, in order to account for sentence processing deficits in IWA across the whole trial, the models should include a parser. The original Lisp implementation of the activation-based model in ACT-R (Anderson et al., 2004; Lewis & Vasishth, 2005) includes a left-corner parsing algorithm that could conceivably be added to our model. However, as far as we are aware, there exists no computational implementation of a parser in the (modified) direct-access model. A future direction would be to incorporate a parser into the (modified) direct-access model, and to fit Bayesian versions of both models that account for individual parsing steps. This is especially challenging because it requires finding a common architecture that supports a parsing algorithm for both models, so that model comparisons can be performed.

Finally, another limitation concerns the use of looks to the target at the critical region as an index of retrieval. An analysis of eye fixation patterns across the entire trial would possibly be more informative regarding the time-course of interference during sentence processing. This is especially true for relative clauses, where our results show that higher fixations to the target at the critical region do not lead to faster reaction times in the picture selection task. Specific modeling techniques for visual-world-data could be considered, such as growth curve analysis (Mirman, 2017) or divergence point analysis (Stone, Lago, & Schad, 2020). Integrating these analyses with our computational modeling approach may be possible, and may yield important insights into the differences between IWA and controls over the course of the trial.

Conclusion

We conducted the first large-scale evaluation of two computational models of sentence processing in individuals with aphasia (IWA) in German. Our study tested two competing models of cue-based retrieval — the activation-based model and a modified version of the direct-access model — against online and offline data from IWA and control participants. The data came from a visual-world eye-tracking experiment with a picture selection task. Similarity-based interference was manipulated in two linguistic constructions, namely pronoun resolution and relative clauses. Reaction times from the picture selection task were modeled as a function of interference, group (IWA vs. control), lexical access speed, and fixations to the target picture at the critical region of the sentence. The results show that in pronoun resolution, IWA experience greater gender interference effects relative to control participants. In relative clauses, the data suggest that IWA exhibit a larger subject-object asymmetry than controls. In IWA, the subject-object asymmetry is much stronger than the effect of number interference, suggesting that in relative clauses, IWA may rely more strongly on case cues compared to number cues. The parameter estimates from both implemented models are in line with the slow syntax and the intermittent deficiencies accounts. In addition, the parameters of the modified direct-access model are also in line with the resource reduction theory. The cross-validation results show that while both models have similar quantitative performance for the pronoun structures, the activation-based model outperforms the modified direct-access model in relative clauses.

Appendix A: Details about the individuals with aphasia

Table 2, adapted from Pregla et al. (2021), shows a summary of the demographics and neurological information about the individuals with aphasia that took part in the experiments.

Table 2 Demographic and neurological data of the individuals with aphasia that took part in the experiment, adapted from Pregla et al. (2021)

IWA	Gender	Age	Education	Years P.O.	Etiology	Localization	T3	T11	Aphasia type
2	F	72	8	7	IMI	L	77	19	Anomic
3	M	76	20	17	IMI	L/R	61	20	Not-classifiable
4	F	47	13	21	IMI	L	78	20	Anomic
6	M	55	14	10	IMI	L	67	20	Anomic
8	F	51	19	7	MA	L	74	20	Anomic
9	M	64	15	2	IMI	L	73	20	Anomic
10	M	58	18	1	IMI	L	52	20	Broca
11	F	63	12	1	IMI	L	73	20	Broca
12	F	46	12	13	IMI	L	65	20	Broca
13	M	74	13	8	IMI	L	57	20	Broca
14	M	66	13	17	IMI	L	75	20	Anomic
15	F	59	21	4	I	L	77	20	Broca
16	M	67	17	26	VH	R	72	19	Broca
17	F	43	14	10	IMI	L	65	20	Broca
18	M	57	13	1	I	L	67	18	Wernicke
19	F	52	19	8	IMI	L	76	20	Broca
20	M	38	13	3	IMI	L	73	19	Broca
21	M	57	18	2	IMI	L	66	18	Broca
22	F	67	16	5	IMI	L	76	20	Anomic
23	M	74	15	7	IMI	L	67	20	Anomic

Note: Age and education are given in years. T3 and T11 are part of the LEMO 2.0 (Stadie et al., 2013) battery of tests and the scores given in the table are raw scores; T3 ($n=80$) = auditory lexical decision, T11 ($n=20$) = auditory word-picture matching. *IWA*, individual with aphasia; *P.O.*, post-onset; *F*, female; *M*, male; *L*, left; *R*, right; *IMI*, ischemic arteria cerebri media infarct; *I*, infarct; *MA*, arteria cerebri media aneurysm; *VH*, vertebrobasilar hemorrhage

