Computational models of sentence comprehension in aphasia

Paula J. Lissón Hernández

2021

Doctoral Thesis submitted to the Faculty of Human Sciences at the University of Potsdam in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Linguistics
University of Potsdam

Supervisors:
Prof. Dr. Shravan Vasishth
Dr. Nicole Stadie
Dr. Dario Paape

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-55548
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-555487

Abstract

It is well-known that individuals with aphasia (IWA) have difficulties understanding sentences that involve non-adjacent dependencies, such as object relative clauses or passives (Caplan, Baker, & Dehaut, 1985; Caramazza & Zurif, 1976). A large body of research supports the view that IWA's grammatical system is intact, and that comprehension difficulties in aphasia are caused by a processing deficit, such as a delay in lexical access and/or in syntactic structure building (e.g., Burkhardt, Piñango, & Wong, 2003; Caplan, Michaud, & Hufford, 2015; Caplan, Waters, DeDe, Michaud, & Reddy, 2007; Ferrill, Love, Walenski, & Shapiro, 2012; Hanne, Burchert, De Bleser, & Vasishth, 2015; Love, Swinney, Walenski, & Zurif, 2008). The main goal of this dissertation is to computationally investigate the processing sources of comprehension impairments in sentence processing in aphasia.

In this work, prominent theories of processing deficits coming from the aphasia literature are implemented within two cognitive models of sentence processing –the activation-based model (Lewis & Vasishth, 2005) and the direct-access model (McElree, 2000)–. These models are two different expressions of the cue-based retrieval theory (Lewis, Vasishth, & Van Dyke, 2006), which posits that sentence processing is the result of a series of iterative retrievals from memory. These two models have been widely used to account for sentence processing in unimpaired populations in multiple languages and linguistic constructions, sometimes interchangeably (Parker, Shvartsman, & Van Dyke, 2017). However, Nicenboim and Vasishth (2018) showed that when both models are implemented in the same framework and fitted to the same data, the models yield different results, because the models assume different data-generating processes. Specifically, the models hold different assumptions regarding the retrieval latencies. The second goal of this dissertation is to compare these two models of cue-based retrieval, using data from individuals with aphasia and control participants. We seek to answer the following question: Which retrieval mechanism is more likely to mediate sentence comprehension?

We model 4 subsets of existing data: Relative clauses in English and German; and control structures and pronoun resolution in German. The online data come from

either self-paced listening experiments, or visual-world eye-tracking experiments. The offline data come from a complementary sentence-picture matching task performed at the end of the trial in both types of experiments. The two competing models of retrieval are implemented in the Bayesian framework, following Nicenboim and Vasishth (2018). In addition, we present a modified version of the direct-acess model that – we argue – is more suitable for individuals with aphasia.

This dissertation presents a systematic approach to implement and test verbally-stated theories of comprehension deficits in aphasia within cognitive models of sentence processing. The conclusions drawn from this work are that (a) the original direct-access model (as implemented here) cannot account for the full pattern of data from individuals with aphasia because it cannot account for slow misinterpretations; and (b) an activation-based model of retrieval can account for sentence comprehension deficits in individuals with aphasia by assuming a delay in syntactic structure building, and noise in the processing system. The overall pattern of results support an activation-based mechanism of memory retrieval, in which a combination of processing deficits, namely slow syntax and intermittent deficiencies, cause comprehension difficulties in individuals with aphasia.

Acknowledgments

First and foremost, I would like to thank Shravan Vasishth for being an excellent researcher and mentor. I am grateful for his detailed feedback, his guidance, and above all, for showing me how to develop the right mindset to do science. I also thank him for giving me the time and patience I needed to properly prepare and (re)learn things at my own pace.

I am grateful to Dario Paape and Bruno Nicenboim, who have also been great mentors. I thank them for their time and effort. I am also grateful to Nicole Stadie and Frank Burchert, for their feedback and support.

I would also like to thank the rest of (former and current) lab members of the Vasishth Lab: Sol Lago, Anna Laurinavichyute, Daniela Mertzen, Kate Stone, Pia Schoknecht, Daniel Schad, Garrett Smith, João Veríssimo, Titus von der Malsburg, and Himanshu Yadav, for the countless hours of interesting discussions and ideas. I am greatly indebted to Dorothea Pregla, as without her, this project would have not been possible; and to Sandra Hanne, for teaching me everything I know about aphasia.

I am extremely grateful to the DFG and the SFB 1287 project, for granting the funding that covered the research in this project as well as the travel costs for several conferences. I am grateful to the SFB 1287 members for organizing activities that have fostered my development as a scientist, and to Michaela Schmitz in particular, for her help throughout these years.

I owe gratitude to Verónica Trujillo-González for encouraging me to become a scientist, and to Nicolas Ballier (DT), who showed me range of opportunities that research has to offer. Finally, I thank Jens-Uwe Junghanns for showing me how to navigate the ups and downs of a PhD, and for standing always by my side. Le doy las gracias a mi hermana porque siempre me ha animado a luchar por lo que quiero, y a mi madre, por ser incondicional.

Contents

Chapter 1

Introduction

Sentence processing is a complex task that relies on several cognitive mechanisms. Individuals with an acquired language disorder, such as individuals with aphasia (IWA), are an interesting case-study for theories of sentence processing. The linguistic performance of IWA offers the possibility to investigate how neurological damage affects the cognitive resources used in language comprehension. IWA exhibit comprehension difficulties (e.g., Caplan et al., 2015; Caplan et al., 2007; Caramazza & Zurif, 1976), which are usually more pronounced in sentences with complex syntactical operations, such as the examples shown in (1). As IWA were once able to process these complex sentences, studying the source of the comprehension deficits may yield important insights with respect to which cognitive resources are involved in sentence processing, and how the language system is organized. This dissertation aims to computationally investigate which cognitive mechanisms are the source of the sentence comprehension difficulties that individuals with aphasia experience.

(1) a. **Object relative:** The boy who the man scratched pushed the girl.
 b. **Reflexive pronoun:** The father of the boy scratched himself.
 c. **Passive:** The man was scratched by the boy.

(Caplan, DeDe, & Michaud, 2006)

Processing a sentence entails resolving relations between words in real time. As the linguistic input unfolds, listeners have to connect incoming words together and build up syntactic and semantic dependencies in real time. This is known as dependency resolution in the sentence processing literature. For instance, verbs have to be linked with their dependent arguments (e.g., in 1a, *the man* is the subject of *scratched*, and *the boy* is the object), and pronouns have to be linked to their antecedents (e.g. *himself* →*the father* in example 1b).

Dependency resolution is commonly assumed to require storing and retrieving information from working memory (Gibson, 2000; Just & Carpenter, 1992; Lewis, 1999; Lewis et al., 2006; McElree, 2006; Van Dyke & Lewis, 2003). A well-known theory of sentence processing in unimpaired populations, cue-based retrieval (Engelmann, Jäger, & Vasishth, 2019; Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, 2006; McElree, Foraker, & Dyer, 2003; Van Dyke, 2007; Van Dyke & McElree, 2006, 2011; Vasishth, Nicenboim, Engelmann, & Burchert, 2019), posits that sentence representations are incrementally built via a series of iterative retrievals from memory. According to this theory, processed words and phrases – often called chunks – are stored in memory with their morphosyntactic and semantic features. Chunks are stored in memory as a bundle of feature-value pairs. For example, in sentence (2), the first noun phrase (NP1) *the boy* is represented in memory as an attribute-value matrix (Pollard & Sag, 1994) with the feature-value pairs shown below.

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

$$\text{NP1: The boy} \quad \begin{pmatrix} \text{nominal - yes} \\ \text{animate - yes} \\ \text{subject - yes} \\ \text{singular - yes} \end{pmatrix}$$

Chunks are retrieved from memory on the basis of their syntactic and semantic features. The features used to search and retrieve a co-dependent in memory are called retrieval cues. For instance, in a subject relative clause such as (2), the verb *greeted* triggers the retrieval of an item in memory whose features match the retrieval cues [+animate] and [+nominative]. When *greeted* is read, the only item in memory that matches these retrieval cues is *boy*. Consider now the sentence below, which is an object relative clause.

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

In (3), when the comprehender reaches the verb *greeted*, there is one item in memory that matches all the retrieval cues set by the verb, *boy*, but there is also another item that matches some of the retrieval cues, *girl* ([+animate, -nominative]). Following Jäger, Engelmann, and Vasishth (2017), the fully-matching item (*boy*) is referred to as the retrieval target, and items with partial feature match (*girl*) are distractors. Similarly, in this case, the verb *greeted* is the retrieval site, i.e., the point at which the retrieval of the co-dependent is triggered.

Memory retrieval is assumed to be affected by time-based decay (Van Dyke & Lewis, 2003). Processing difficulty is predicted to arise with increasing distance between the two co-dependents, because as time goes by, chunks become less available for retrieval. Another key assumption is that memory retrieval is subject to interference (Lewis, 1996; Lewis et al., 2006). Processing difficulty is predicted to arise if multiple chunks in memory match the same retrieval cues set at the retrieval, because these chunks become difficult to distinguish from each other. This effect is known as the fan effect in memory research (Anderson et al., 2004). For instance, the verb *greeted* should be more difficult to process in (3) than in (2), due to the distractor (*boy*) in (2). In cue-based retrieval, this effect is called similarity-based interference (Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006), and is indexed by a slow-down at the retrieval site (the verb *greeted* in 3 and 2) and/or by the occasional misretrieval of a distractor item (*boy* in 3), which results in misinterpretation. Similarity-based interference has been attested in multiple linguistic constructions across different languages (e.g., Dillon, Mishler, Sloggett, & Phillips, 2013; Engelmann et al., 2019; Gordon, Hendrick, Johnson, & Lee, 2006; Jäger et al., 2017; Jäger, Mertzen, Van Dyke, & Vasishth, 2020; A. E. Martin, Nieuwland, & Carreiras, 2012; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Vasishth & Engelmann, 2021; Vasishth et al., 2019).

One important question that arises is, can the cue-based retrieval theory account for the sentence comprehension deficits in IWA? A large body of research shows that IWA exhibit sentence comprehension difficulties in sentences that involve long-distance dependencies, such as wh-questions, object relatives, passives, object-cleft sentences, and pronoun-antecedent dependencies (e.g., Burkhardt, Avrutin, Piñango, & Ruigendijk, 2008; Caplan et al., 2006; Caplan et al., 2015; Caramazza & Zurif, 1976; Choy & Thompson, 2010; Dickey, Choy, & Thompson, 2007; Ferrill et al., 2012; Luzzatti et al., 2001; Meyer, Mack, & Thompson, 2012; Piñango & Burkhardt, 2005; Thompson, Choy, Holland, & Cole, 2010). In the aphasia literature, several theories have been proposed that aim to explain the source of these comprehension difficulties. For instance, several theories suggest that the source of IWA's impairments is a limitation in processing capacities, such as a lower working memory capacity, a delay in lexical retrieval, or a delay in syntactic structure building (Burkhardt et al., 2008; Caplan et al., 2006; Caplan et al., 2015; Caramazza & Zurif, 1976; Piñango & Burkhardt, 2005). These theories can be straightforwardly integrated in the framework of cue-based retrieval, because cue-based retrieval theory puts memory and other elements from the general cognitive domain at the center of the sentence comprehension process (Lewis, 1999). In addition, IWA's comprehension difficulties arise in long-distance dependencies, which are prone to memory decay and similarity-based interference in the retrieval process. Therefore, cue-based retrieval theory could explain IWA's comprehension difficulties by assuming that IWA's processing deficits disrupt the retrieval process.

One advantage of expanding the cue-based retrieval theory to account for sentence processing in IWA is that existing computational models of retrieval processes can be used to implement the different processing-based theories of impairments in aphasia (Mätzig, Vasishth, Engelmann, Caplan, & Burchert, 2018; Patil, Hanne, Burchert, De Bleser, & Vasishth, 2016). The alternative theories can be evaluated against experimental data from IWA, and control participants, whose performance is taken as a baseline. Thus, cue-based retrieval provides a unified architecture for modeling both unimpaired and impaired sentence comprehension.

Within the cue-based retrieval framework, two different models of retrieval processes have been proposed –the activation-based model (Lewis & Vasishth, 2005, henceforth LV05), and the direct-access model (McElree, 2000)–. The two models posit that retrieval cues mediate access to items in memory. However, they make different assumptions regarding the retrieval process. The activation-based model holds that when retrieving a co-dependent from memory, both the retrieval probability and the retrieval latency depend on sentence complexity. By contrast, the direct-access model holds that retrieval latencies are constant and independent from sentence complexity.

An open question in psycholinguistics concerns how linguistic representations in memory are stored, searched, and retrieved during sentence comprehension (Parker et al., 2017). Given that the activation-based and the direct-access model assume two different underlying mechanisms for retrieval latencies, yet both models can explain sentence processing in unimpaired populations (e.g., Jäger et al., 2017; A. E. Martin & McElree, 2008; McElree, 2006; Van Dyke & Johns, 2012; Van Dyke & McElree, 2011; Vasishth & Engelmann, 2021; Vasishth et al., 2019), one key question that arises is whether one of the two models is better able to model linguistic data from IWA. Investigating which model of retrieval processes is more likely to mediate sentence comprehension in IWA could help understanding how linguistic representations are stored and accessed in IWA, and by extension, in unimpaired populations.

1.1 Aims and outline

This thesis has two main goals. The first goal is to computationally investigate the role of different processing deficits in IWA (e.g., delays in lexical access and in syntactic structure building) under the framework of the cue-based retrieval of sentence processing. The second goal is to compare two competing models of retrieval that assume different underlying latent retrieval processes, using data from IWA and control participants. Two aspects in this dissertation are novel: First, this is the first work in which competing models of retrieval are evaluated using data from a range of syntactic constructions in IWA and control participants in English and German. Second, a systematic approach to computationally investigate deficits in sentence processing in IWA is developed.

Figure 1.1 shows a schematic representation of the models and theories that are investigated in this dissertation, and how these are interconnected. The theories of processing deficits in aphasia are mapped onto the models' parameters. The estimates for control participants are taken as a baseline, and the estimates for IWA are used to test the theories. Thus, we evaluate whether the two models of cue-based retrieval provide a good fit to experimental data from IWA by assuming a series processing-based impairments that have been proposed in the aphasia literature. Finally, the two models of retrieval processes are quantitatively compared against each other, in order to assess which model provides a better fit.

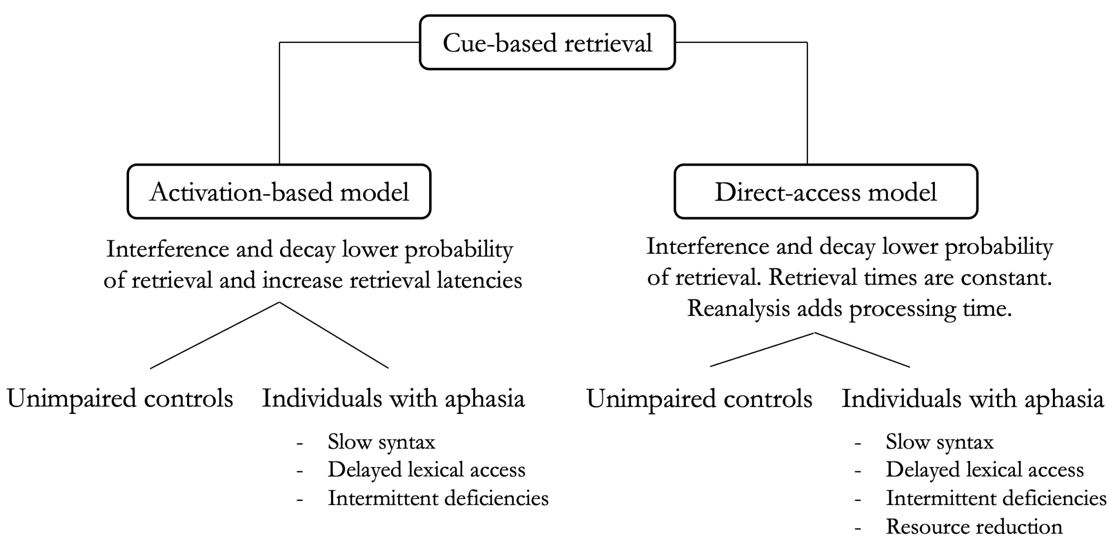

Figure 1.1: Graphical representation of the theories and models tested in this dissertation.

The structure of the thesis is as follows: Chapter 2 presents a short overview on aphasia and details the theories of processing deficits in aphasia that will be implemented computationally. These include *delayed lexical access* (Ferrill et al., 2012), *slow syntax* (Burkhardt et al., 2003), *resource reduction* (Caplan, 2012), and *intermittent deficiencies* (Caplan et al., 2007). Chapter 3 contains the implementation details of the computational models. Chapter 4, 5 and 6 constitute the empirical part of the dissertation, and are dedicated to the evaluation of processing deficits in aphasia in the framework of two computational models of sentence comprehension: the activation-based model (Lewis & Vasishth, 2005), and the direct access model (McElree, 2000).

In Chapter 4, the following question is addressed: Can sentence comprehension impairments in aphasia be explained by difficulties arising from dependency completion processes in parsing? The activation-based and the direct-access model are fitted to listening times coming from an experiment testing the comprehension of English subject and object relative clauses (Caplan, Michaud, & Hufford, 2013). The predictive performance of the models is evaluated using k-fold cross-validation. This chapter shows that for both IWA and controls, the activation-based model furnishes a slightly better quantitative fit to the data relative to the direct-access model. Yet, the quantitative difference between the predictive performance of these two models is not decisive. Model comparisons using Bayes factors show that, assuming an activation-based model, intermittent deficiencies may be the best explanation for the cause of impairments in IWA, although slowed syntax and lexical delayed access may also play a role.