Appendix B: Descriptive statistics

Figure 11 shows descriptive statistics for the pronoun sub-experiment by group and condition. Figure 12 shows descriptive statistics for the relative clause sub-experiment by group and condition.
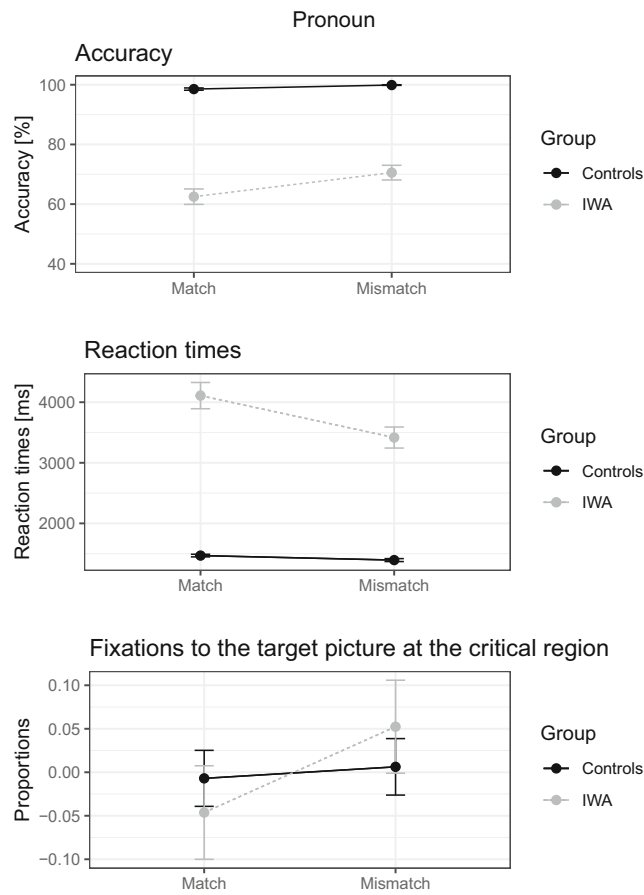

Fig. 11 Descriptive statistics for the pronoun sub-experiment. The dots stand for the means, and the error bars show the standard error of the means. In the plot showing the fixations to the target picture at the critical region, the proportions are centered and scaled within groups

Fig. 12 Descriptive statistics for the relative clauses sub-experiment. The dots stand for the means, and the error bars show the standard error of the means. In the plot showing the fixations to the target picture at the critical region, the proportions are centered and scaled within groups
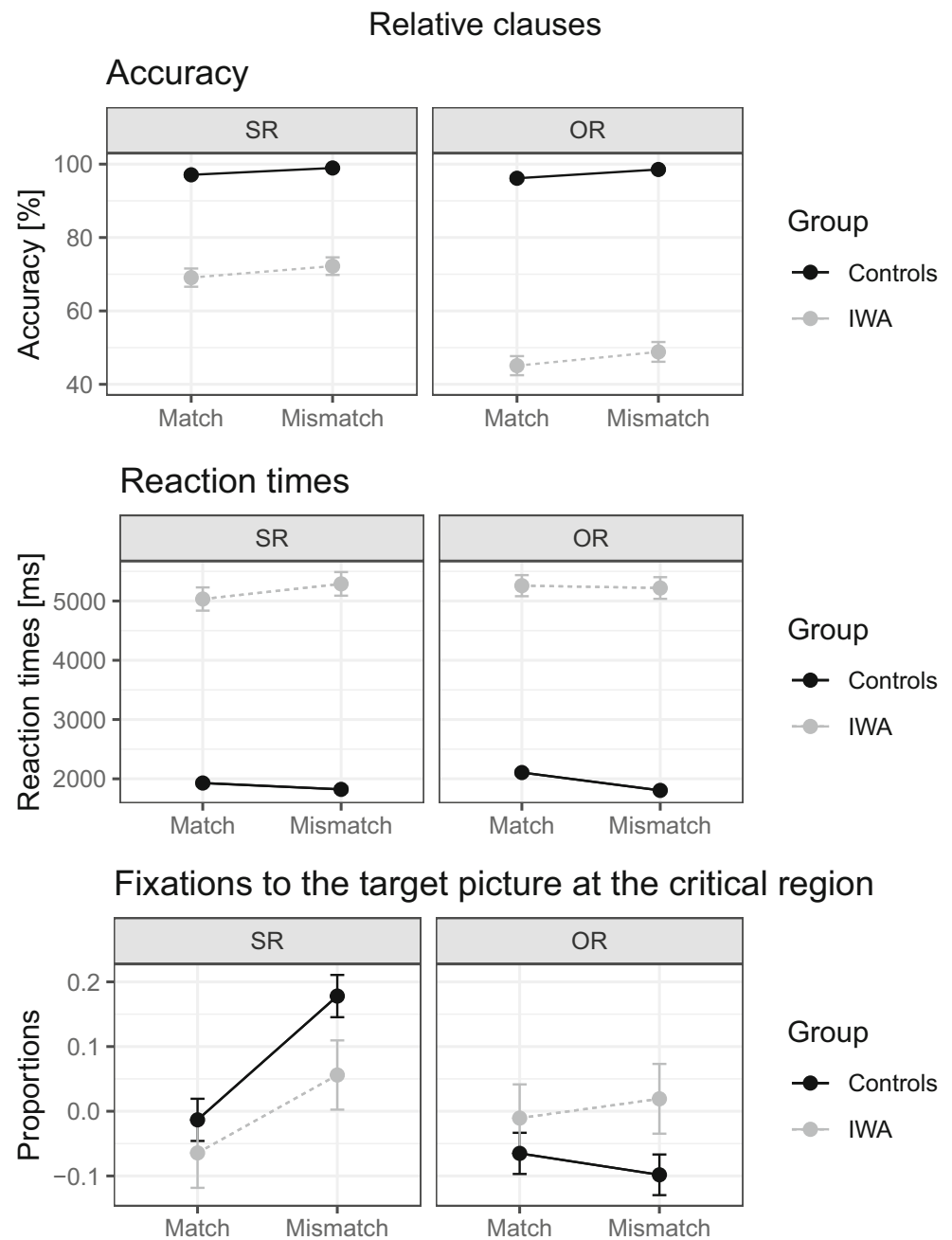
Appendix C: Details about the computational implementations

For both the activation-based and the direct-access model, the varying intercepts and slopes for subject, **u**, come from a multivariate normal distribution with x dimensions (where x = number of by-subject adjustments), abbreviated as MVN_x. The varying intercepts and slopes for items, **w**, also come from a multivariate normal distribution with y dimensions, MVN_y (where y = number of by-item adjustments). In the equations below, **0** is a column vector of zeros with the x or y dimensions. The σ are the variance-covariance matrices of the multivariate normal distributions.

$$\mathbf{u} = MVN_x(\mathbf{0}, \Sigma_u) \quad \mathbf{w} = MVN_y(\mathbf{0}, \Sigma_w) \tag{1}$$

Activation-Based Model For each trial i, values from two independent distributions are sampled, as shown in Eq. (2). The lowest value becomes the estimated reaction time for trial i (see Eq. (3)).

Accumulator NP1

$$FT_{\text{NP1}_i} \sim \text{lognormal}(\mu_1, \sigma)$$

Accumulator NP2

$$FT_{\text{NP2}_i} \sim \text{lognormal}(\mu_2, \sigma) \tag{2}$$

$$RT_i = \min(FT_{\text{NP1}_i}, FT_{\text{NP2}_i}) \tag{3}$$

The hierarchical structure in the location parameter of the lognormal distribution for each accumulator (μ_1 and μ_2) and in the standard deviation (σ) is shown in Eq. (4).

Accumulator NP1
$$\mu_{\text{NP1}} = \alpha_1 + u_{\alpha_1} + w_{\alpha_1} + (\beta_1 + w_{\beta_1}) \cdot \text{group} + (\beta_3 + u_{\beta_3}) \cdot \text{condition} + \beta_5 \cdot \text{group} \cdot \text{condition} + \beta_7 \cdot LDT + \beta_9 \cdot \text{group} \cdot LDT + \beta_{11} \cdot \text{fixations} + \beta_{13} \cdot \text{fixations} \cdot \text{group}$$