In Chapter 5, a modified version of the original direct-access model is presented. In the original direct-access model, an initial misretrieval followed by reanalysis leads to the retrieval of the target from memory. In the modified version, reanalysis can lead to the retrieval of the target, or to a misretrieval. Self-paced listening data from IWA and controls participants in German are used to test the comprehension of control structures (Chomsky, 1981; Müller, 2002). The performance of the augmented model is compared to the original model. Model comparisons using Bayes factors reveal that the original and the augmented model provide a comparable fit to the data from IWA and controls. Despite the inconclusive result, we argue that the modified direct-access model is better suited to fit data from impaired populations because only the augmented model can account for slow incorrect responses, which are frequent in the aphasia literature.

In Chapter 6, the modified direct-access model developed in Chapter 5 and the activation-based model are compared. This chapter presents the first large-scale computational evaluation of interference effects in two models of sentence processing in

German in IWA and control participants. The models are tested against two linguistic constructions: Pronoun resolution and relative clauses. The results show that both control participants and IWA are susceptible to retrieval interference, and that a combination of theoretical explanations (intermittent deficiency, and slow syntax) can explain IWA's deficits in sentence processing. Both models have a similar predictive performance in pronoun resolution, but the activation-based model outperforms both the original and the direct-access model in relative clauses.

Finally, Chapter 7 summarizes the main conclusions of this dissertation.

Chapter 2

Sentence processing in aphasia

2.1 Introduction

Aphasia is a neurological acquired disease that causes impairments in language production and comprehension due to brain damage (Harley, 2013, p. 68). Until the second half of the 20th century, aphasia was generally studied clinically by physicians and neurologists (see Eling & Whitaker, 2009, for a review on the history of aphasia). Initially, the study of aphasia focused on studying the anatomophysiological localization of language-related functions in the brain. Broca (1863) was the first to suggest that the frontal gyrus in the left hemisphere was responsible for speech comprehension and production. A few years later, Wernicke (1874) drafted a model that conceived language as a psychological process, in which several components of the brain played a role. As of today, Broca's and Wernicke's aphasia are two of the prototypical classic clinical syndromes. These syndromes were followed by the categorization of conduction aphasia, global aphasia, and others (see Alexander & Hillis, 2008, for a description of the different types of aphasia). Later on, during the second half of the 20th century, the focus in aphasia research started to shift from the brain localization to neurolinguistically-motivated models of language production and comprehension (see Goodglass & Wingfield, 1998). However, there is a great deal of controversy regarding the use of clinical categories of aphasia in cognitive psychology (Caramazza, 1984) because such classifications do not take into account the specific details of the linguistic deficits in IWA (Caplan, 2001). Early studies reporting sentence comprehension deficits in aphasia focused on studying the specific processing difficulties associated with a given type of aphasia (e.g., Caramazza & Zurif, 1976; Goodglass et al., 1979; Marin, Saffran, & Schwartz, 1976). For instance, many studies in the 1980s argue that patients with Broca's aphasia have a loss in syntactic knowledge (for an overview see R. C. Martin, 2006). These studies are rooted in the neurophysiological tradition, in which the anatomical locus of linguistic functions was a major focus. As individuals with Broca's aphasia usually displayed syntactic comprehension difficulties, and the lesion in Broca's aphasics is located in the left inferior frontal cortex, it was inferred that that was the part of the brain responsible for the so-called syntactic processing module (Berndt & Caramazza, 1980). However, later studies (e.g., Berndt, Mitchum, & Haendiges, 1996; Caramazza, Capitani, Rey, & Berndt, 2001) have shown that Broca's aphasia is not related to a concrete pattern in sentence comprehension. Instead, patients with Broca's aphasia show variable comprehension patterns. Although Broca's aphasics usually suffer from agrammatism, there are Broca's aphasics that do not suffer from agrammatism, and conversely, there is agrammatism without Broca's aphasia (De Bleser, Burchert, Holzinger, & Weidlich, 2012).

More generally, some studies have shown that performance in sentence comprehension does not necessarily correspond to the classical categories of aphasia (Caplan et al., 1985; Caplan & Hildebrandt, 1988; Dronkers, Wilkins, Van Valin Jr, Redfern, & Jaeger, 2004). Similarly, Badecker and Caramazza (1986, p.278) claim that clinical categories do not assure group homogeneity in the performance of aphasics "in terms of the intact cognitive system that subserves language processing". In their view, methodological issues arise when the validation of theories of language processing relies in the differences between the clinical categories, i.e., when they are syndrome-based. Due to the diversity in performance across and within different aphasia types, some researchers (e.g., Badecker & Caramazza, 1985; Caplan et al., 2015), propose that claims about syntactic disorders in specific aphasia types should be based on objective measures of performance. That is, in psycholinguistic studies, patients should not be classified according to their type of aphasia, but rather according to their linguistic performance. We follow the approach in Caplan and colleagues (Caplan et al., 2015), and throughout this thesis, we will use the term IWA to refer to individuals with aphasia, irrespective of the aphasia type.

2.1.1 Sentence comprehension deficits in aphasia

An important process in sentence comprehension is to correctly establish who did what to whom. Failure to do this can lead to misinterpretations or to comprehension errors. In terms of generative grammar, this is known as theta-role assignment: The verb determines the so-called thematic roles, i.e., the agent (the doer of the action) and the patient or theme (the recipient of the action). In a simple, canonical English sentence, such as in (4a), the verb assigns the thematic roles to its arguments in a linear order. Canonical sentences are those in which constituents appear in the base canonical order of the language (i.e., subject-verb-object in English). By contrast, in non-canonical sentences, constituents are assumed to undergo movement (Chomsky, 1957, 1981). A constituent is a word or a group of words that functions as a single unit within the sentence. Movement refers to the displacement of constituents in a given sentence. The dependency between the original position of the constituent (known as gap, or trace) and its actual position in the sentence is represented by co-indexation. In (4), the trace and its co-indexation are represented by *t*. In non-canonical sentences, such as in the passive in (4b), the verb assigns the thematic role to the trace or gap, which then gets passed on to the argument. In English, examples of non-canonical sentences are passives, object relatives, object-clefts, and wh-questions.

(4) a. The dog chased the cat.
  b. The dog_t was chased (t) by the cat.

In the sentences shown in (4), the thematic roles are semantically reversible, i.e., semantically, both noun phrases could in principle be the agent and the patient. This contrasts with the sentences in (5), where the second noun phrase is inanimate, and therefore cannot be assigned the agent theta-role.

(5) a. The woman found the box.
  b. The box_t was found (t) by the woman.

Although IWA's performance is variable (Berndt et al., 1996), IWA are generally reported to have more comprehension difficulties in reversible, non-canonical sentences, such as (4b), in contrast to reversible, canonical sentences (Caplan et al., 2006; Caramazza & Zurif, 1976), such as (4a). This effect is commonly known in the aphasia literature as the *canonicity* effect (Caplan & Hildebrandt, 1988).

Besides constructions involving movement and reversible thematic role assignment, IWA also show difficulties processing binding relations, i.e., pronouns and reflexives (e.g., Choy & Thompson, 2010; Edwards & Varlokosta, 2007; Engel, Shapiro, & Love, 2018; Thompson & Choy, 2009). Referential expressions such as pronouns and reflexives lack meaning on their own, and refer to a noun in the sentential context, known as antecedent. Therefore, processing structures that contain binding relations involves resolving the dependency between the antecedent and the referential expression. In Chomskys' framework of government and binding (Chomsky, 1981), binding relations involve a co-indexation between the referential expression and their antecedent. Chomsky's binding principles establish that whereas reflexives are bound in their local domain, pronouns are not. For instance, consider the sentences in (6), taken from Edwards and Varlokosta (2007). The co-indexational relation between the referential expression and their antecedents is shown with subscripts. In sentence (6a), the antecedent of *himself* cannot be bound by *John*, as it is outside of its local domain. By contrast, in (6b), the pronoun is bound by the subject of the matrix clause, which is outside of the local domain.

(6) a. John$_{\mathrm{i}}$ thinks that Bill$_{\mathrm{j}}$ likes himself$_{\mathrm{j}}$
b. John$_{\mathrm{i}}$ thinks that Bill$_{\mathrm{j}}$ likes him$_{\mathrm{j}}$
(Edwards & Varlokosta, 2007)

A great deal of (psycho)linguistic theories of impairments in aphasia have been developed over the last decades. Many of these theories were built on experiments testing the comprehension of sentences involving movement and binding relations (e.g. Caplan et al., 2015; Caplan et al., 2007; Choy & Thompson, 2010; Ferrill et al., 2012; Grodzinsky & Reinhart, 1993; Piñango & Burkhardt, 2005). Broadly speaking, theories of sentence comprehension in aphasia fall within one of these two categories: Representational accounts, and processing deficits accounts. Representational accounts assume that the cause of comprehension difficulties in IWA is related to a loss in grammatical knowledge, i.e., in Chomsky's framework, in competence rather than in performance. The most well-known theory within this framework is the Trace Deletion Hypothesis (Grodzinsky, 1995), which claims that IWA cannot represent traces of syntactic movement. However, there are a number of studies that speak against representational accounts. One example is that in grammaticality judgments, IWA have been reported to detect complex grammatical violations, even with structures involving traces (Zurif, Swinney, Prather, & Love, 1994). Another example comes from studies testing the same linguistic construction across different tasks. For instance, Caplan et al. (2006) and Caplan et al. (2013) tested constructions involving movement and binding relations in three tasks, and they found that most comprehension deficits arose in only one task, rather than systematically across the tasks. Moreover, there are cases in which IWA perform better at comprehending syntactically complex structures that involve movement relative to simple sentences (e.g., Caplan et al., 2007). These results cannot be accounted for by a theory that posits a loss of grammatical knowledge. Instead, the breakdown in parsing could be the result of processing difficulties.

Processing accounts assume that IWA's grammatical knowledge is intact, and that the comprehension deficits are caused by limitations in processing capacities, such as a lower working memory capacity, or a delay in lexical access. Some of the processing accounts are also somewhat grounded on linguistic theory. For instance, the delayed lexical access theory posits that a processing deficit in IWA slows down the formation of syntactic structure (Burkhardt et al., 2003; Love et al., 2008). Other theories hinge on the role of more general cognitive processes, such as working memory (e.g., Caplan, 2012; Caplan et al., 2006). As of today, there is no consensus on the source of comprehension deficits in IWA, mainly because a single deficit cannot account for the range of impairments reported in IWA. Recently, it has been proposed that the performance of IWA may be better explained by a combination of processing deficits, rather than by assuming a single source of impairment (Caplan et al., 2015; Mätzig et al., 2018).

One important methodological question that arises is, assuming that an interplay of several processing deficits is what causes comprehension difficulties in IWA, how can we assess the extent to which each of these deficits are responsible for IWA's impairments? One way to evaluate the different theories is to implement the different accounts in a cognitive model, and to test them against experimental data from IWA. In this dissertation, cue-based retrieval theory, which is designed to explain dependency completion in sentence processing, is the framework used to implement and assess the different theories of processing deficits in aphasia. Specifically, it is proposed that different processing-based deficits hinder retrieval from memory, which is needed for dependency completion. This leads to (a) more susceptibility to competing interpretations, and (b) more failures in parsing. The goal of this dissertation is to computationally investigate the extent to which different processing deficits play a role in IWA, in the framework of two distinct cognitive models of cue-based retrieval.

2.2 Theories of sentence processing deficits in aphasia

In this section, the different processing deficit theories that will be computationally evaluated in this dissertation are presented. These are: slow syntax, delayed lexical access, intermittent deficiencies, and resource reduction.

2.2.1 Slow syntax

Most of the research that supports the slow syntax theory (Burkhardt et al., 2008; Burkhardt et al., 2003; Piñango & Burkhardt, 2005) comes from experiments showing that IWA correctly use grammatical information in real-time sentence comprehension, but in a delayed manner. The slow syntax theory proposes that IWA do not have an impairment at the level of syntactic building proper, but that IWA's syntactic computations are slowed down. As the syntactic formation chain is delayed, the upcoming information cannot be parsed, and extra-syntactic heuristics, based on semantics or on frequency, may arise. Competition between syntactic and extra-syntactic information can lead to the correct interpretation of the sentence if IWA establish the dependency using the syntactic structure, or to a misinterpretation if the dependency completion is based on extra-syntactic heuristics, such as semantic information. Therefore, the slow syntax account predicts that IWA should exhibit comprehension deficits in thematic role assignment when a full structural representation is needed; i.e., when comprehenders cannot rely on extra-syntactic information, such as in semantically reversible non-canonical sentences.

The first studies testing the slow syntax theory used a cross-modal syntactic priming (CMSP) paradigm. In this task, participants hear a sentence preamble, and at a given time point, a letter string (visual probe) is presented in a computer screen. The visual probe is the continuation of the sentence preamble, and participants have to decide whether the continuation is acceptable or not by pressing a yes/no button. For example, Friederici and Kilborn (1989) report two experiments with 5 IWA and 28 controls using a CMSP task. The authors tested sentences such as in (7) in Dutch, although the examples are shown here in English for the sake of simplicity. The visual probe for the decision task is shown in bold. The sentence preambles contained a variety of auxiliaries in Dutch (e.g., passives: *is being*), and tense (perfect: *has*, *is*).

(7) a. The indian is being **helped**.
    b. *The poet is being **climbed**.

The results in Friederici and Kilborn (1989) yield two main findings: In Experiment 1, IWA exhibited faster decision times in grammatical conditions (7a) relative to ungrammatical conditions (7b). In Experiment 2, which tested the same items, IWA exhibited significantly faster decision times when the interstimulus interval was longer (i.e., when there was more time between the end of the sentence preamble and the appearance of the visual probe; from 0 ms in Experiment 1, to 200 ms in Experiment 2). On the basis of these results, the authors claim that syntactic knowledge in IWA is preserved, but critically delayed in comparison with unimpaired controls.

Haarmann and Kolk (1991b) also tested sentences similar to (7) in Dutch in a CMSP task with 13 IWA and 13 controls, and their items included a greater number of grammaticality violations (e.g. modal verbs: We can talk/ *nose; preposition: On the cupboard/ *smoke). In addition, the visual probes for the lexical decision task were presented at different time intervals after the sentence preamble (300 ms, 700 ms, and 1100 ms). Overall, controls showed syntactic effects of priming (i.e., faster decision times in grammatical vs. ungrammatical conditions) at the three time points.

By contrast, IWA only showed significant effects of syntactic priming at the 1100 ms time point. The authors concluded that the late appearance indicates that IWA's syntactic information is activated at a slower-than-normal rate, which is consistent with the slow syntax theory.

More recent studies investigating the time-course of syntactic information have used a cross-modal lexical decision (CMLD) paradigm. In this paradigm, subjects have to simultaneously perform a comprehension task, and a lexical decision task. Sentences are presented auditorily, and subjects have to ask comprehension questions. In addition, at some point during the auditory presentation of the sentence, a letter string is visually presented in the monitor. Subjects decide whether the letter string is a word or a non-word, by pressing a yes/no button. This method is particularly designed to investigate the time-course of the activation of a specific element in the sentence (target word) because the visual probe semantically related or unrelated to the target word.

For instance, Burkhardt et al. (2003) investigated wh- and NP-movement in English using the CMLD paradigm. The main goal of the study was to investigate gap-filling in IWA, i.e., whether IWA could reactivate traces by showing, in real-time processing, reactivation of the antecedent at the trace (or gap) position. Experiment 1 (3 IWA and 10 controls) targeted the comprehension of object relative clauses, such as the example sentence (8).

(8) The kid loved the cheese$_{j}$ which $_{j/i}$ the brand new microwave *melted* t$_{i}$ yesterday afternoon while the entire family was watching TV.

For the example (8), the corresponding visual probes were *cheddar* (semantically related) and *album* (semantically unrelated). Since priming effects were taken as an indication for antecedent reactivation, these visual probes were shown at different time points during the sentence, so as to determine at which particular point the antecedent was being primed. The results of Experiment 1 show that unimpaired controls already experience priming effects at 100 ms after the verb, whereas IWA's priming effects emerged at 650 ms after the verb.

In Experiment 2 (2 IWA, 23 controls), Burkhardt et al. (2003) tested the comprehension of active sentences with unergative (9a) vs. unaccusative verbs (9b). According to the Split Intransitivity Hypothesis (Burzio, 1986; Perlmutter, 1978), unergative verbs are assumed to base-generate their argument preverbally, whereas unaccusative verbs base-generate their argument postverbally. In English, active sentences with unaccusative verbs are therefore assumed to undergo movement from the postverbal argument trace to the subject position. Crucially, the authors hypothesize that in sentences such as (9b), a priming effect would be expected sometime after the verb, at the trace. By contrast, no priming effect is expected at the same point in (9b).

(9) a. The graduate with a passion for movies *celebrated* after the last of the official ceremonies was over.

b. The butter$_i$ in the small white dish *melted* t$_i$ after the boy turned on the brand new microwave.

In sentences with unaccusative verbs, controls showed activation of the antecedent at 650 ms after the verb, but, as predicted, not in sentences with unergative verbs. IWA showed the expected pattern of priming at around 800 ms in sentences with unaccusative verbs, although this effect did not reach significance. No priming effects were detected in sentences with unergative verbs.

Considering the results from both experiments, Burkhardt et al. (2003) concluded that IWA can successfully process movement, but in a protracted manner. That is, IWA do not reactivate the antecedent at the trace, as controls do, but they do reactivate the antecedent at a later point in the sentence, as shown in the delayed priming effects. The authors concluded that syntactic operations involving moment are critically delayed, both in movement involving wh-traces and object-NP traces.

Finally, the slow syntax theory has been also tested with reflexives and logophors. Piñango and Burkhardt (2005) and Burkhardt et al. (2008) tested reflexives and logophors in English (2 IWA, 13 controls) and Dutch (3 IWA, 16 controls), respectively. Their results show that IWA can resolve reflexive-antecedent dependencies, but in a protracted manner. Specifically, Burkhardt et al. (2008) propose that the Merge operation (Chomsky, 1995) is delayed. Merge consists on the combination of two syntactic elements that form a constituent. The authors argue that a delay in Merge would have the effect that the arguments of a predicate would become available at a later-than normal point. In the meantime, competing extra-syntactic sources of information may appear, leading to possible misinterpretations.

All in all, all the studies testing the slow syntax theory show that syntactic operations in IWA are delayed. If given enough time, IWA could, in principle, understand sentences involving movement, as their syntactic knowledge is unimpaired. However, due to the slower-to-normal formation of syntactic dependencies, non-syntactic heuristics may kick in, providing two competing interpretation alternatives. This would explain the comprehension pattern that is reflected in the aphasia literature: In complex sentences, in some trials, IWA show slow reaction times, and yet correct comprehension responses. However, in trials in which extra-syntactic heuristics lead the processing, slow processing times and incorrect responses are expected. Although the exact nature of the syntactic parsing operation that is delayed is not clear (e.g., general processing slow-down in dependency completion, delay in trace reactivation, delay in the Merge operation), these studies show that syntactic structure building is generally delayed in IWA.

2.2.2 Delayed lexical access

The delayed lexical access theory (DLA) posits that a delay in lexical access is the main cause of comprehension difficulties in IWA. According to this theory, a slower-to-normal access to lexical items causes breakdowns in the syntactic structure building, leading to syntactic comprehension difficulties. The DLA theory has been tested with priming experiments, and mostly in filler-gap dependencies.

The first studies investigating DLA appeared in the 90s. Prather, Zurif, Love, and Brownell (1997), Prather, Zurif, Stern, and Rosen (1992), and Prather (1994)[1] carried out a series of case-studies in which they showed that lexical activation in IWA is delayed in comparison with language unimpaired populations. These case-studies analysed the time-course of lexical activation using list priming paradigm experiments. In this task, participants are shown letter strings on a computer screen, and they have to perform a lexical decision task as quickly as possible. As soon as the participant presses the yes/no button, the word disappears from the screen and after a predetermined time, known as interstimulus interval, the next word is shown. The list of items contains experimental word pairs (e.g., chair-table) and filler items. The core idea behind this task is that response times at the lexical decision task should be faster when the target word is preceded by a prime word semantically associated with it. So, for instance, participants are expected to have faster RT when *table* is preceded by *chair*. By manipulating the interstimulus interval, this task allows for the investigation of the time-course of priming, as the shortest and longest intervals at which priming effects emerge can be determined. Typically, unimpaired elderly and college-aged adults are known to start showing priming effects at the interstimulus interval of 500 ms (Stern, Prather, Swinney, & Zurif, 1991). By contrast, the case-studies by Prather and colleagues revealed that priming in IWA started to emerge at the 1500 ms interstimulus interval. In these case-studies, Prather and colleagues discuss that the delay in lexical activation could lead to comprehension difficulties in sentences involving traces, as it could cause a failure to complete the antecedent-trace

[1] Prather et al. (1997) tested two individuals with aphasia, one of them had non-fluent Broca's aphasia, whereas the other one had Wernicke's aphasia. Whereas the results for the both patients were different (the idividual with Broca's aphasia did not show priming effects until the 1500 ms interval, the individual with Wernicke's aphasia showed priming effects from 300 ms to 1100 ms. The early priming effects of the patient with Wernicke's aphasia emerges as early as priming emerges in unimpaired control participants.

dependency in time.

Online performance in gap-filling structures in IWA was subsequently assessed in Swinney, Zurif, Prather, and Love (1996), Zurif, Swinney, Prather, Solomon, and Bushell (1993), using a CMLP task (see also Zurif et al., 1994, for a review). Zurif et al. (1993) tested subject relatives, and Swinney et al. (1996) tested object relatives. Both studies revealed that individuals IWA did not show priming effects at the pre-gap and at the gap positions.[2] These results are compatible with the idea that the incapacity to reactivate the antecedent is what drives the comprehension problem in sentences involving movement.

Building on the work by Zurif and colleagues on the use of the CMLP paradigm to assess filler-gap dependencies in real time processing, Love et al. (2008) claim that the slower-than-normal speed of lexical activation disrupts syntactic operations, and specifically, gap-filling. Love et al. (2008) report three experiments; the first two involve the cross-modal lexical priming paradigm, and the third one is a sentence-picture matching task.

The first two experiments tested object relative clauses such as in (10). The visual probes for the lexical decision task were semantically related to the moved constituent (*fighter* for example 10) or unrelated (*climber* for example 10). The visual probe words from the lexical decision task were presented at different time points during the auditory presentation of the sentence, in order to investigate at which time-points the moved constituent primes the visual probe. Priming was measured by comparing the RT from the lexical decision task in the related vs. unrelated conditions.

(10) The audience liked the wrestler that the parish priest condemned (t) for foul language.

In Experiment 1 (8 IWA and 4 controls), the sentences were presented auditorily with a normal speech rate: 4.47 syllables per second. IWA showed a delayed effect of priming after the antecedent (about 300 ms after) and a delayed effect of priming approximately 500 ms after the gap (marked as t in exdample 10). Therefore, IWA showed both delayed lexical activation when encountering the movement element and also delayed reactivation of the element at the trace.

In Experiment 2 (9 IWA and 6 controls), which tested the same items than Experiment 1, the sentences were presented auditorily at the slowed-down rate of 3.4 syllables per second. In this case, priming in IWA emerged at the gap, and also 500

[2] Swinney et al. (1996), Zurif et al. (1993) report case studies for Wernicke's and Broca's patients and focus in the differences between the two types of aphasia. As in Prather et al. (1997), the results are slightly different for patients with different aphasia types. However, as discussed previously, the difference in performance between Broca's and Wernicke's aphasia is beyond the scope of this thesis.

ms after the gap. In contrast to Experiment 1, IWA did show reactivation at the gap. The authors take this results as evidence that IWA are able to work out traces of syntactic movement if given enough time. Controls showed no effect of priming at the gap, but priming emerged 500 ms after the gap. This suggests that that slower rate of the speech input disrupts sentence comprehension in controls.

Experiment 3 (8 IWA and 10 controls) consisted of a sentence-picture matching task that targeted the comprehension of the thematic roles in subject and object relatives, and active and passive declarative sentences. The sentences were presented auditorily at a normal speed rate (5.5. syllable per second) and at a slow rate (3.8 syllables per second). For canonical sentences, in conditions with both normal and slow rates, IWA's accuracies were around 80%. By contrast, for non-canonical sentences, the mean accuracy in the conditions with a normal speech rate was 61%, and in conditions with slow speech rate, 71%.

Overall, considering the results from the three experiments, Love et al. (2008) concluded that with a normal speech rate, IWA show delayed lexical activation when hearing the moved constituents (antecedents) and delayed reactivation at the gap site. By contrast, with a slow speech rate, IWA show reactivation at the gap site. The authors argue that syntactic constructions involving constituent movement are difficult to process for IWA because because lexical reactivation may be completed at a slower-than-normal pace. Due to this delay, non-grammatical heuristics may emerge and provide a conflicting interpretation that does not depend on syntax.

Further evidence for the delayed lexical access hypothesis comes from the study by Ferrill et al. (2012). As mentioned above, most of the studies testing the delayed lexical access hypothesis targeted a) lexical delays at the word level in lexical decision tasks, or b) lexical delays in filler-gap dependencies. Ferrill et al. extend the scope of the delayed lexical access theory by showing that delays in lexical access can also be observed in sentence comprehension in simple, non-canonical constructions.

Ferrill et al. (2012) used a cross modal picture priming paradigm, in which participants listened to sentences such as in (11). At different time points in the sentences, black-and-white line drawings (visual probes) were presented on a computer screen, and participants had to make a binary decision about the drawing: Human (yes) / not human (no). The visual probes were either related (they depicted the noun phrase in the direct object position of the sentence, i.e., *golfer* in 11) or unrelated to any of the words in the sentence.

(11) The boxer punched the *golfer* after the tremendously antagonistic discussion about the title fight

Their results revealed that controls showed a significant priming effect at the offset of the target noun (i.e., *golfer*), and no significant priming effects later on, which suggests that the meaning of the noun is activated immediately after encountering it, and that it decays rapidly. By contrast, IWA showed no priming effect at the offset of the noun, and a significant effect of priming 400 ms after the noun. Therefore, Ferrill and colleagues conclude that the lexical delay feeds syntactic structure building too slowly, causing misinterpretations and failures in parsing in complex structures that rely on syntactic structure. Essentially, Ferrill et al. (2012) argue that a delay in lexical access is the source of the syntactical disorders.

Overall, all of the studies testing the delayed lexical access hypothesis show that lexical access is critically delayed in IWA, and that such a delay impacts real-time sentence processing. In addition, the work by Ferrill et al. (2012) shows that a theory such as slow syntax cannot fully explain the full range of sentence comprehension deficits in aphasia, as it cannot account for delays in simple, declarative canonical sentences.

2.2.3 Resource reduction

The resource reduction hypothesis was developed by Caplan and colleagues in a series of studies. The work by the Caplan group is in many aspects novel in comparison with previous studies in aphasia in several aspects: First, Caplan and colleagues test a large number of participants (40 to 56 IWA) in several tasks, usually self-paced listening, picture-sentence matching, grammaticality judgments, and object manipulation. Second, in each study, Caplan and colleagues investigate a variety of linguistic structures. These structures are usually tested in sentence pairs that contain an experimental and a baseline condition, and the results are interpreted in terms of the differences in processing times and/or accuracies between the two conditions. Overall, the experiments by the Caplan group have shed important insights on the variability, the dissociations, and the associations of performance within and between IWA, and across tasks and sentence types.

The resource reduction hypothesis (Caplan, 2012) claims that the source of impairments in IWA is a reduction in the resources that are necessary for parsing. Processing resources may be related to language-specific mechanisms, such as slower lexical access, or a limitation in a specific verbal processing system, as proposed in Caplan and Waters (1999); or to general mechanisms in the cognitive architecture (e.g., working memory capacity, encoding, perception, action planning). Resource demands are determined by the complexity of a sentence structure and the specific task in which the structure is being tested. Caplan postulates that more complex sentences (and more complex tasks) require the use of more parsing resources. Therefore, sentences involving movement, such as object relatives, should be more difficult to process than sentences without movement, because the former require a greater amount of resources. The resource reduction hypothesis relies on three main components: Sentence demands, task demands, and the resources of the individual.

If the resources of an IWA reach the demands of the sentence and task, the resource reduction hypothesis predicts normal-like processing. However, if the processing demands of a sentence structure in a given task exceed the available processing resources of an individual, sentence comprehension is predicted to be impaired. Therefore, IWA are more likely to experience difficulties in complex sentences because these are more likely to exceed their resources. In addition, IWA with a higher degree of resource reduction are assumed to be more affected by sentence complexity.

The resource reduction hypothesis was tested in Caplan et al. (2006) with data from 42 IWA and 25 unimpaired controls in two tasks: Object manipulation, and sentence-picture matching. The experimental items contained several complex constructions and their corresponding baseline constructions, namely object vs. subject relative clauses, passives vs. actives, and reflexive-antecedent dependencies. Within each construction, several sentences were tested: Full and truncated passives, cleft and center-embedded relative clauses, and reflexives in genitive (*the father of the boy scratched himself*) vs. possessive constructions (*the boy's father scratched himself*). The aim of the study was to investigate the occurrence of specific syntactic structure deficits across both tasks. Caplan et al. (2006) evaluated the data from each participant at the individual level, so the data from the 42 IWA are presented as 42 case studies, rather than as a group.

Overall, all IWA showed processing difficulties in more than one structure, but not in all sentence types. For 31 out of the 42 IWA, most comprehension deficits were task-specific (i.e., emerged in only one of the two tasks) and construction-specific (i.e., the deficit appeared in the two sentences that tested the same syntactic construction, e.g., truncated and full passives). The work by Caplan et al. (2006) therefore confirms that processing deficits in aphasia seem to be dependent on an interaction between the syntactic constructions and the task demands. This pattern of associations and dissociations in performance of IWA across tasks and across syntactic structures was later replicated in Caplan et al. (2013), with 61 IWA and 46 controls. Importantly, Caplan et al. (2013) carried out an extended replication of Caplan et al. (2006). Instead of testing 10 sentences of each type, Caplan et al. (2006) tested 20. This replication, with a larger number of participants and experimental items, confirmed that task-independent structure-independent deficits are almost non-existent in aphasia. The results of both studies speak in favor of a reduction in processing resources as the main source of comprehension deficits, together with an interaction between task demands and a reduction in parsing and interpretative abilities. The conclusions of both studies indicate that IWA may not be able to meet the demands required by a particular task on top of parsing certain constructions.

One crucial test for the resource reduction hypothesis is that it predicts that correct trials are the result of unimpaired processing. That is, in correct trials, IWA are not assumed to resort to guessing, or to require other extra-linguistic strategies. Therefore, the online data in correct trials should reflect normal-like processing. As hypothesized, Caplan et al. (2007) found that in self-paced listening times, IWA's pattern for correct trials was quantitatively similar to controls' pattern; whereas the patterns of both groups differed in incorrect trials. This finding has been replicated in later studies using the visual-world eye-tracking paradigm (e.g., Choy & Thompson, 2010; Dickey et al., 2007; Dickey & Thompson, 2009; Hanne et al., 2015; Hanne, Sekerina, Vasishth, Burchert, & De Bleser, 2011; Meyer et al., 2012). Yet, in correct trials in visual-world studies, IWA generally show delayed fixations to the target, relative to controls. Several studies (Hanne et al., 2015; Hanne et al., 2011; Meyer et al., 2012; Pregla, Vasishth, Lissón, Stadie, & Burchert, 2021; Schumacher et al., 2015) have attributed this pattern to a general processing slowdown. Such a slowdown is also compatible with the resource reduction hypothesis, as processing speed could be a reduced resource in IWA (see Pregla, Vasishth, et al., 2021).

The resource reduction also posits that IWA with lower resource capacities should be more affected by sentence complexity (Caplan et al., 2013; Caplan et al., 2007). Support for this assumption comes from two-self paced listening experiments, in which Caplan and colleagues found that IWA who had overall lower accuracies were more affected by complexity effects (Caplan et al., 2015; Caplan et al., 2007). These studies found a super-additive interaction between the demand of a specific sentence type and different groups of IWA, classified according to their accuracy performance: At chance, above chance but below normal, and within the normal range.

The resource reduction hypothesis has been further investigated by Caplan et al. (2015). This study reports self-paced listening data from 61 IWA and 46 controls, from 13 sentence types. Examples from the sentences, taken from Caplan et al. (2015), are shown in (12). The authors compared the residual corrected listening times[3] at the critical word in the experimental sentences to their corresponding baseline sentences.

(12) a. **Active:** The girl hugged the boy.

[3]Listening times were corrected (word duration was subtracted) and adjusted for word frequency by regressing them against word frequency within each participant and calculating the residual of this regression for each word for each participant.

b. **Passive:** The boy was hugged by the girl.
c. **Three noun phrases:** The niece said that the girl hugged the boy.
d. **Pronoun:** The niece said that the girl hugged the her.
e. **Reflexive:** The niece said that the girl hugged the herself.
f. **Cleft subject:** It was the girl who hugged the boy.
g. **Cleft object:** It was the boy who the girl hugged.
h. **Subject object:** The boy who the girl hugged washed the woman.
i. **Subject subject:** The girl who hugged the boy washed the woman.
j. **Subject subject, pronoun:** The woman who hugged the girl washed her.
k. **Subject object, pronoun:** The woman who the girl hugged touched her.
l. **Subject subject, reflexive:** The woman who hugged the girl washed herself.
m. **Subject object, reflexive:** The woman who the girl hugged touched herself.

Their analysis of the pooled data[4] revealed that object-extracted sentences elicited higher listening times than subject-extracted sentences for both IWA and controls; and an indication of higher listening times for pronouns than for three-noun phrase structures, although this effect did not reach significance. Overall, IWA showed significantly higher listening times in object-extracted sentences than in any other sentence type.

Caplan et al. (2015) claim that their data is in line with the resource reduction hypothesis for several reasons. First, there was a super-additive interaction between the demand of a specific sentence type and the performance of the different groups of IWA: IWA at or above chance level had higher listening times than controls in object relatives, but such difference did not emerge between IWA at normal range and controls. In addition, no interaction between groups of IWA and controls in other sentence types was found, which suggests that object-extracted relatives require a greater resource demand. Second, the difference in listening times between the experimental and baseline conditions increased as accuracy decreased from the normal range to above chance to at chance, and overall, the difference was greater in sentences with lower accuracies. The authors propose that this pattern can be seen as a gradient of on-line performance that depends on the complexity of the sentence, and the degree of resource reduction in IWA. Moreover, Caplan et al. (2015) hold that slowed lexical access and slowed syntax may be the core mechanisms that are affected by resource reductions. This claim is based on the online data: Residual corrected listening times at the critical word in correct trials in all experimental sentences correlated

[4] We use the term *pooled data* to refer to the data from all IWA and controls in all sentences. Notice that Caplan et al. (2015) performed a variety of analysis, including analysis that categorized IWA's performance between at chance, below chance, and above chance; and an analysis of IWA's performance at the individual level.

positively with reation times from a lexical decision task that had been administered independently from the experiment. In addition, the residual corrected listening times in non-critical words in the sentence also correlated positively with the listening times at the critical regions, where the syntactic parsing operations of interest are expected to happen. These correlations speak in favor of delays in lexical access and in syntactic structure building.