Accumulator NP2
$$\mu_{\text{NP2}} = \alpha_2 + u_{\alpha_2} + w_{\alpha_2} + (\beta_2 + w_{\beta_2}) \cdot \text{group} + (\beta_4 + u_{\beta_4}) \cdot \text{condition} + \beta_6 \cdot \text{group} \cdot \text{condition} + \beta_8 \cdot LDT + \beta_{10} \cdot \text{group} \cdot LDT + \beta_{12} \cdot \text{fixations} + \beta_{14} \cdot \text{fixations} \cdot \text{group}$$

Noise parameter
$$\sigma = \sigma_0 + \beta_{15} \cdot \text{group} \tag{4}$$

In addition to the adjustments shown in Eq. (4), in the model for relative clauses, the following fixed effects were also added to both accumulators: RC type, RC type × condition, RC type × group, RC type group × group ×

condition, fixations × RC type, fixations × group × RC type. The slope for the effect of RC type also had a by-subject adjustment. Equation (5) shows the priors used in the activation-based model, all in log space.

$$\alpha_{1,2} \sim \text{normal}(7.5, 0.6)$$
$$\beta_{1,\dots,15} \sim \text{normal}(0, 0.5) \tag{5}$$
$$\sigma_0 \sim \text{normal}_+(0, 0.5)$$

Direct-Access Model In Eq. (6), the parameter θ stands for the probability of retrieval of the target; μ is the location of the distribution from which the reaction times are sampled. P_b is the probability of backtracking, δ is the cost of backtracking (in log ms), and σ is the noise parameter. In addition to the adjustments shown in (6), in the model for relative clauses, the following fixed effects were added to θ: RC type, RC type × condition, RC type × group, and RC type × group × condition. The slope for the effect of RC type also had a by-subject adjustment.

$$\mu = \mu_0 + u_{\mu 0} + w_{\mu 0} + \beta_1 \cdot \text{group}$$
$$\theta = \alpha + u_\alpha + w_\alpha + \beta_2 \cdot LDT + \beta_3 \cdot LDT \cdot \text{group}$$
$$(\beta_4 + w_{\beta_4}) \cdot \text{group} + (\beta_5 + u_{\beta_5}) \cdot \text{condition} +$$
$$\beta_6 \cdot \text{group} \cdot \text{condition} + \beta_7 \cdot \text{fixations} +$$
$$\beta_8 \cdot \text{group} \cdot \text{fixations} \tag{6}$$
$$\theta_b = \alpha_b + u_{\alpha_b} + \beta_9 \cdot \text{group}$$
$$P_b = \gamma + u_\gamma + \beta_{10} \cdot \text{group}$$
$$\delta = \delta_0 + \beta_{11} \cdot \text{group}$$
$$\sigma = \sigma_0 + \beta_{12} \cdot \text{group}$$

Equation (7) shows the priors used in the direct-access model. The priors for α, α_b, and γ are in logit space, and the rest of priors are in log space.

$$\alpha \sim \text{normal}(1, 0.5)$$
$$\alpha_b \sim \text{normal}(0, 1)$$
$$\beta_{1,\dots,12} \sim \text{normal}(0, 0.5)$$
$$\mu_0 \sim \text{normal}(7.5, 0.6) \tag{7}$$
$$\gamma \sim \text{normal}(-1, 0.5)$$
$$\delta_0 \sim \text{normal}(0, 1)$$
$$\sigma_0 \sim \text{normal}(0, 0.5)$$

Appendix D: Model estimates (activation-based model)

Pronouns

Table 3 shows model estimates for the activation-based model for the pronoun sub-experiment.

Table 3 Model estimates for the fixed effects on μ_{NP1} and μ_{NP2} and corresponding credible intervals, backtransformed to ms

Parameter	Estimate	95% CrI	Accumulator (μ)
Group	2006 ms	CrI: [1604, 2453] ms	NP1
Group	− 797 ms	CrI: [− 2459, 629] ms	NP2
Condition	310 ms	CrI: [193, 433] ms	NP1
Condition	− 1478 ms	CrI: [− 2820, -431] ms	NP2
Condition × group	199 ms	CrI: [81, 322] ms	NP1
Condition × group	1424 ms	CrI: [400, 2720] ms	NP2
LDT	836 ms	[539, 1152] ms	NP1
LDT	588 ms	[− 621, 1811] ms	NP2
LDT × group	671 ms	[377, 973] ms	NP1
LDT × group	51 ms	[− 1147, 1243] ms	NP2
Fixations	41 ms	[− 77, 160] ms	NP1
Fixations	518 ms	[− 141, 1212] ms	NP2
Fixations × group	96 ms	[− 23, 219] ms	NP1
Fixations × group	251 ms	[− 430, 949] ms	NP2

Relative Clauses

Table 4 shows model estimates for the activation-based model for the relative clauses sub-experiment.