In German, the resource reduction hypothesis has been investigated recently, with visual-world data from 21 IWA and 50 controls in Pregla, Vasishth, et al. (2021). In this study, participants heard a sentence at a normal speech rate (4.79 syllables per second) while two pictures (target and foil) were displayed in a computer screen, all through the trial. At the end of the auditory presentation of the sentence, participants had to choose which picture depicted the meaning of the sentence. Participants were tested using the same stimuli in two separate sessions (test-retest) with a gap of a month in between. The experimental sentences included simple declarative sentences (SVO vs. OVS), relative clauses (subject vs. object relatives), subject and object control structures, and pronoun resolution. The visual-world data was analysed using a time bin and time windows analyses. The authors found that, as expected under the resource reduction hypothesis, processing difficulty is more frequent in complex sentences relative to simple sentences, and that IWA show a general slowdown in processing, in all sentence structures. However, in contrast with the predictions of the resource reduction hypothesis, Pregla and colleagues found some quantitative differences between IWA and controls in correct trials. While IWA's gazes reflect a preference for the target picture across the trial (which confirms that IWA are not guessing, as predicted by resource reduction), IWA's maximum target fixations remained below controls' maximum target fixations at the picture-selection task. The authors take this as an indication that IWA show more uncertainty than controls.

Overall, the studies investigating resource reductions show that sentence comprehension in IWA is modulated by both sentence complexity and task demands, and that variability between and within IWA is unsystematic, which speaks in favor of a random-error generating component in the processing system of IWA. The role of this component, usually referred to as *noise*, will be further explored in the intermittent deficiencies account, which is presented next.

2.2.4 Intermittent deficiencies

The intermittent deficiencies account claims that the source of comprehension difficulties in IWA are intermittent reductions in IWA's processing capacities. This account, also developed by Caplan and colleagues, can be seen as a combination of resource reduction and noise. Caplan (2012) characterizes resource reduction and noise in the following way:

> The level of random error generation [noise] may be considered one way to conceive of resource reduction. Since it postulates the fewest possible constructs, a theory that postulates only demand associated with each sentence type and task and a random error-generating factor is to be preferred on metatheoretical grounds, if it is workable. The next most severely restricted model would be a three-parameter model that recognizes resource demand associated with sentence types and tasks and both noise and resource reduction in individual patients. Models of this sort could be explored in many ways. *Noise might be held constant and resource reduction allowed to vary, or both may vary; if the latter is the case, the level of noise may be related to the degree of resource reduction or not* [emphasis added]. The important point is that two features of a model of aphasic deficits in syntactically based comprehension are clearly necessary: resource demand associated with different sentence types and variable amounts of noise in different patients. (Caplan, 2012, p.46).

In this thesis, we follow Caplan (2012), and implement resource reductions and intermittent deficiencies as two separate sources of comprehension deficits. Specifically, the intermittent deficiencies theory will be associated to the noise component in our models.

Most of the findings in the studies previously presented in the resource reductions section also speak in favor of the intermittent deficiencies theory. But specifically, the results of two studies (Caplan et al., 2006; Caplan et al., 2007) are directly interpreted in terms of intermittent deficiencies. These two studies report data from experimental sentences that included the following contrasts: Actives vs. passives, subject vs. object extracted relative clauses, and sentences with and without reflexive pronouns. The comprehension of these sentences was tested in object manipulation, sentence picture matching, and grammaticality judgments. The latter tasks were tested in two different modalities: Whole sentence auditory presentation, and self-paced listening. Accuracies in all tasks, as well as RT for the picture-selection task and grammaticality judgments were recorded.

The findings of Caplan et al. (2006) have been summarized in the previous section, in the context of the resource reduction hypothesis. However, one specific aspect of the data adds support to the claim that there is a random error-generating process (Caplan et al., 2006, p. 920) in the processing system that can also cause comprehension deficits in simpler syntactic structures: In about 5% of the trials, IWA experienced a reversed-pattern, i.e., abnormal performance in baseline sentences, and normal-like performance in the related experimental counterpart, in the same task.

The results of Caplan et al. (2007) show that across the three tasks, both online (self-paced listening) and offline data (accuracies) indicate that object relative clauses are more difficult to process than subject relatives. The processing cost appears at the verb of the object relative clauses. Similarly, the online data also indicate that IWA experience some processing load at the verb of the passives compared to the actives. In addition, greater effects were found for IWA with lower accuracy. Given that normal-like online processing was found in correct trials, Caplan et al. (2007) claim that the source of the failures in parsing are not in the syntactic structure building proper (as in correct trials the syntactic structure is built correctly), but that the parsing system fails intermittently, with more frequent failures in complex sentences, where processing load is higher, and especially when resource availability is low. Therefore, Caplan et al. (2007) argue that IWA's performance is affected by the complexity of the sentence, the task, and stochastic noise inherent to the participant.

Support for the intermittent deficiencies theory in German-speaking IWA comes from Hanne, Burchert, and Vasishth (2016), who tested subject-extracted and object-extracted who-questions in German in 8 IWA and 40 controls using the visual-world paradigm and and sentence-picture matching. Controls showed incremental processing of case-marking cues, and their gaze patterns revealed that they resolve the filler-gap dependency at the gap position. By contrast, IWA's online data show that gap-filling is delayed. The offline results for IWA reveal three different patterns: Similar comprehension and same impairments for both question types; b) better comprehension of subject compared to object who-questions and c) reversed asymmetry, i.e., better comprehension of object compared to subject questions. The authors attribute the comprehension errors of IWA to intermittent failures in parsing the wh-dependencies in wh-questions and in integrating / using case cues.

Finally, the work by Pregla, Lissón, Vasishth, Burchert, and Stadie (2021) and Pregla, Vasishth, et al. (2021) in German also provides some support for intermittent deficiencies. In both online and offline measures, Pregla and colleagues found unsystematic variability in IWA: Changes in performance between test and retest sessions were unsystematic between and within IWA, whereas controls showed practise effects from exposure to more complex sentences (i.e., controls performed better in complex sentences in the retest session relative to the test session). The authors attribute this variability to the role of noise in the processing system.

2.3 Previous modeling of sentence processing deficits in aphasia

In this section, we present a short overview on studies that have computationally investigated the role of the previously presented processing deficits in aphasia.

One of the first computational models of sentence processing in aphasia, SYNCHRON (Haarmann & Kolk, 1991a), is a timing-deficit based model. The core assumption in this model is that parsing fails due to an incapacity to maintain syntactic elements simultaneously co-active. The model builds phrase-structure representations using a series of retrievals from working memory. The final retrieval, the phrasal category, can only happen if all the constituent categories are simultaneously available in memory. The model assumes that IWA have a time-based impairment, which could be due to a) a capacity reduction of in syntactic working memory, b) a decrease in the activation rate of items in memory, or c) an increase in the decay of representational elements. Such impairment is the source of comprehension difficulties in complex sentences, because it causes a disruption in the simultaneity of the phrasal category. The model can also account for degrees of severity among different IWA. However, this model lacks the structure to account for impairments in thematic role assignment, which is one of the most studied phenomena in comprehension difficulties in aphasia (Caramazza & Zurif, 1976).

The Capacity Constrained Resourced Deficit (CCRD) model, proposed by Haarmann, Just, and Carpenter (1997), is based on the assumption that the source of impairments in aphasia is a low working memory capacity (see also Just & Carpenter, 1992). Model components include lexical access, parse tree building, and thematic role assignment. The CCRD model provided a good fit to the accuracy data reported in Caplan et al. (1985) and in Kolk and Van Grunsven (1985), and could account for sentence complexity effects by varying the working memory parameter, as more complex sentences are assumed to require higher working memory demands. Other time-deficit based computational models proposed for aphasia include the HOPE model (Gigley, 1983, 1988), and the Unification Space model (Kempen & Vosse, 1989). In general, all of these models assume that successful language comprehension requires the co-activation of two or more representations, and that IWA experience some sort of time-based deficit that disrupts such co-activation (Haarmann et al., 1997), such as faster memory decay, higher fluctuations, lower activation values, and/or slower times in parsing steps.

The resource reduction hypothesis has been computationally investigated in the studies by Gutman, DeDe, Michaud, Liu, and Caplan (2010) and Gutman, DeDe, Caplan, and Liu (2011), who used the offline data from Caplan et al. (2006) and implemented (extended) Rasch models to investigate the relationship between task demands, resource reduction, and syntactic structures. Rasch models compute the probability of providing a correct response as a linear function of individual capacity (i.e., degree of impairment of each IWA), and sentence complexity. Different models with increasing complexity were implemented and compared against each other by assessing their goodness of fit. IWA's individual accuracy performance on all sentences in one or the two tasks (object manipulation and picture-sentence matching) were taken as an index of IWA's level of resource reduction (i.e., degree of impairment). Sentence demands were determined by computing the total correct responses on each sentence across all IWA in one or the two tasks. The best fitting model included task and patient groups, but not sentence types. This suggests that overall, sentence performance in IWA may be explained by the different levels of resource reduction in IWA and task demands. Moreover, the results in Gutman et al. (2010) speak against syntactic-based deficits (e.g., the trace deletion hypothesis, Grodzinsky, 1995), as sentence type was not a determinant factor in the model that provided a better fit for the data. One caveat of these studies, however, as pointed out in Caplan et al. (2013), is that the Rasch models may not have had enough power to detect an effect of sentence type.

Finally, the cue-based retrieval model of Lewis and Vasishth (2005) has been adapted to model aphasic sentence processing in the studies by Patil et al. (2016) and Mätzig et al. (2018). Patil et al. (2016) modeled the visual-world data of Hanne et al. (2011), which tested SVO vs. OVS sentences in a sentence-picture matching task. Eye movement patterns, accuracy, and response time were considered, and the data from the 7 IWA was modeled individually. Patil and colleagues developed a series of models based on the Lewis and Vasishth (2005), in which different theories of processing difficulties in aphasia were implemented. Specifically, Patil et al. (2016) implemented a model with slower procedural memory (interpreted as the implementation of slow syntax and delayed lexical access), a model with extra noise in the parsing steps (interpreted as the implementation of intermittent deficiencies), and a model containing both modifications. In addition, Patil et al. also implemented two LV05-based models that simulated the predicted parsing behavior under two different versions of the trace-deletion hypothesis (Grodzinsky, 1995). The results show that IWA's behavior in sentence processing is better captured by the model that combined the two processing deficit accounts: Slowed processing and intermittent deficiencies. Their results also highlight that patients may be affected by these deficits to a different extent, which suggests the existence of considerable variability among IWA.

Mätzig et al. (2018) investigated the variability in processing deficits in IWA by estimating specific parameters of the Lewis and Vasishth (2005) model at the individual level. Slowed processing (understood as delayed lexical access and sow syntax), intermittent deficiencies, and resource reduction were mapped onto different parameters of the model. Mätzig et al. (2018) fitted their model to the accuracy data from the subject vs. object relative clauses data reported in Caplan et al. (2015). Overall, IWA's range of parameters show great variability, whereas controls' parameters are less variable, and closer to the default parameters in the original Lewis and Vasishth (2005) model. These results confirm the conclusions in Caplan et al. (2015): Several processing deficits may be responsible for IWA's comprehension difficulties, and deficits may lie on a graded continuum, depending on the degree of impairment of each individual.

In general, previous studies implementing computational models of sentence processing in aphasia have yielded the following insights: First, time-based deficits (either due to a delay in lexical access or in syntax / parsing operations) are a crucial component in sentence comprehension in aphasia. Second, individual IWA may suffer from several processing deficits, and to different degrees. Models of sentence processing in IWA should account for a range of possible processing deficits in IWA, and for variability along the continuum of deficits. Third, models of sentence processing should also account for sentence and task complexity, as well as a possible interaction between these two factors.

Finally, computational modeling in aphasia has shown that constraining theories by formalising their principles and assumptions fosters theory development. As an example, consider the results in Patil et al. (2016) and Gutman et al. (2010), which indicate that specific syntactic impairments cannot account for the data from IWA. Instead, they point towards a combination of processing deficits as the source of comprehension problems in IWA. Moreover, adapting existing models of sentence comprehension in unimpaired populations to IWA has helped developing a connection between general processing resources and impairments in aphasia, which, in turn, allows for a better understanding of comprehension disorders in aphasia (De Bleser et al., 2012). For instance, both Mätzig et al. (2018) and Patil et al. (2016) have shown that the Lewis and Vasishth (2005) model of unimpaired sentence processing can account for IWA's performance by fine-tuning specific parameters that can be mapped onto theoretically-informed deficits. This dissertation builds on their work, and examines the role of processing deficits in the context of two computational models of sentence processing: The Lewis and Vasishth (2005) model, and McElree's direct-accecss model (McElree, 2000).

Chapter 3

Computational models of retrieval processes

This chapter details the computational implementation of the models of cue-based retrieval that are evaluated in this dissertation. Section 3.1 is an introduction to the general computational framework in which the models are implemented. The next sections present the implementations of the activation-based model (Lewis & Vasishth, 2005) and the direct-access model (McElree, 2000), as well as a modified version the direct-access model, which has been developed in this dissertation. Finally, the last part of the chapter is dedicated to model comparisons.

3.1 Bayesian cognitive modeling

The competing models of sentence processing are implemented in the Bayesian probabilistic programming language Stan (Carpenter et al., 2017). The package *rstan* (Stan Development Team, 2021a) was used to fit the models through R.

The objective of using Bayesian modeling is to get a posterior distribution of the parameters of the models, conditional on the data and the model itself. Bayesian inference uses Baye's rule to compute the posterior distribution of the parameters, as shown in Equation (3.1), where $p(\Theta|y)$ is the posterior distribution, Θ is a vector of parameters, y is the data, and p is a probability density function for continuous variables, or a probability mass function for discrete variables.

$$p(\Theta|y) = \frac{p(y|\Theta) \cdot p(\Theta)}{p(y)} \tag{3.1}$$

Equation (3.1) is commonly rewritten as Equation (3.2), where the likelihood is the probability density function or probability mass function expressed as a function of Θ. The prior is the probability distribution that is initially assigned to the parameters; and the marginal likelihood is a normalizing constant used to standardize the posterior distribution so that it sums up to 1. The unnormalized posterior distribution is proportional to the numerator of Equation (3.1), as shown in Equation (3.3).

$$\text{Posterior distribution} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Marginal Likelihood}} \tag{3.2}$$

$$\text{Unnormalized posterior distribution} \propto \text{Likelihood} \cdot \text{Prior} \tag{3.3}$$

In complex models, deriving the posterior distribution analytically becomes impossible, because computing the marginal likelihood involves integrating the numerator in Bayes rule (Equation 3.4), which is often mathematically intractable. As a result, Bayesian statistics rely on sampling algorithms that draw samples from the unnormalized posterior density, as the histogram of these samples approximate the target posterior distribution. Stan implements a variant of the No-U-Turn sampler (NUTS; Hoffman & Gelman, 2014), based on advanced Hamiltonian Monte Carlo algorithm (HMC), which is a type of Markov Chain Monte Carlo method.

$$p(\Theta|y) = \frac{p(y|\Theta) \cdot p(\Theta)}{\int_{\Theta} p(y|\Theta) \cdot p(\Theta) d\Theta} \tag{3.4}$$

The HMC algorithm uses the shape of the target posterior density, $(p(\theta|y)$ to inform the exploration of the parameter space. The algorithm uses Hamiltonian dynamics simulation followed by a Metropolis acceptance step (Stan Development Team, 2021b, chapter 15). Hamiltonian dynamics can be better understood by visualizing a particle of mass m sliding over a frictionless surface. Important variables to consider are the position of the particle, the momentum (ρ), the potential energy of the particle (which is proportional to the height of the surface at its position), and the kinetic energy of the particle. As explained by Neal (2011):

> If it encounters a rising slope, the puck's [particle's] momentum allows it to continue, with its kinetic energy decreasing and its potential energy increasing, until the kinetic energy (and hence p) is zero, at which point it will slide back down (with kinetic energy increasing and potential energy decreasing). (Neal, 2011, p.114)

In the context of sampling in the Bayesian framework, the position corresponds to a vector that contains the parameters of interest, Θ, and its potential energy $V(\Theta)$, is defined as the negative logarithm of the unnormalized posterior. The momentum variables are introduced artificially. Specificaly, in Stan, they are drawn from a multinormal distribution (Stan Development Team, 2021b).

The HMC algorithm work as follows: First, it starts with randomly-generated initial values for the set of parameters θ. For each iteration, a new state $\theta*$ is proposed based on the current position θ and a randomly drawn value of momentum p from a multinormal distribution. The transition to a new state $\theta*$ and a new momentum $p*$ are calculated using Hamiltonian equations, which are numerically approximated by discretizing time. Stan uses the leapfrog method to discretize time in L number of small time steps ϵ (ϵ is also known as the step-size parameter). Finally, a Metropolis acceptance step is performed. If the proposal is not accepted, there is no transition to the proposed state (θ^*, p^*) and the previous parameter value is used to initialize the next iteration (see Stan Development Team, 2021b, chapter 15). One problem with the HMC algorithm is that it is sensitive to two user-specified parameters: The step size (ϵ) and the number of leapfrog steps (L). For instance, a large L can lead to unnecessary computation, while a low L can lead to random-walk behavior (Hoffman & Gelman, 2014). Stan implements a variant of the NUTS algorithm (Hoffman & Gelman, 2014), which estimates these two parameters, thus removing the need to hand-tune HMC's parameters.[1]

The Stan syntax is based on C++. Stan models contain several blocks of code. The mandatory blocks include `functions` (for user-written functions), `data` (for declaring of the known variables, e.g., observations), `parameters` (for declaring the unknown variables, i.e., the model parameters), `model` (for specifying the prior distributions and the likelihood to define the posterior). The optional blocks include `transformed data` (to declare transformed data), `transformed parameters` (to declare transformed parameters), and `generated quantities` (to generate quantities from the estimated parameters in the model). The expression `target+=` is used to add terms to the unnormalized log posterior probablity, which is the same as multiplying a term in the numerator of the unnormalized posterior, as shown in Equation (3.3). Therefore, the expression `target+=` is normally used in the model block, to declare the prior distributions of the model parameters, as well as the likelihood.

[1] A detailed explanation of HMC and NUTS are beyond the scope of this dissertation. For the interested reader, some useful references are Betancourt (2018), Leimkuhler and Reich (2004), Neal (2011).

3.2 The activation-based model as a race of accumulators

In all the experiments modeled in this thesis, there are two dependent variables: Accuracy in a sentence-picture matching task, and reaction times or listening times. The reaction times come from the sentence-picture matching task, while the listening times come from self-paced listening tasks. The main assumptions linked to these dependent variables are that the accuracy reflects the interpretation of thematic roles in the sentence, and that the reaction times/listening times serve as a proxy for retrieval latencies. Consider sentence (13a), which is a subject relative clause, and sentence (13b), an object relative clause. These examples are similar to the relative clause conditions modeled in Chapter 4.

(13) a. The **dog**$_{\text{NP1}}$ who beat the cat$_{\text{NP2}}$ chases the rat.

b. The dog$_{\text{NP1}}$ who the **cat**$_{\text{NP2}}$ beat chases the rat.

According to cue-based retrieval, when encountering the verb *beat*, comprehenders start a retrieval process to retrieve the subject of *beat*. In (13a), the target of the dependency is the first noun phrase (NP1), whereas in (13b), the target is the second noun phrase (NP2), and NP1 is a distractor.

The main assumption in the activation-based model as implemented in this dissertation is that the process of accumulation of evidence can be equated to the process of activation accrual in the original Lewis and Vasishth (2005) model of sentence comprehension. In Lewis and Vasishth (2005), the probability of retrieval of an item and its retrieval latency are determined by its activation value. The activation of an item decreases as a function of interference and decay. In sentences with distance or interference manipulations, such as (13b), the target item would be less likely to be retrieved, and retrieval latencies would be slower, relative to a sentence like (13b), in which there is no distractor item. All the items modeled in this dissertation consist of a pair of sentences such as (13a) and (13b), where one sentence contains the experimental manipulation, and the other one serves as a baseline condition.

In our implementation of Lewis and Vasishth (2005) as a race of accumulators, which is based on the original implementation by Nicenboim and Vasishth (2018), there are two accumulators of evidence, each one corresponding to the retrieval of a noun phrase in a given sentence. One of these noun phrases is the target of the dependency, and the other one is the distractor. The race process between the two accumulators consists on sampling finishing times from two distributions with mean μ_{NP1}, and μ_{NP2}, respectively; and standard deviation σ. For each trial, the accumulator with the lower finishing time wins the race and its value becomes the estimated listening time/reaction time. This is exemplified graphically in Figure 3.1.

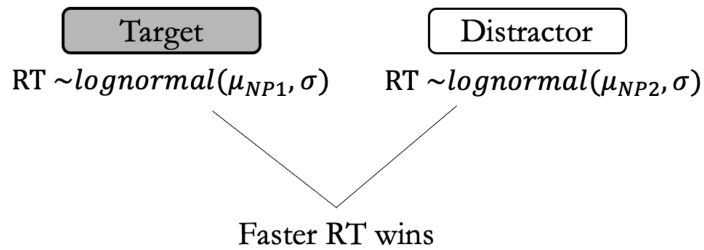

Figure 3.1: Graphical representation of the activation-based model as a race of accumulators.

The race process is written in a user-defined function. The function takes as arguments the answer and reaction time (RT) for each trial from the empirical data; as well as the estimated parameters μ_{NP1}, μ_{NP2} and σ. The answer for a given trial corresponds to the winner of the trial, i.e., the actual noun phrase that was retrieved. In the picture-selection task, accuracy selecting the right picture was coded as 1 (correct picture was selected) or 0 (incorrect picture was selected). Our assumption is that the picture selection task also reflects the retrieval process. Participants were shown two pictures. The target picture depicts depicts the meaning of the sentence that they hear. The foil picture depicts the action expressed by the verb in the sentence, but with reversed thematic roles. Therefore, if participants chose the target picture, they retrieved the target noun from memory as the agent of the action depicted in the picture, because that is what the target picture shows. If the incorrect picture was chosen, we assume that the distractor was retrieved as the agent of the action depicted in the picture. For example, in correct trials in object relative clauses, we assume that the NP2 was retrieved, and in incorrect trials, NP1 was retrieved. The variable *answer* maps, by trial, the accuracy with the retrieved noun phrase. This variable is coded in the data. The race function checks, for each trial, the *answer* variable. If the answer corresponds to NP1, the function adds to the log likelihood the probability that the RT_i is drawn from a lognormal distribution with mean μ_{NP1}, for the current parameter values. This is done by using the `lognormal_lpdf` function in Stan. The accumulator that does not win in trial i will have a slower (or higher) finishing time. The probability that the RT_i comes from a lognormal distribution with mean μ_{NP2} is added to the log likelihood using the complementary cumulative distribution (`lognormal_lccdf`) in Stan[2]. In trials in which the answer is NP2, the probability that the RT_i is drawn from a lognormal distribution with mean μ_{NP2} is

[2] This code builds on the original code from Nicenboim and Vasishth (2018) and from van het Nederend (2018).

added to the log likelihood using the lognormal probability distribution, and from μ_{NP1}, using the complementary cumulative distribution distribution.

Listing 3.1: Race function

```
real race(int answer, real RT, real accum_NP1, real accum_NP2,
    real accum_sigma){

    real log_lik;
    log_lik = 0;

    if(answer==1){
      log_lik += lognormal_lpdf(RT| accum_NP1, accum_sigma);
      log_lik += lognormal_lccdf(RT| accum_NP2, accum_sigma);
    }
    else {
      log_lik += lognormal_lpdf(RT| accum_NP2, accum_sigma);
      log_lik += lognormal_lccdf(RT| accum_NP1, accum_sigma);
    }
    return(log_lik);
  }
```

The means of the accumulators, μ_{NP1}, and μ_{NP2} include the fixed and random effects relevant for the structure of each experimental data-set. For each data-set, the specific hierarchical structure will be detailed in their corresponding chapters. The same number of fixed and random effects are estimated for μ_{NP1} and μ_{NP2}. Listing 3.2 shows the most complex model structure fitted in this dissertation, which corresponds to the data modeled in Chapter 6. In addition to the main effect of group and of the linguistic manipulations (rctype and num variables, these will be explained in Chapter 6), the main effects also include lexical decision times from a lexical decision task (LDT variable), proportions of gazes to the target picture in a sentence-picture matching task (fixations), and all relevant interactions between the different predictors.

Listing 3.2: Example of fixed and random effects added to the accumulators.

```
model {
 for (n in 1:N_obs) {
  real accum_NP1 = alpha[1] + u[subj[n],1] + w[item[n],1] +
                   group[n]*(beta[1]+w[item[n],3]) +
                   num[n]*(beta[3]+u[subj[n],3]) + group[n]*num[n]*beta[5] +
                   LDT[n]*beta[7] + group[n]*LDT[n]*beta[9] +
                   LDT[n]*rctype[n]*beta[11] +
                   group[n]*LDT[n]*rctype[n]*beta[13] +
                   fix[n]*beta[15] + fix[n]*group[n]*beta[17] +
```

```
                      fix[n]*rctype[n]*beta[19] +
                      group[n]*fix[n]*rctype[n]*beta[21] +
                      rctype[n]*(beta[23]+u[subj[n],5]) +
                      rctype[n]*group[n]*beta[25] + rctype[n]*num[n]*beta[27] +
                      rctype[n]*num[n]*group[n]*beta[29];

    real accum_NP2 = alpha[2] + u[subj[n],2] + w[item[n],2] +
                      group[n]*(beta[2]+w[item[n],4]) +
                      num[n]*(beta[4]+u[subj[n],4]) + group[n]*num[n]*beta[6] +
                      LDT[n]*beta[8] + group[n]*LDT[n]*beta[10] +
                      LDT[n]*rctype[n]*beta[12] +
                      group[n]*LDT[n]*rctype[n]*beta[14] +
                      fix[n]*beta[16] + fix[n]*group[n]*beta[18] +
                      fix[n]*rctype[n]*beta[20] +
                      group[n]*fix[n]*rctype[n]*beta[22] +
                      rctype[n]*(beta[24]+u[subj[n],6]) +
                      rctype[n]*group[n]*beta[26] + rctype[n]*num[n]*beta[28] +
                      rctype[n]*num[n]*group[n]*beta[30];

    real accum_sigma = sigma_0 + group[n]*beta[31];
  }
}
```

The model implements by-subject and by-item correlated varying intercept and varying slopes in the two accumulators. Given the complexity of the model, non-centered parametrization with the Cholesky decomposition was used for the random effects. This is a reparametrization often used in complex hierarchical models in Stan (e.g., see Nicenboim, Schad, & Vasishth, 2021, chapter 11). Uncorrelated vectors for subjects (z_u) and items (z_w) are sampled from a normal distribution with mean 0 and standard deviation 0.5. These vectors are multiplied by the Cholesky factor (L_u and L_w, respectively) in the transformed parameter block of code. These multiplications result in two respective matrices that contain vectors of correlated variables. By multiplying the matrices by the diagonal matrix τ_u and τ_w, respectively, using the `diag_pre_multiply` function, the vectors of correlated values are scaled back to the corresponding standard deviations. In Listing 3.3, n_u and n_w are arguments previously declared in the data block, which stand for the number of random effects for subjects and items, respectively.

Listing 3.3: Non-centered parametrization.

```
parameters{
  vector[7] beta;
  real alpha[2];
```

```
  real<lower=fabs(beta[7])> sigma;

  cholesky_factor_corr[n_u] L_u;
  cholesky_factor_corr[n_w] L_w;
  vector<lower=0>[n_u] tau_u;
  vector<lower=0>[n_w] tau_w;
  matrix[n_u,N_subj] z_u;
  matrix[n_w,N_item] z_w;
}

transformed parameters {
  matrix[N_subj, n_u] u;
  matrix[N_item, n_w] w;
  u = (diag_pre_multiply(tau_u, L_u) * z_u)';
␣␣w␣=␣(diag_pre_multiply(tau_w,␣L_w)␣*␣z_w)';
}
```

3.2.1 Priors

All the parameters (which are on the log scale because of the lognormal likelihood) have regularizing priors, shown in Equation (3.5). These priors allow for a wide range of values and down-weight extreme values. In the context of psycholinguistics, prior specification in hierarchical models is discussed in Sorensen, Hohenstein, and Vasishth (2016), Nicenboim and Vasishth (2016), Vasishth, Nicenboim, Beckman, Li, and Kong (2018), Schad, Betancourt, and Vasishth (2021), and in Chapter 6 from Nicenboim et al. (2021). The priors are shown graphically in Figure 3.2.

$$\begin{aligned}
\alpha_{1,2} &\sim normal(7.5, 0.6)\\
\beta_{1,\ldots,n} &\sim normal(0, 0.5)\\
\sigma &\sim normal_{+}(0, 0.5)\\
\tau_{u_{1,\ldots,n}} &\sim normal_{+}(0, 0.1)\\
\tau_{w_{1,\ldots,n}} &\sim normal_{+}(0, 0.1)\\
L_u &\sim LKJcorr(2)\\
L_w &\sim LKJcorr(2)\\
z_u &\sim normal(0, 0.5)\\
z_w &\sim normal(0, 0.5)
\end{aligned} \tag{3.5}$$

where the subscript n stands for the number of dimensions of the vector of parameters, the subscript u indexes subjects, and w indexes items. Accordingly, τ_u and τ_w are the sd of the by-subject and by-item random effects, and L_u and L_w stand for the Cholesky factor of the correlation matrices for the random effects. Finally, z_u and z_u are the random uncorrelated variables that are multiplied by the Cholesky factor (L_u and L_w, respectively) and by the diagonal matrices (τ_u and τ_w, respectively) in order to obtain the vectors of correlated random effects using a non-centered parametrization. In the prior for the residual sd (σ), the subscript $+$ in the normal distribution prior stands for a normal distribution truncated at 0 (reflecting the fact that standard deviations can never be less than 0).

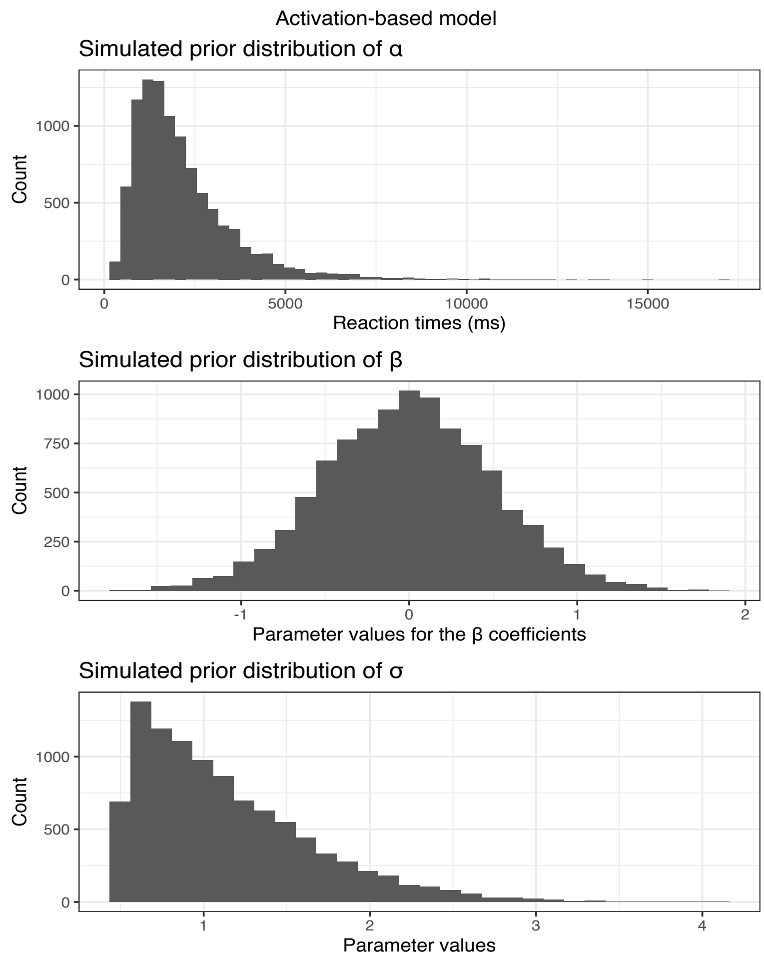

Figure 3.2: Simulated prior distributions for main parameters in activation-based model.

The prior on the intercepts of the accumulators, $\alpha_{1,2}$, assumes that a priori, the intercept is somewhere between $exp(7.5 - 0.6 \times 2) = 544$ ms and $exp(7.5 + 0.6 \times 2) \approx$ 6000 ms. The interpretation of the prior for the effect sizes, indexed by $\beta_{1,\ldots,n}$, depends on the intercept α. For instance, for an intercept of 7.5 log ms ($\approx$ 1800 ms), the prior for $\beta_{1,\ldots,n}$ would estimate a change in α in the range from $exp(7.5) \times exp(-1) = 665$

ms, to $exp(7.5) \times exp(1) = 4914$ ms. The effect size strongly changes depending on the intercept, so the prior allows for a wide range of effect sizes. The same prior was used for σ. Finally, the LKJ (Lewandowski, Kurowicka, & Joe, 2009) prior was used for the correlation of the random effects, with parameter 2, which disfavors extreme correlations like ± 1 (Carpenter et al., 2017).

3.2.2 Posterior predictive checks

In order to assess the adequacy of the model, posterior predictive checks are required. This model check consists on simulating data based on the posterior samples of the parameters in the model (Gelman et al., 2013). This is achieved by creating a user-generated function in Stan. Listing 3.4 shows the function corresponding to the modeling for relative clauses. The function first samples from the two accumulators. The if else statement works as follows: If the sampled value from accumulator 1 is lower than the sample from accumulator 2, and the trial corresponds to a subject relative, generate accuracy 1 (correct), and if the trial corresponds to an object relative, generate accuracy 0 (incorrect). Else if the sampled value from the accumulator 2 is faster than accumulator 1, generate accuracy 1 if trial corresponds to an object relative, and accuracy 0 if trial corresponds to a subject relative. This mapping works because NP1 is the target in subject relatives, and NP2 is the target in object relatives. The function outputs a vector which contains the generated RT and the generated accuracies.

Listing 3.4: Function to generate posterior predictive checks.

```
vector race_rng(real mu_1, real mu_2, real sigma, int rctype){
    vector[2] gen;
    real accum_NP1_RT = lognormal_rng(mu_1, sigma);
    real accum_NP2_RT = lognormal_rng(mu_2, sigma);

    if(accum_NP1_RT < accum_NP2_RT){
      gen[1] = accum_NP1_RT;
      if(rctype == -1){ //SR -1 in contrast coding
        gen[2] = 1;}
      else {
        gen[2] = 0;}
    }
    else {
      gen[1] = accum_NP2_RT;
      if(rctype == 1){ // OR +1 in contrast coding
        gen[2] = 1;
      }
      else {
        gen[2] = 0;}
```

```
    }
    return(gen);}
}
```

This function is used to generate simulated data from the posterior of the model parameters. The function is called in the `generated quantities` block, and the parameters `accum_NP1, accum_NP2` and `sigma` are passed as arguments. These parameters, as shown in Listing 3.2, include the full structure of random and fixed effects. So the posterior predictive checks are done with simulated data that include all the estimated parameters in the model.

3.2.3 Parameter recovery

One important test to check the validity of the model is to assess whether the model is able to recover its own parameters. This procedure involves (a) fitting the activation-based model to the empirical data, (b) extracting the estimated mean parameter values from this fit, (c) generating a relatively large amount of simulated data based on the mean estimated parameter values following the generative process assumed by the model, (d) fit the activation-based model to the simulated data, (e) check whether the mean estimated parameter values from the empirical data fall within the range of parameter values estimated by the model fit to simulated data. This procedure ensures that the model is working properly and that it can recover its own parameters. Failure to recover its own parameters could indicate that the model is misspecified. As this is a custom-built model whose complexity can quickly increase depending on the number of random and fixed effects, parameter recovery is a necessary step to ensure the validity of the model estimates. The steps of parameter recovery are detailed below.