Table 4 Model estimates for the fixed effects on μ_{NP2} and μ_{NP2} and corresponding credible intervals, backtransformed to ms

Parameter	Estimate	95% CrI	Accumulator (μ)
Group	2348 ms	[1687, 3055] ms	NP1
Group	4057 ms	[3279, 4916] ms	NP2
RC type	2417 ms	[2076, 2807] ms	NP1
RC type	− 1897 ms	[− 2372, -1459] ms	NP2
Condition	− 68 ms	[− 262, 122] ms	NP1
Condition	− 228 ms	[− 506, 40] ms	NP2
Condition × group	− 37 ms	[− 224, 153] ms	NP1
Condition × group	86 ms	[− 180, 355] ms	NP2
RC type × group	− 1670 ms	[− 2019, -1366] ms	NP1
RC type × group	1847 ms	[1443, 2293] ms	NP2
Condition × RC type	− 150 ms	[− 323, 19] ms	NP1
Condition × RC type	631 ms	[385, 884] ms	NP2
Condition × RC type × group	103 ms	[− 68, 276] ms	NP1
Condition × RC type × group	− 249 ms	[− 491, -10] ms	NP2
LDT	779 ms	[243, 1337] ms	NP1
LDT	812 ms	[252, 1398] ms	NP2
LDT × group	− 50 ms	[− 594, 496] ms	NP1
LDT × group	− 460 ms	[− 1045, 121] ms	NP2
LDT × RC type	− 5 ms	[− 258, 255] ms	NP1
LDT × RC type	19 ms	[− 347, 385] ms	NP2
LDT × RC type × group	5 ms	[− 251, 262] ms	NP1
LDT × RC type × group	291 ms	[− 72, 677] ms	NP2
Fixations	− 37 ms	[− 205, 130] ms	NP1
Fixations	289 ms	[61, 521] ms	NP2
Fixations × group	− 46 ms	[− 211, 118] ms	NP1
Fixations × group	57 ms	[− 173, 285] ms	NP2
Fixations × RC type	69 ms	[− 97, 235] ms	NP1
Fixations × RC type	− 27 ms	[− 248, 204] ms	NP2
Fixations × RC type × group	75 ms	[− 96, 244] ms	NP1
Fixations × RC type × group	− 111 ms	[− 341, 112] ms	NP2

Appendix E: Cross-validation results by group and construction

Figure 13 shows cross-validation results by group and construction.
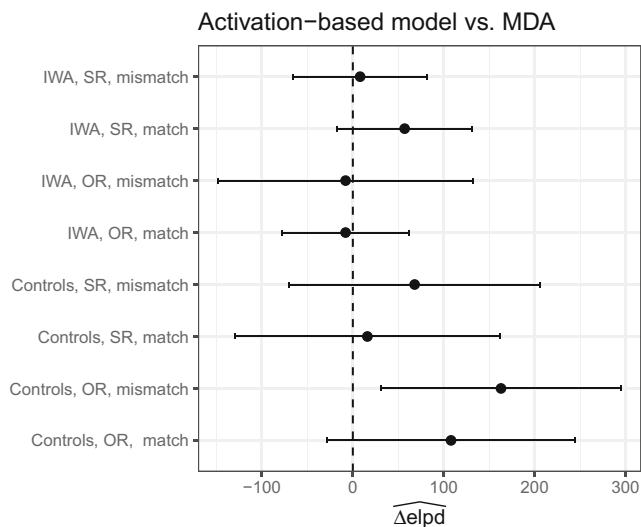

Fig. 13 Graphical representation of the $\widehat{\Delta elpd}$ between the activation-based and the modified direct-access model across groups and conditions. The dot stands for the $\widehat{\Delta elpd}$ and the bars indicate to the 95% confidence interval. Positive values indicate an advantage for the activation-based model, and negative values indicate an advantage for the modified direct-access model (MDA)

Appendix F: Comparisons with original direct-access model

The $\widehat{\Delta elpd}$ between the original direct-access model and the modified direct-access model is -41, SE 134 for pronoun resolution; and -175, SE 166 for relative clauses. The negative $\widehat{\Delta elpd}$ estimates indicate that the modified direct-access model may have a better predictive performance, but given the large SE, the $\widehat{\Delta elpd}$ are inconclusive. When comparing the activation-based model with the original direct-access model, the difference is not conclusive in the pronoun experiment ($\widehat{\Delta elpd}$ -68, SE 140), but in relative clauses, the activation-based model performs better ($\widehat{\Delta elpd}$ 578, SE 173). Thus, in relative clauses, the activation-based model outperforms both the original and the modified-direct access model.