First, a data-set of simulated data is created. In order to do this, a data-frame that contains the same variables than the empirical data is created in R. Listing 3.5 shows an example of this, corresponding to the data modeled in chapter 6. In order to ensure that the simulated data has a relatively large sample size, as a rule of thumb, the simulated data consisted of at least three times the amount of observations in the empirical data. For instance, in this case, the number of simulated trials, N, was 16,000. The variable `LDT` stands for lexical decision task. In the empirical data it corresponds to the mean lexical decision RT from a lexical decision task, averaged by participant. In order to generate simulated data in the same unit than it appears in the empirical data, simulated RT are sampled from a normal distribution and then scaled. There is only one LDT value per participant, so in the simulated data, 100 data points are sampled, as we assume that 100 is the number of subjects. Similarly, the variable `fixations`, which stands for the proportions of looks to the target picture, contains scaled simulated data from a uniform distribution with parameters 0 and 1. The rest of the code creates simulated data that mimics the design in the empirical data: Two conditions in the relative clause variable (subject vs. object relatives), two conditions in the number variable (match vs. mismatch), and two groups (controls vs. IWA). All of these variables are coded with sum contrasts: +1 and −1. The specific design of the empirical data will be explained in chapter 6. Here it is simply shown as an example for the recovery of the parameters. Finally, the columns corresponding to RT and answer are empty, as these will be filled later on with values generated from the generative process.

Listing 3.5: Simulated data.

```
N <- 16000
nsubj <- 100
LT_LDT <- rnorm(nsubj,1500,500)
LDT <- scale(LT_LDT)

ACT.data.sim <- data.frame(
        rctype = rep(c(rep(-1, 1000), rep(1, 1000)),8),
        num = rep(c(rep(-1, 500), rep(1, 500)),16),
        item = rep(1:10, 1600),
        subj = rep(1:nsubj, each=10, times=16),
        group = rep(c(rep(-1, 10), rep(1, 10)), 800),
        LDT = rep(LDT, each=10, times=16),
        fixations = scale(runif(N,0,1)),
        rt = rep(0, N),
        answer = rep(0, N))
```

To assess the recovery of the parameter estimates from the random effects, the estimates are extracted from the model fitted to the empirical data. In the code below, `ACT` contains all the parameter values from the model fitted to the empirical data[3]. After extracting the estimates for the Cholesky factor L_u, and the diagonal matrix τ_u, random uncorrelated variables are generated from a normal distribution with mean 0 and sd 0.5. The random variables are then multiplied by τ_u and L_u, which yields the final matrix of correlated adjustments for subjects. The same procedure is done for the by-item random adjustments.

Listing 3.6: Code for by-subject random effects.

```
# random uncorrelated variables
z_u1 <- rnorm(ACT.data.sim.nsubj, 0, .5)
```

[3] The code for simulating data builds on the original code from van het Nederend (2018).

```
z_u2 <- rnorm(ACT.data.sim.nsubj, 0, .5)
z_u3 <- rnorm(ACT.data.sim.nsubj, 0, .5)
z_u4 <- rnorm(ACT.data.sim.nsubj, 0, .5)
z_u <- matrix(c(z_u1, z_u2, z_u3,z_u4),
    ncol = ACT.data.sim.nsubj, byrow = T)

# Cholesky
L_u <- matrix(c(
ACT$`L_u[1,1]`,ACT$`L_u[2,1]`,ACT$`L_u[3,1]`,ACT$`L_u[4,1]`,
                0,ACT$`L_u[2,2]`,ACT$`L_u[3,2]`,ACT$`L_u[4,2]`,
                0,0,ACT$`L_u[3,3]`,ACT$`L_u[4,3]`,
                0,0,0,ACT$`L_u[4,4]`),nrow=4,ncol=4)

# Diagonal matrix Tau
Tau_u <- matrix(c(ACT$`tau_u[1]`,0,0,0,
                  0,ACT$`tau_u[2]`,0,0,
                  0,0,ACT$`tau_u[3]`,0,
                  0,0,0,ACT$`tau_u[4]`),nrow=4,ncol=4)

# correlated adjustments
ACT.u <- Tau_u %*% L_u %*% z_u %>% t
```

Once the random effects have been generated, the next step is to produce simulated data that follows the generative process assumed by the model. This involves simulating μ_{NP1}, μ_{NP2} and σ using the parameter estimates from the model fit to the real data. Listing 3.7 shows an example of how this is achieved for μ_{NP1}. The same procedure is done for μ_{NP2} and σ. The code below is enclosed in a for loop with range `1:nrow(ACT.data.sim)`,which ensures that the data generating process is done for each simulated trial. Then, one data point is sampled from a lognormal distribution with location μ_{NP1} and μ_{NP2}, respectively, and sd σ. Finally, the variables *answer*, *RT* and *accuracy* are generated and incorporated to the data-frame of simulated data as shown in Listing 3.8. The RT from the winner accumulator (i.e., the lower RT) becomes the estimated RT for a given trial. If the sampled value of μ_{NP1} is lower, then 1 is assigned to the variable `answer`, otherwise 2 is assigned. This mapping of 1 and 2 follows the same mapping that is coded in the empirical data, 1 corresponds to NP1 and 2 corresponds to NP2.

Listing 3.7: Simulated data for μ_{NP1}.

```
mu_NP1 <- ACT$`alpha[1]`+ACT.u[sub,1]+ACT.w[item,1]+
          grp*(ACT$`beta[1]`+ACT.w[item,3]) +
          num*(ACT$`beta[3]`+ACT.u[sub,3]) +
          grp*num*ACT$`beta[5]`+ LDT*ACT$`beta[7]` +
          grp*LDT*ACT$`beta[9]` + LDT*rctype*ACT$`beta[11]` +
```

```
    grp*LDT*rctype*ACT$`beta[13]` + fix*ACT$`beta[15]`+
    grp*fix*ACT$`beta[17]`+ fix*rctype*ACT$`beta[19]` +
    grp*fix*rctype*ACT$`beta[21]` +
    rctype*(ACT$`beta[23]`+ ACT.u[sub,5]) +
    grp*rctype*ACT$`beta[25]` +
    rctype*num*ACT$`beta[27]` +
    rctype*num*grp*ACT$`beta[29]`
```

Listing 3.8: Simulated data for the activation-based model.

```
rt1 <- rlnorm(1, mu_NP1, sigma);
rt2 <- rlnorm(1, mu_NP2, sigma);

ACT.data.sim$answer[n]  <- ifelse(rt1 < rt2, 1, 2)
ACT.data.sim$rt[n]      <- ifelse(rt1 < rt2, rt1, rt2)
```

Once the dataframe is complete with the RT and responses simulated using the estimated parameters form the empirical data and the generative process assumed by the model, the simulated data is fitted to the model. The estimates of the model fit with simulated data can be used to check whether the estimates of the model fit with empirical data can be recovered. This can be done graphically, by plotting the discrepancies between the posterior means of the model parameters and the values extracted from the original fit with real data, for those parameters. An example is shown in Figure 3.3[4], which corresponds to the model of relative clauses fitted in Chapter 6. The black dots indicate the difference between the posterior means and the extracted values from the model fitted to real data. The lines show the 95% credible interval of the difference (i.e., the 2.5th and 97.5th percentiles of the posterior draws). For the model to correctly recover its parameters, all or nearly all of the intervals should include zero. Simulated data and parameter recovery were carried out for each model fitted in this dissertation, in order to check that the models were adequate. These plots are available at the online repositories linked at the appendix.

3.3 The direct-access model as a mixture process

We implement the direct-access model as a Bayesian two-component mixture model, following Nicenboim and Vasishth (2018). These are the main assumptions of the original direct-access model that we implement:

1. Retrieval cues enable direct-access to elements from memory. Retrieval times μ_{da} are assumed to be independent from sentence complexity (i.e.,

[4] This plot is adapted from the original approach presented in Furr (2017), available here: https://mc-stan.org/users/documentation/case-studies/rasch_and_2pl.html

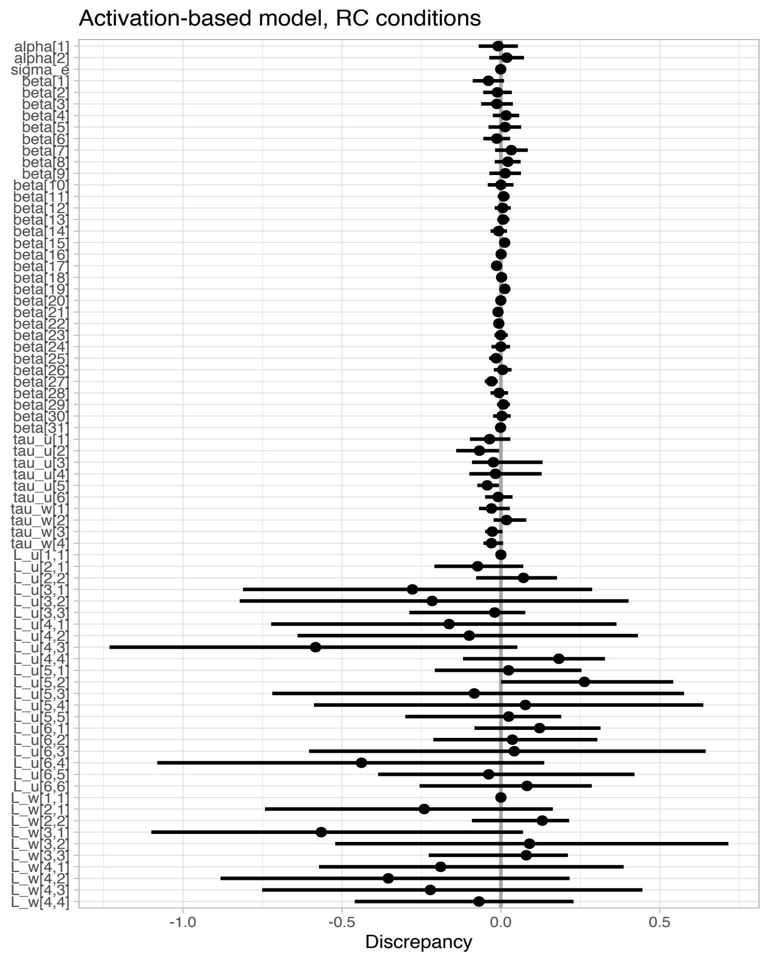

Figure 3.3: Discrepancies RC.

increased distance between the target and the retrieval site, similarity-based interference).

2. Sentence complexity can lower the availability of the target NP in memory. Low availability can lead to the retrieval of the distractor, which is known as a misretrieval. After an initial misretrieval, if a backtracking process is initiated, the target will be eventually retrieved from memory.

3. Backtracking requires extra processing time that is independent from the retrieval time μ_{da}. The backtracking time (δ) needs to be added to μ_{da} in order to account for the processing time in trials that include backtracking.

In order to implement the assumptions of the direct-access model of memory retrieval, we assume that the response selection has a certain probability of retrieval of the target (θ) which is affected by the probability of backtracking (P_b). The implementation assumes that responses (*accuracy*) follow a Bernoulli distribution, in which additional probability mass is added to the distribution of the correct responses if backtracking is performed. The probability of retrieving the target or the distractor for trial i is shown in Equations (3.6) and (3.7), respectively. In the equations, *target* and *distractor* stand for the selection of the target or the distractor NP from memory, θ is the probability of initial retrieval of the target; and P_b is the probability of backtracking.

$$P(answer_i = target|\theta, P_b) = \theta + (1-\theta)\cdot P_b \tag{3.6}$$

$$P(answer_i = distractor|\theta, P_b) = (1-\theta)\cdot(1-P_b) \tag{3.7}$$

Response times are drawn from a lognormal distribution. If the response for trial i is correct (accuracy = 1), the RT for trial i come from a mixture distribution that is estimated as shown in Equation (4.6). The mixture distribution reflects the two processes that can lead to a correct response: Direct access of the target, or initial misretrieval followed by backtracking.

$$RT_{correct} \sim \begin{cases} lognormal(\mu, \sigma), & \text{initial retrieval succeeds} \\ lognormal(\mu+\delta, \sigma), & \text{initial retrieval fails + backtracking} \end{cases} \tag{3.8}$$

$$RT_{incorrect} \sim lognormal(\mu, \sigma) \tag{3.9}$$

where $RT_{correct}$ stand for the RT in correct trials, μ is the location of the distribution of RT, σ is the standard deviation of the lognormal distribution, and δ is the processing time needed for backtracking. The RT corresponding to incorrect trials ($RT_{incorrect}$) are assumed to come from the same distribution than the correct trials without backtracking, as detailed in Equation (3.9).

As shown in Equation (3.6), the probability of a correct answer is yielded by $\theta + P_b\cdot(1-\theta)$. It follows that:

1. The first component of the mixture model shown in Equation (3.8) corresponds to correct answers that come from an initial correct retrieval, which occur with probability

$$\frac{\theta}{\theta + P_b\cdot(1-\theta)} \tag{3.10}$$

2. The second component in Equation (3.8) corresponds to correct answers that come from an initial incorrect retrieval followed by backtracking, which occur with probability

$$\frac{P_b \cdot (1-\theta)}{\theta + P_b \cdot (1-\theta)} \tag{3.11}$$

3. Finally, the response for trial i is incorrect (accuracy $= 0$), with probability

$$(1-\theta) \cdot (1-P_b) \tag{3.12}$$

These probabilities are coded in Stan, in log space, in a user-generated function that is called in the model block. This function is shown in Listing 3.9[5]. In log space, the multiplications become sums, the divisions become subtractions, and the `log_sum_exp` function is used to add two terms. The second part in the function implements the mixture model as shown in Equations 3.8 and 3.9.

Listing 3.9: Function defining the likelihood for the direct-access model.

```
real direct_access(int accuracy, real RT, real theta, real P_b,
    real mu, real delta, real sigma){

    // Equation 3.7
 real log_p_answer_correct = log_sum_exp(log(theta), log(P_b) + log1m(theta));
    // Equation 3.11
 real log_p_answer_correct_direct_access = log(theta) - log_p_answer_correct;
    // Equation 3.12
 real log_p_answer_correct_backtrack = log(P_b) + log1m(theta)
                                        - log_p_answer_correct;
    // Equation 3.13
 real log_p_answer_incorrect = log1m(theta) + log1m(P_b);

 if(accuracy==1) {
  return(log_p_answer_correct +
    log_sum_exp(
     log_p_answer_correct_direct_access + lognormal_lpdf(RT| mu, sigma),
     log_p_answer_correct_backtrack + lognormal_lpdf(RT| mu + delta, sigma)));
    } else {
     return(log_p_answer_incorrect + lognormal_lpdf(RT| mu, sigma));
    }
  }
```

In order to model the effect of sentence complexity (i.e., similarity-based interference, decay) and the differences in performance between the individuals with aphasia

[5]This code builds on the original code from Nicenboim and Vasishth (2018) and from van het Nederend (2018).

and control participants, the fixed effects *condition*, *group*, and *condition* $\times$ *group* were added to the parameters of interest. Below there is a general description of the main effects that were added to all of the models, based on the assumptions about sentence processing deficits in aphasia.

- μ only has a main effect of group, as μ is a proxy for retrieval times, and in the direct-access model, retrieval times are not dependent on sentence complexity. However, it is assumed that there could be a difference in retrieval times between the two groups.
- θ has main effects for group, condition, and group $\times$ condition. This parameter stands for the initial probability of correct retrieval, and is expected to reflect the effects of sentence complexity and group.
- P_b and δ have a main effect of group because IWA are expected to show an impairment in the process of backtracking.
- σ also has an adjustment by group because the retrieval process is expected to be noisier for IWA.

The model structure becomes more complex with more conditions (e.g., the four relative-clauses conditions in Chapter 6) and with more predictors (e.g., fixations and data from a lexical decision task in 5). In these cases, the adjustments were added to μ and/or θ, according to the specific theoretical predictions for that model. A more thorough explanation of the different parameters in relation to the processing deficits in aphasia will be given when presenting the specific assumptions for each modeled data-set. As an example, Listing 3.10 shows the structure of fixed and random effects implemented for the data modeled in Chapter 6.

Listing 3.10: .

```
real mu = mu_0 + u[subj[i],1]+ w[item[i],1] + group[i]*beta[1];
real theta =  inv_logit(alpha + u[subj[i],2] + w[item[i],2] +
                        LDT[i]*beta[2] + group[i]*LDT[i]*beta[3] +
                        group[i]*(beta[4] + w[item[i],3]) +
                        rctype[i]*(beta[5]+ u[subj[i],3]) +
                        rctype[i]*group[i]*beta[6] +
                        num[i]*(beta[7]+ u[subj[i],4]) +
                        num[i]*group[i]*beta[8] +
                        num[i]*rctype[i]*beta[9] +
                        num[i]*rctype[i]*group[i]*beta[10]+
                        fix[i]*beta[11] +
                        fix[i]*group[i]*beta[12] +
                        fix[i]*rctype[i]*beta[13] +
```

```
                    fix[i]*rctype[i]*group[i]*beta[14]);

real P_b = inv_logit(gamma + u[subj[i],5] + group[i]*beta[15]);
real delta = delta_0 + group[i]*beta[16];
real sigma = sigma_0 + group[i]*beta[17];
```

3.3.1 Priors

In the direct-access model, the same regularizing priors were used for the location and the sd of the distribution of RT (μ and σ) and for the effect sizes ($\beta_{1,\ldots,n}$) as in the activation-based model. Similarly, the same non-centered parametrization for the random effects was implemented, and the same priors were used for the random effects. These have been discussed in Section 3.2.1. The rest of the priors are shown in Equation 3.13 and can be seen graphically in Figure 3.4.

$$\begin{aligned}
\alpha &\sim normal(1, 0.5) \\
\beta_{1,\ldots,n} &\sim normal(0, 0.5) \\
\mu_0 &\sim normal(7.5, 0.6) \\
\gamma &\sim normal(-1, 0.5) \\
\delta_0 &\sim normal(0, 1) \\
\sigma_0 &\sim normal(0, 0.5)
\end{aligned} \tag{3.13}$$

The prior for θ and for γ (which is the intercept for P_b) encode some general information taken from a previous study. Nicenboim and Vasishth (2018), in their multivariate implementation of the direct-access model, estimated that the probability of retrieval of the target at the first retrieval attempt (equivalent to θ in the present model) was around 70% (95% CrI: [60, 75]). Accordingly, our prior assumes that, a priori, the mean probability of retrieval of the target is $\approx$ 70%, and that it varies between 50% to $\approx$ 90%.[6] The estimate for the probability of backtracking in Nicenboim and Vasishth (2018) was 48% (95% CrI: [40, 55]). Our prior for the intercept of P_b assumes, a priori, that the probability of backtracking is within the range 12% to 50%, with mean $\approx$ 30%. Our prior covers the estimates in Nicenboim and Vasishth (2018), but it generally assumes a lower probability of backtracking. The reason is that while Nicenboim and Vasishth (2018) modeled data from unimpaired controls, we also model data from IWA, who are assumed to perform reanalysis less often (e.g. Mack, Wei, Gutierrez, & Thompson, 2016).

[6] This can be calculated in R by using the following code: `mean = plogis(1), lower = plogis(1-1), upper = plogis(1+1)`

Finally, the effect of δ in ms depends on μ, as this is an multiplicative effect in log scale. The prior for δ accounts for our uncertainty about this parameter, because as far as we know, no previous study has measured the cost of backtracking in IWA. We hypothesize that δ should be higher for IWA than for controls, whose estimate in Nicenboim and Vasishth (2018) was 120 ms (95% CrI: [30, 55]). However we have no assumption regarding how much higher. Accordingly, we chose a regularizing prior that allows for a wide range of parameter values. For instance, with $\mu = 7$ (log ms), the range of the cost of backtracking in ms would oscillate between $exp(7{+}0.1){-}exp(7)$ $= 115$ ms and $exp(7+2) - exp(7) = 7000$ ms.

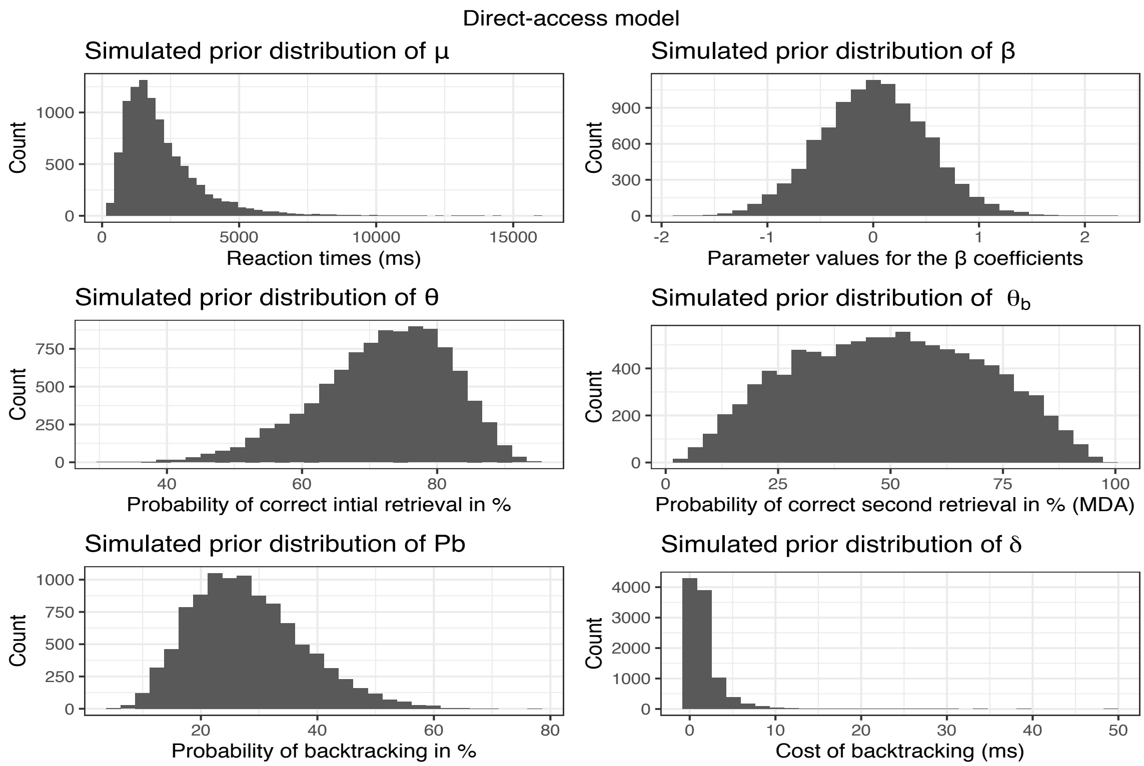

Figure 3.4: Simulated prior distributions for main parameters in the direct-access and modified direct-access model.

3.3.2 Posterior predictive checks

As in the activation-based model, in order to generate simulated data from the model parameters, a user-defined function was created. The function, shown in Listing 3.11, takes all the main parameters of the direct-access model as arguments. First, the initial accuracy is sampled from a Bernoulli distribution with probability θ. If the initial accuracy is 0 (incorrect), a value is sampled from a Bernoulli distribution with probability P_b. If backtracking was estimated to happen (i.e., the sampled value was

1), the estimated accuracy for that trial is 1 (correct), else it is 0. If `backtracking` = 1, the RT from that trial are sampled from $lognormal(\mu+\delta, \sigma)$, otherwise, the RT are sampled from $lognormal(\mu, \sigma)$. The function is called in the `generated quantities` block of code, and outputs a vector of generated accuracies and RT. This vector can be used to visually inspect the simulated data and assess whether the empirical data falls under the range of predicted data by the model.

Listing 3.11: Function to generate posterior predictive checks in the direct-access model.

```
vector direct_access_rng(real theta, real P_b, real mu,
    real delta, real sigma){
    int init_acc;
    int backtrack;
    vector[2] gen;

    init_acc = bernoulli_rng(theta);
    backtrack = 0;
    if (init_acc!=1) backtrack = bernoulli_rng(P_b);
    // Change the answer to 1 if backtracking = 1:
    gen[2] = backtrack ? 1 : init_acc;
    { real mu_rng; // adds delta if there is backtracking:
      mu_rng = mu + (backtrack ? delta : 0);
      gen[1] = lognormal_rng(mu_rng, sigma);
    }
    return(gen);
}
```

3.3.3 Parameter recovery

For the assessment of model adequacy, the recovery of the parameters of the model was carried out in the same fashion than in the activation-based model. That is, (a) fitting the model with the empirical data, (b) extracting the model parameters and using them to create simulated data, (c) fitting the model to the simulate data, and (d) evaluating the discrepancies between the mean parameter values in the model fitted to the empirical data against the model estimates in the model fitted to simulated data. Listing 3.12 exemplifies how the parameters are generated using data extracted from the direct-access model fit to the empirical data. The random effects were simulated in the exact same way as in the activation-based model. What changes is the way in which the accuracies and RT are generated, as the direct-access model assumes a different generative process. The last part of the code shows how the data is sampled from the parameters. This is analogous to the `race_rng` function created in Stan to perform posterior predictive checks.

Listing 3.12: Function to generate posterior predictive checks in the direct-access model.

```
mu <- DA$mu_0  + DA.u[sub,1] + DA.w[item,1] + grp*DA$`beta[1]`

theta <- inv.logit(DA$alpha + DA.u[sub,2] + DA.w[item,2] +
                   LDT*DA$`beta[2]` + grp*LDT*DA$`beta[3]`+
                   grp*(DA$`beta[4]`+ DA.w[item,3]) +
                   rctype*(DA$`beta[5]`+DA.u[sub,3])+
                   rctype*grp*DA$`beta[6]`+
                   num*(DA$`beta[7]`+DA.u[sub,4]) +
                   grp*num*DA$`beta[8]`+ grp*rctype*DA$`beta[9]`+
                   grp*num*rctype*DA$`beta[10]`+
                   fix*DA$`beta[11]`+ fix*grp*DA$`beta[12]`);

P_b <- inv.logit(DA$gamma + DA.u[sub,5] + grp*DA$`beta[13]`);
delta <- DA$delta_0 + grp*DA$`beta[14]`;
sigma <- DA$sigma_0 + grp*DA$`beta[15]`;


 init_acc <- rbinom(1,1,theta)
  if(init_acc){
    DA.data.sim$acc[i] <- init_acc
    DA.data.sim$rt[i] <- rlnorm(1, mu, sigma);
  } else {
    backtrack <- rbinom(1,1,P_b)
    DA.data.sim$acc[i] <- backtrack
      if(backtrack){
       DA.data.sim$rt[i] <- rlnorm(1, mu + delta, sigma);
     } else {
       DA.data.sim$rt[i] <- rlnorm(1, mu, sigma);
     }
  }
```

3.4 The modified direct-access model

The main difference between the direct-access model and our proposed modification is that in the original model, backtracking always leads to the retrieval of the target, whereas in the augmented model, backtracking can lead to the retrieval of the target or to a misretrieval. The rationale for this modification will be explained in Chapter 5. Here we detail the computational details of the implementation of the modified direct-access model. The probability of retrieving the target or the distractor for trial i is shown in Equations (3.14) and (3.15), respectively. In these equations, θ is the probability of retrieval of the target, P_b is the probability of backtracking, and θ_b is the probability of retrieving the target after backtracking.

$$P(answer_i = target|\theta, \theta_b, P_b) = \theta + [(1-\theta) \cdot P_b \cdot \theta_b] \tag{3.14}$$

$$\begin{aligned} P(answer_i = distractor|\theta, \theta_b, P_b) = & [(1-\theta) \cdot (1-P_b)] + \\ & [(1-\theta) \cdot P_b \cdot (1-\theta_b)] \end{aligned} \tag{3.15}$$

From these equations, it follows that:

1. Correct responses that come from an initial correct retrieval occur with probability
$$\frac{\theta}{\theta + (1-\theta) \cdot P_b \cdot \theta_b} \tag{3.16}$$
2. Correct responses that come from an initial incorrect retrieval followed by backtracking occur with probability
$$\frac{P_b \cdot (1-\theta) \cdot \theta_b}{\theta + (1-\theta) \cdot P_b \cdot \theta_b} \tag{3.17}$$
3. Incorrect responses that come from an initial incorrect retrieval occur with probability
$$\frac{(1-\theta) \cdot (1-P_b)}{[(1-\theta) \cdot (1-P_b)] + [(1-\theta) \cdot (P_b) \cdot (1-\theta_b)]} \tag{3.18}$$
4. Incorrect responses that come from an initial incorrect retrieval followed by failed backtracking occur with probability
$$\frac{(1-\theta) \cdot P_b \cdot (1-\theta_b)}{[(1-\theta) \cdot (1-P_b)] + [(1-\theta) \cdot (P_b) \cdot (1-\theta_b)]} \tag{3.19}$$

The function that implements these probabilities in Stan (in log space), is shown in Listing 3.13.

Listing 3.13: Likelihood for the modified direct-access model

```
    // CORRECT responses
    //  Equation 3.15
log_p_answer_correct = log_sum_exp(log(theta), log1m(theta) +
                                   log(P_b) + log(theta_b));
    //  Equation 3.17
```

```
log_p_answer_correct_direct_access = log(theta) - log_p_answer_correct;
    //  Equation 3.18
log_p_answer_correct_reanalysis = log1m(theta) + log(P_b) + log(theta_b) -
                                                  log_p_answer_correct;


    // INCORRECT responses
    // Equation 3.16
log_p_answer_incorrect = log_sum_exp(log1m(theta) + log1m(P_b), log1m(theta) +
                                                  log(P_b) + log1m(theta_b));
    // Equation 3.19
log_p_answer_incorrect_direct_access = log1m(theta) + log1m(P_b) -
                                                  log_p_answer_incorrect;
    // Equation 3.20
log_p_answer_incorrect_reanalysis = log1m(theta) + log(P_b) + log1m(theta_b) -
log_p_answer_incorrect;
```

This function is called in the model block (code shown in Listing 3.14), after generating the model parameters, which as in the direct-access model, include both fixed and random effects. The extra parameter θ_b has a varing intercept by subject, and a main effect of group, as shown in Equation (3.20).

$$\theta_b = \alpha_b + u_{\alpha_b} + \beta \cdot group \tag{3.20}$$

Listing 3.14: Modified-direct access model function in the model block.

```
for (n in 1:N_obs) {
    target += m_direct_access(accuracy[i], RT[i], theta,
                              P_b, theta_b, mu, delta, sigma_e);
}
```

3.4.1 Priors

The same priors as in the direct-access model were used. The prior for the intercept (α_b) of the extra parameter θ_b is shown in Equation (3.21). This is a regularizing prior that assumes that the mean intercept for the retrieval of the target after backtracking is within the range of 12% to 90%.

$$\alpha_b \sim normal(0, 1) \tag{3.21}$$

3.5 Cross-validation

For model comparisons, k-fold cross-validation is the main method used in this dissertation. Cross-validation is a standard procedure in machine learning, and it quantifies the ability of a given model to predict a set of held-out data, i.e., unseen data, based on the training set of observed data (Vehtari, Gelman, & Gabry, 2017). In order to achieve this, a given dataset with N observations is split in a number of (balanced) k subsets. One of the subsets is held out, and the remaining $k-1$ subsets are used as training set (i.e., the model is fit to these $k-1$ subsets). The posterior distributions of the resulting model are used to compute predictive accuracy on the held-out subset. This procedure is repeated k times, so that all the subsets have been held-out. The expected log pointwise predictive density, $\widehat{elpd}$, is calculated as a measure of predictive accuracy. $\widehat{elpd}$ is a measure of the deviation between the held-out and the predicted values. This deviation is summed across the k subsets, so that all the data from a given dataset is held out. Equation (3.22), taken from Gelman, Hwang, and Vehtari (2014, p. 1004) shows how to calculate $\widehat{elpd}$ for a given k, where N is the number of observations, S is the number of posterior draws, and $p(y_i|\theta^s)$ is the probability of obtaining a data point y_i given the parameter draws Θ^s.

$$\widehat{elpd} = \sum_{i=1}^{N} log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\theta^S)\right) \tag{3.22}$$

Models can be compared by computing the difference in $\widehat{elpd}$, with higher $\widehat{elpd}$ indicating better predictive fit. Because $\widehat{elpd}$ is an estimate, the standard error of the difference in $\widehat{elpd}$ can be calculated, as shown in Equation (3.23). The SE has the standard frequentist interpretation: $\widehat{\Delta elpd} \pm 2 \times$ SE gives a 95% confidence interval. Therefore, if the difference in $\widehat{elpd}$ between the models is larger than $2 \times SE$, we conclude that it is decisive.

$$se(\widehat{elpd}_{M1} - \widehat{elpd}_{M2} = \sqrt{N \cdot Var(\widehat{elpd}_{M1,i} - \widehat{elpd}_{M2,i})} \tag{3.23}$$

where M1 and M2 stand for model 1 and model 2, and N is number of data points; and $\widehat{elpd}_{M1,i}$ and $\widehat{elpd}_{M2,i}$ represent a vector that contains the $\widehat{elpd}$ for each data point, for model *M1* and *M2* respectively.

We set K to 10, which is standardly done (Vehtari et al., 2017). For the implementation of the cross-validation, we followed the approach detailed in Vasishth, Nicenboim, Chopin, and Ryder (2017). The data are pseudo-randomly partitioned in 10 folds in a balanced fashion, i.e., making sure that subjects were similarly represented within each fold. A list of 10 data-frames is created. In each dataframe, a new column called heldout is added. Some observations are marked as held-out data (heldout == 1) and some are marked as training data (heldout == 0). Using a for loop, the model is be fitted to each one of these 10 data-frames, and the log likelihood of the observations marked as training data is extracted, in order to calculate the predictive accuracy on the held-out data.

The heldout variable is declared within the data block in the Stan programs. The following pieces of code are also added: First, in the transformed parameter block, a vector of length equal to the number of data points is created. The log likelihood of each data point is then stored in this vector. In the model block, the code shown in Listing 3.15 is added. This code ensures that only the likelihood corresponding to the data points selected as training data is evaluated.

Listing 3.15: Code added to all models in the model block for the cross-validation.

```
for (n in 1:N_obs) {
  if(heldout[n]==0){target += log_lik[n];}
  }
```

3.6 Summary

In this chapter, the implementation of the activation-based, the direct-access, and the modified-direct access model of retrieval have been detailed. The next chapter presents the evaluation of the activation-based model and the direct-access model in subject and object relative clauses in English.

Chapter 4

Modeling subject and object relative clauses in English

The contents of this chapter are published in:

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

4.1 Introduction

Within the cue-based retrieval framework, two distinct models of sentence processing have been proposed: The Lewis and Vasishth (2005) model of sentence processing (LV05), and the direct-access model (DA) developed by McElree (2000). The two models share the assumption that retrieval is driven by a cue-based mechanism, and both predict that a distractor disrupts the retrieval of the target when the retrieval cues match the distractor and the target. Despite these similarities, the two models assume fundamentally different underlying processes for the access of representations in memory. In the LV05 model, retrieval time for an item depends on the activation of the item in memory, with reduced discriminability of an item leading to lower activation and therefore longer retrieval times. By contrast, in the direct-access model, retrieval time is assumed to be constant, and reduced discriminability only affects the probability of correct retrieval of the target.

Nicenboim and Vasishth (2018) were the first to formally implement these two competing models and compare their relative predictive performance. Using self-paced reading data from a number interference experiment in German (Nicenboim, Vasishth, Engelmann, & Suckow, 2018), Nicenboim and Vasishth implemented the LV05 and DA models in a Bayesian framework. They showed that (a) the direct-access model has better predictive performance than the activation-based model, but (b) the activation-based model yields a comparable performance to the direct-access model when the variance of the retrieval times is allowed to be different for correct and incorrect retrievals. The computational implementations of the two competing models of retrieval make it possible, for the first time, to investigate their relative performance using a broader range of experimental data.