Author Contribution The first draft of the manuscript was written by Paula Lissón and all authors commented on previous versions of the manuscript. Dario Paape carried out revisions after the first round of review. All authors read and approved the final manuscript. Data analysis and modeling were carried out by Paula Lissón and supervised by Shravan Vasishth and Dario Paape. Nicole Stadie, Frank Burchert, and Dorothea Pregla contributed to the study conception and design. Material preparation and data collection were performed by Dorothea Pregla.

Code Availability The data generated and/or analyzed during the current study are available in the OSF repository, https://tinyurl.com/yz9acste.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

References

Adelt, A., Burchert, F., Adani, F., & Stadie, N. (2020). What matters in processing German object relative clauses in aphasia–timing or morpho-syntactic cues? *Aphasiology*, *34*(8), 970–998.

Adelt, A., Stadie, N., Lassotta, R., Adani, F., & Burchert, F. (2017). Feature dissimilarities in the processing of German relative clauses in aphasia. *Journal of Neurolinguistics*, *44*, 17–37.

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.

Arantzeta, M., Webster, J., Laka, I., Martínez-Zabaleta, M., & Howard, D. (2018). What happens when they think they are right? Error awareness analysis of sentence comprehension deficits in aphasia. *Aphasiology*, *32*(12), 1418–1444.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748.

Bader, M., & Meng, M. (1999). Case attraction phenomena in German. *Unpublished Manuscript. University of Jena*.

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178–1198.

Burchert, F., & de Bleser, R. (2004). Passives in agrammatic sentence comprehension: A German study. *Aphasiology*, *18*(1), 29–45.

Burchert, F., de Bleser, R., & Sonntag, K. (2003). Does morphology make the difference? Agrammatic sentence comprehension in German. *Brain and Language*, *87*(2), 323–342.

Burchert, F., Hanne, S., & Vasishth, S. (2013). Sentence comprehension disorders in aphasia: The concept of chance performance revisited. *Aphasiology*, 27(1), 112–125.

Burkhardt, P., Avrutin, S., Piango, M. M., & Ruigendijk, E. (2008). Slower-than-normal syntactic processing in agrammatic Broca's aphasia: Evidence from Dutch. *Journal of Neurolinguistics*, 21(2), 120–137.

Burkhardt, P., Piñango, M. M., & Wong, K. (2003). The role of the anterior left hemisphere in real-time sentence comprehension: Evidence from split intransitivity. *Brain and Language*, 86(1), 9–22.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28.

Caplan, D. (2012). Resource reduction accounts of syntactically based comprehension disorders. In C. K. Thompson, & R. Bastiannse (Eds.) *Perspectives on Agrammatism, pp. 34–48*.

Caplan, D., Michaud, J., & Hufford, R. (2013). Dissociations and associations of performance in syntactic comprehension in aphasia and their implications for the nature of aphasic deficits. *Brain and Language*, 127(1), 21–33.

Caplan, D., Michaud, J., & Hufford, R. (2015). Mechanisms underlying syntactic comprehension deficits in vascular aphasia: New evidence from self-paced listening. *Cognitive Neuropsychology*, 32(5), 283–313.

Caplan, D., Waters, G., DeDe, G., Michaud, J., & Reddy, A. (2007). A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*, 101(2), 103–150.

Caramazza, A., & Zurif, E. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4), 572–582.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M. ,..., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.

Chomsky, N. (1977). On Wh-Movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.) *Formal syntax*. New York: Academic Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5, 630.

Choy, J. J., & Thompson, C. K. (2010). Binding in agrammatic aphasia: Processing to comprehension. *Aphasiology*, 24(5), 551–579.

Comrie, B. (1985). Reflections on subject and object control. *Journal of Semantics*, 4(1), 47–65.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, 75, 117–139.

Czypionka, A., Dörre, L., & Bayer, J. (2018). Inverse Case attraction: Experimental evidence for a syntactically guided process. *The Journal of Comparative Germanic Linguistics*, 21(2), 135–188.

Dickey, M. W., Choy, J. J., & Thompson, C. K. (2007). Real-time comprehension of wh-movement in aphasia: Evidence from eyetracking while listening. *Brain and Language*, 100(1), 1–22.

Dickey, M. W., & Thompson, C. K. (2009). Automatic processing of wh- and NP-movement in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics*, 22(6), 563–583.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69(2), 85–103.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, 36(2), 147–164.

Edwards, S., & Varlokosta, S. (2007). Pronominal and anaphoric reference in agrammatism. *Journal of Neurolinguistics*, 20(6), 423–444.

Engel, S., Shapiro, L. P., & Love, T. (2018). Proform-antecedent linking in individuals with agrammatic aphasia: A test of the Intervener Hypothesis. *Journal of Neurolinguistics*, 45, 79–94.

Engelmann, F. (2016). *Toward an integrated model of sentence processing in reading*. (Doctoral dissertation, University of Potsdam).

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association in retrieval processes: A computational account. *Cognitive Science*, 43(12), e12800.

Fanselow, G., Schlesewsky, M., Cavar, D., & Kliegl, R. (1999). Optimal parsing: Syntactic parsing preferences and optimality theory, vol. 367. Rutgers State University of New Jersey.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54(4), 541–553.

Ferrill, M., Love, T., Walenski, M., & Shapiro, L. P. (2012). The time-course of lexical activation during sentence comprehension in people with aphasia. *American Journal of Speech-Language Pathology*, 21(2), S179.

Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, 47(3), 193–221.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society*, 182, 389–402.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

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.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, 29, 261–290.

Hanne, S., Burchert, F., de Bleser, R., & Vasishth, S. (2015). Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction. *Journal of Neurolinguistics*, 34, 83–111.

Hanne, S., Sekerina, I. A., Vasishth, S., Burchert, F., & de Bleser, R. (2011). Chance in agrammatic sentence comprehension: What does it really mean? Evidence from eye movements of German agrammatic aphasic patients. *Aphasiology*, 25(2), 221–244.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, 6, 617.

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.

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, 104063.

Laurinavichyute, A., Jäger, L. A., Akinina, Y., Roß, J., & Dragoy, O. (2017). Retrieval and encoding interference: Cross-linguistic evidence from anaphor processing. *Frontiers in Psychology*, 8, 965.

Lewis, R. L. (1999). Specifying architectures for language processing: Process, control, and memory in parsing and interpretation. In M. W. Crocker, M. Pickering, & C. Clifton (Eds.) *Architectures and mechanisms for language processing, pp. 56–89*.

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

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

Lissón, P., Pregla, D., Nicenboim, B., Paape, D., Van Het Nederend, M. L., Burchert, F. ,..., & Vasishth, S. (2021a). A computational evaluation of two models of retrieval processes in sentence processing in aphasia. *Cognitive Science*, *45*(4), e12956.

Lissón, P., Pregla, D., Paape, D., Burchert, F., Stadie, N., & Vasishth, S. (2021b). Modeling sentence comprehension deficits in aphasia: A computational evaluation of the direct-access model of retrieval. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, NAACL, pp. 177–185*.

Logačev, P., & Vasishth, S. (2012). Case matching and conflicting bindings interference. In M. Lamers, & P. de Swart (Eds.) *Case, word order and prominence, vol. 40, pp. 187–216*.

Logačev, P., & Vasishth, S. (2016). Understanding underspecification: a comparison of two computational implementations. *Quarterly Journal of Experimental Psychology*, *69*(5), 996–1012.

Love, T., Swinney, D., Walenski, M., & Zurif, E. (2008). How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. *Brain and Language*, *107*(3), 203–219.

Mack, J. E., Ji, W., & Thompson, C. K. (2013). Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics*, *26*(6), 619–636.

Mack, J. E., Wei, A. Z.-S., Gutierrez, S., & Thompson, C. K. (2016). Tracking sentence comprehension: Test-retest reliability in people with aphasia and unimpaired adults. *Journal of Neurolinguistics*, *40*, 98–111.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, *58*(3), 879–906.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language*, *64*(4), 327–343.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage*, *59*(2), 1859–1869.

Mätzig, P., Vasishth, S., Engelmann, F., Caplan, D., & Burchert, F. (2018). A computational investigation of sources of variability in sentence comprehension difficulty in aphasia. *Topics in Cognitive Science*, *10*(1), 161–174.

McAllister, T., Bachrach, A., Waters, G., Michaud, J., & Caplan, D. (2009). Production and comprehension of unaccusatives in aphasia. *Aphasiology*, *23*, 989–1004.

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, *32*(4), 536–571.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111–123.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, *46*, 155–200.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

Mirman, D. (2017). Growth curve analysis and visualization using R. Boca Raton, FL: Chapman & Hall/CRC.

Müller, S. (2002). *Complex predicates: verbal complexes, resultative constructions, and particle verbs in German*. Stanford: center for the study of language and information.

Nicenboim, B., Schad, D. J., & Vasishth, S. (2021). *Introduction to Bayesian data analysis for cognitive science*. Under contract with Chapman and Hall/CRC statistics in the social and behavioral sciences series.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas - Part II. *Language and Linguistics Compass*, *10*(11), 591–613.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, *42*, 1075–1100.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569–587.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, *53*(3), 238–283.

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272–290.

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). 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, pp. 121–144. Newcastle: Cambridge Scholars Publishing*.

Patil, U., Hanne, S., Burchert, F., de Bleser, R., & Vasishth, S. (2016). A computational evaluation of sentence processing deficits in aphasia. *Cognitive Science*, *40*(1), 5–50.

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*, 329.

Pregla, D., Lissón, P., Vasishth, S., Burchert, F., & Stadie, N. (2021). Variability in sentence comprehension in aphasia in German. *Brain and Language*, *222*, 105008.

Pregla, D., Vasishth, S., Lissón, P., Stadie, N., & Burchert, F. (2022). A visual world study of sentence processing in aphasia in German: The resource reduction hypothesis revisited. PsyArXiv preprint: psyarxiv.com/c456b.

R Core Team, R. (2020). R: A language and environment for statistical computing (Version 4.0.2). Vienna, Austria.

Runner, J. T., & Head, K. D. (2014). What can visual world eye-tracking tell us about the binding theory? *Empirical Issues in Syntax and Semantics*, *10*, 269–286.

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103–126.

Schlesewsky, M. (1996). *Kasusphänomene in der Sprachverarbeitung* (Doctoral dissertation, University of Potsdam).

Schumacher, R., Cazzoli, D., Eggenberger, N., Preisig, B., Nef, T., Nyffeler, T. ,..., & Müri, R. M. (2015). Cue recognition and integration–eye tracking evidence of processing differences in sentence comprehension in aphasia. *Plos One*, *10*(11), e0142853.

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, *12*(3), 175–200.

Stadie, N., Cholewa, J., & de Bleser, R. (2013). *LEMO 2.0: Lexikon modellorientiert: Diagnostik für Aphasie, Dyslexie und Dysgraphie*. Hofheim: NAT.

Stan Development Team, S. (2020). RStan: The R interface to Stan.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Staub, A., Dillon, B., & Clifton Jr, C. (2017). The matrix verb as a source of comprehension difficulty in object relative sentences. *Cognitive Science*, *41*, 1353–1376.

Stiebels, B., McFadden, T., Schwabe, K., Solstad, T., Kellner, E., Sommer, L., & Stoltmann, K. (2018). ZAS database of clause-embedding predicates, release 1.0. In *OWID Plus*.

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

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.

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 misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263.

Van Maanen, L., Katsimpokis, D., & Van Campen, A. D. (2019). Fast and slow errors: Logistic regression to identify patterns in accuracy–response time relationships. *Behavior Research Methods*, *51*(5), 2378–2389.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Vasishth, S., & Engelmann, F. (2022). *Sentence comprehension as a cognitive process: A computational approach*. Cambridge, UK: Cambridge University Press.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, *82*(4), 767–794.

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*, 968–982.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Yadav, H., Paape, D., Smith, G., Dillon, B., & Vasishth, S. (2021). Individual differences in cue-weighting in sentence comprehension: An evaluation using approximate Bayesian computation.

Yadav, H., Smith, G., & Vasishth, S. (2021). Is similarity-based interference caused by lossy compression or cue-based retrieval? A computational evaluation. In *Proceedings of the International Conference on Cognitive Modeling*.