Both LV05 and DA are meant to account for retrieval processes in sentence comprehension in unimpaired populations. An open question is whether these models, which have until now only been investigated in connection with unimpaired processing, can also characterize retrieval difficulty in impaired populations. That is, can the models account for impaired processing through parametric variation? And if they can, what do the changes in the parameters tell us about the impairments? In this chapter, we focus on an important and under-studied problem, the underlying nature of retrieval difficulty in individuals with aphasia.

One question we seek to answer is: Given the two competing models of retrieval processes, which one better characterizes processing difficulty in IWA? As data, we use the largest dataset currently in existence on sentence comprehension in IWA. This dataset, reported in Caplan et al. (2015), provides listening times and picture-selection accuracies from IWA and matched unimpaired controls. The full dataset involves a range of syntactic constructions and methods, but in this chapter, we focus on self-paced listening data on the subject vs. object relative clause construction, which is a very well-studied construction in psycholinguistics.

This chapter is structured as follows. We begin by reviewing prior work on modeling retrieval processes in aphasia. Next, we present the data, our implementation of the activation-based and the direct-access models, the results of the model comparisons, and a Bayes factors analysis.

4.1.1 Modeling retrieval processes in aphasia

There are several theories about why language processing deficits arise in IWA. In this dissertation we focus on processing deficit theories that can be implemented within the framework of cue-based theory and that are of relevance for our modeling work.[1] These theories were reviewed in Chapter 2, but to remind the reader, we focus on the following accounts: *delayed lexical access* (Ferrill et al., 2012), *slow syntax* (Burkhardt et al., 2008), *resource reduction* (Caplan, 2012), and *intermittent deficiencies* (Caplan et al., 2015).

[1]For alternative accounts see, for example, Grodzinsky (1995), Grillo (2009), or Engel et al. (2018). A complete summary of the theories of processing deficits in aphasia can be found in Caplan et al. (2015).

The *delayed lexical access* theory claims that lexical access is delayed in IWA, which can cause a slowdown in the formation of a syntactic dependency. Evidence supporting this theory comes from a series of cross-modal lexical priming studies, which combine a listening comprehension and a lexical decision task. Love et al. (2008) and Ferrill et al. (2012) (inter alia) found that IWA showed slower lexical activation relative to controls. Some cross-modal lexical priming studies have also revealed that IWA build syntactic dependencies at a slower-than-normal speed. This has been taken as support for the *slow syntax* theory (Burkhardt et al., 2008; Burkhardt et al., 2003), which posits that a slowdown in syntactic structure building can cause a delayed interpretation or a failure to interpret the sentence. Under this account, the impairment is at the level of syntactic structure formation.

Caplan et al. (2007) and Caplan et al. (2015) present online and offline data that support the hypothesis that IWA have a deficit in the resources used in parsing, what they refer to as *resource reduction* (Caplan, 2012). Complex sentences demand more resources, such as a higher memory load or attention, and therefore, IWA are more likely to misinterpret complex sentences. Finally, Caplan et al. (2013) argue that in addition to a *resource reduction*, IWA may exhibit intermittent breakdowns in the parsing system, a theory known as *intermittent deficiencies*.

Some of these accounts have been implemented in the framework of LV05. Patil et al. (2016) developed several LV05-based models that implement theories of processing deficits in aphasia. They found that IWA's processing was better characterized by a model that combined the implementation of slowed processing (understood as a "pathological slowdown in the processing system") and intermittent deficiencies, relative to models that included only one of these deficits. Building on the conclusions of Patil et al. (2016), Mätzig et al. (2018) investigated variability among IWA by implementing slowed processing, intermittent deficiencies, and resource reduction within the LV05 model. The range of parameters estimated for IWA showed a broad variability, whereas the parameters for control participants were closer to the default parameters of the original LV05 model, and displayed a smaller range of variability. These results imply that IWA are very variable in the extent and nature of their deficits along these three hypothesized dimensions (slowed processing, intermittent deficiencies, and resource reduction). The broader conclusion here is that deficits may lie on a continuum, and along different dimensions.

Although Patil et al. (2016) only modeled data from 7 IWA, and Mätzig et al. (2018) modeled offline measures (accuracies), both studies showed that LV05 can account for IWA's behavior by modifying specific parameters that can be mapped onto theoretically-informed assumptions. By doing so, they derived quantitative predictions under the assumptions of theories of deficits in aphasia. It remains to be tested whether the LV05 model can account for the different hypothesized deficits in a larger dataset with online measures.

However, there exists another competing model of retrieval processes, the direct-access model. The crucial difference between these two models is that they assume different underlying mechanisms for the access of items in memory. Yet, the relative predictive performance of the activation model and of the direct-access model has never been compared using data from both unimpaired and impaired populations. By comparing these two models' predictions with data from IWA, we aim to investigate the following questions:

(a) Can the direct-access mechanism of retrieval also account for sentence processing in IWA?

(b) How do the different parameters of these two models relate to theories of processing deficits in IWA?

(c) Which model provides a better fit to data from IWA and controls?

Investigating these questions would provide new insight into the nature of the dependency completion process in impaired and unimpaired populations. The Caplan et al. dataset makes such a model comparison possible. Below, we begin by revisiting the characteristics of the subset of the Caplan et al. dataset that we model in this chapter.

4.2 The Caplan et al. dataset: Self-paced listening times in relative clauses

The empirical data we consider here consist of listening times and picture-selection accuracies from 33 IWA and 46 controls matched by age and years of education. The original dataset reported in Caplan et al. (2015) included 56 IWA, but we discarded data from 8 IWA because they were in the early post-acute phase (less than four months post-stroke), and from 15 other individuals who had been classified as IWA but showed no symptoms of aphasia in the Boston Diagnostic Aphasia Exam (Goodglass, Kaplan, & Barresi, 2001).

Out of the 11 sentence types in the dataset, we selected the subject relative (SR) and object relative (OR) constructions (see examples 14a and 14b). This choice was motivated by the fact that relative clauses (RCs) have been extensively studied in psycholinguistics, and a great deal is known about relative clause processing. In English and many other languages, ORs have been uniformly found to be more difficult

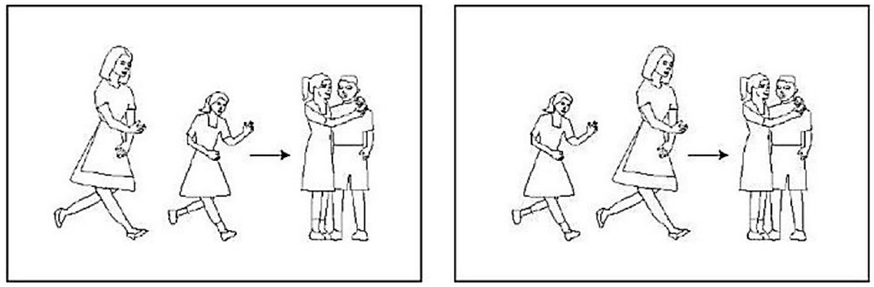

Figure 4.1: Example of the images shown in the picture-selection task. In the subject relative condition, the picture on the right is the target, whereas the picture on the left is the foil. In the object relative condition, the picture on the left is the target, and the one on the right is the foil.

to process than SRs (Grodner & Gibson, 2005). Moreover, IWA are known to experience difficulties in the comprehension of object relative clauses (Caramazza & Zurif, 1976; Hanne et al., 2011), especially when the thematic roles of the nouns can be reversed, as in the sentences shown below.

(14) a. **Subject Relative (SR):** The girl who chased the mother hugged the boy.

b. **Object Relative (OR):** The girl who the mother chased hugged the boy.

In the experiment reported by Caplan et al. (2015), participants listened to sentences presented word by word, and pressed a computer key whenever they were ready to hear the next word. This yielded an online measure of comprehension: Listening times (LT) per segment, in milliseconds. At the end of the sentence, participants had to choose which of two pictures displayed on the screen matched the meaning of the sentence they had just heard. This choice yielded accuracy data (correct/incorrect response). An example of the pictures shown in the picture-selection task is displayed in Figure 4.1. These pictures correspond to the sentences (14a) and (14b).

Of the twenty items corresponding to the SR and OR conditions in Caplan et al. (2015), we only used items 11-20 for our data analysis and modeling. The modeling is limited to these items because it was only in these items that the pictures in the picture-selection task tested the participant's understanding of the meaning of the verb inside the relative clause (e.g., who chased whom in 14a and 14b). For cue-based retrieval theory, in relative clauses, the retrieval of the agent of the action expressed by the verb within the relative clause is the first and key retrieval event (Lewis & Vasishth, 2005).

In English, the verb of the subordinate clause (*chased* in 14a and 14b) does not appear in the same position in subject and object relative clauses, and therefore the listening times corresponding to the verb region are not directly comparable. To make the two sentences comparable, we followed the procedure in Traxler, Williams, Blozis, and Morris (2005) and added up the listening times of the noun phrase (*the mother*) and the verb (*chased*) inside the subject/object relative clause. Trials with listening times shorter than 200 ms were discarded (around 2% of the data).

In the following section we present descriptive statistics and a Bayesian analysis of the data used for modeling. We analyze the data using the Bayesian framework because this allows us to quantify uncertainty about the estimates of interest (e.g., the difference in listening times for subject and object relative clauses). Our statistical inferences are based on 95% credible intervals and means of the estimates; the credible intervals show the range over which plausible values of the parameter lie with 95% probability, given the data and the model.

4.3 Bayesian analysis of the Caplan et al 2015 relative clause data

The mean accuracy for controls and IWA across the two conditions are shown in Figure 4.2. For controls, accuracy is above 90% in both conditions, whereas for IWA accuracy in SRs is 75%, and 63% in ORs. Figure 4.3 shows the mean listening times across conditions and groups. IWA are slower than controls in both conditions. For both IWA and controls, responses in the OR condition are slower relative to responses in the SR condition.

We fit a Bayesian hierarchical model with a lognormal likelihood to the listening times and a Bayesian logistic mixed model to the accuracy data. The analyses were carried out with correct and incorrect trials pooled. We used R (R Core Team, 2020) and the package brms (Bürkner, 2017), which is a front-end for Stan (Carpenter et al., 2017). For both models, the factors *group* (controls/IWA), *condition* (SR/OR), and their interaction were fit as fixed effects. These factors were sum-coded (Schad, Vasishth, Hohenstein, & Kliegl, 2020): SR were coded as -1 and OR as +1; controls as -1 and IWA as +1. Random intercepts by subjects and items were included, a slope by item was added to the group effect, and a slope by subject was added to the effect of condition. The varying intercepts and slopes were allowed to be correlated.

We used so-called regularizing priors, which allow a broad range of parameter values but disallow implausible (or impossible) values. The priors for the model of the accuracies, listed in Equation (4.1), are on the logit scale, whereas the priors for the listening times model, listed in Equation (4.2), are on the log scale. For the correlation matrix of the random effects, we used the so-called LKJ(2) prior

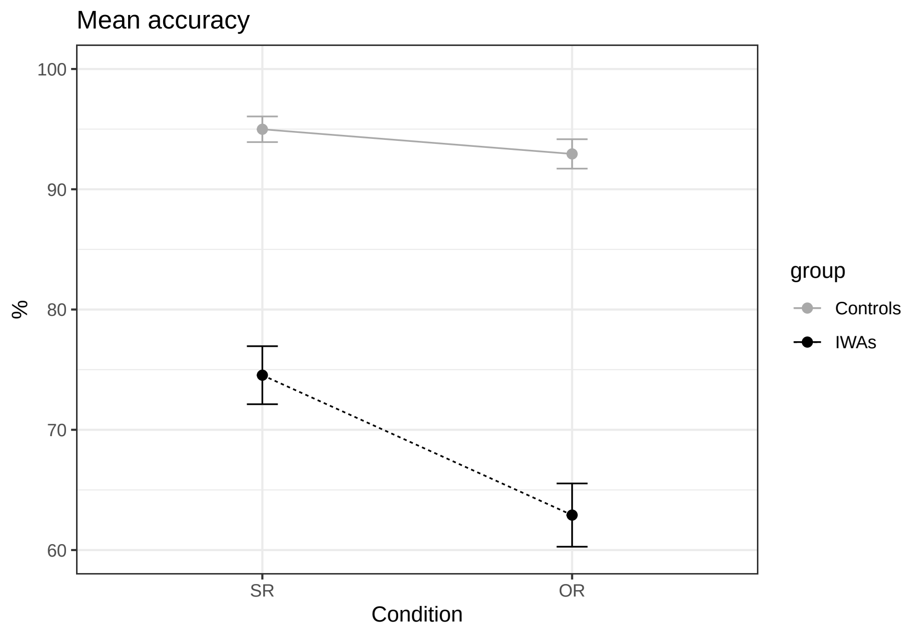

Figure 4.2: Mean accuracy across conditions and groups. Error bars show the standard error of the mean.

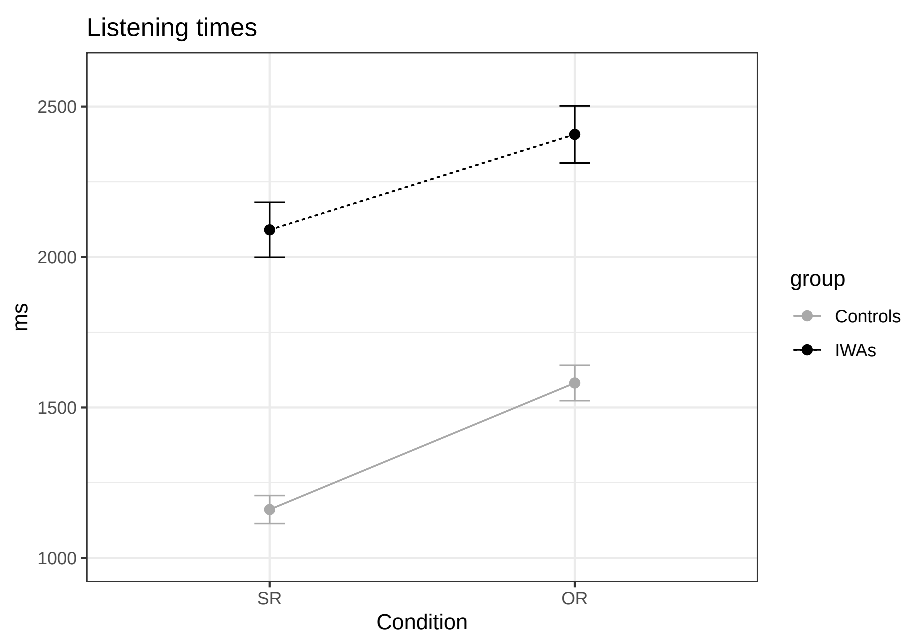

Figure 4.3: Mean listening times across conditions and groups. The listening times correspond to the sum of the listening times for the verb and noun phrase of the relative clause. Error bars show the standard error of the mean.

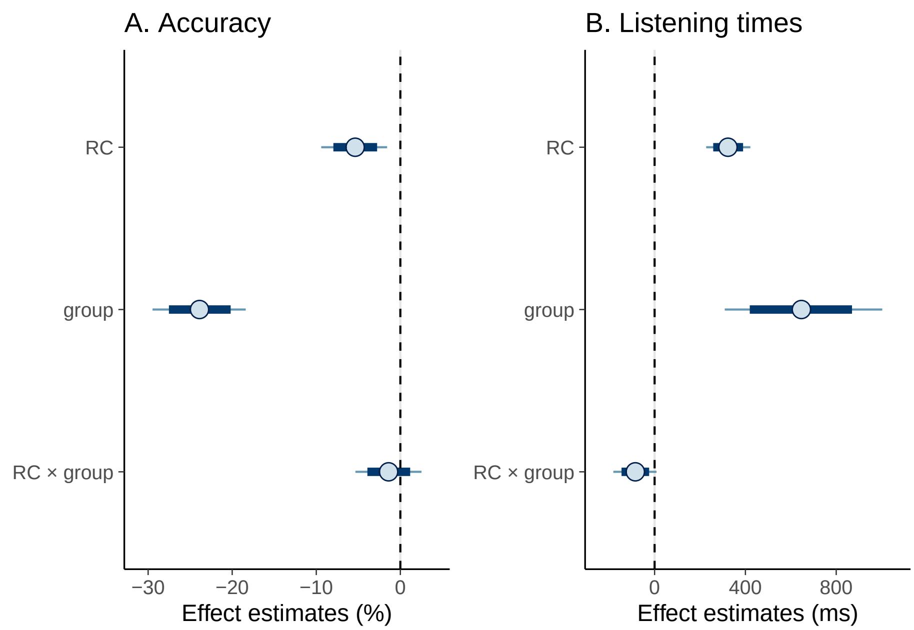

Figure 4.4: Posterior probability distributions of the different effect sizes for the effect of group (controls/IWA), condition (SR/OR), and their interaction. The dot corresponds to the mean of the distribution, the thick lines are 80% credible intervals, and the thin lines show 95% credible intervals. The dashed line stands for an effect size of zero.

(Lewandowski et al., 2009). Priors are explained in detail in Chapter 3. The models were fit with four chains and 2000 iterations, of which 1000 were warm-up iterations.

$$\begin{aligned} \alpha &\sim normal(0, 1) \\ \beta_{1,\ldots,3} &\sim normal(0, 0.5) \\ \sigma &\sim normal_+(0, 0.5) \end{aligned} \tag{4.1}$$

$$\begin{aligned} \alpha &\sim normal(7.5, 0.6) \\ \beta_{1,\ldots,3} &\sim normal(0, 0.5) \\ \sigma &\sim normal_+(0, 0.5) \end{aligned} \tag{4.2}$$

Figure 4.4 shows the posterior distributions of the parameters of interest. In a Bayesian model, the posterior distribution indicates the most likely parameter values given the data and the model. We report the mean estimate for each effect of interest, as well as their corresponding 95% credible interval (CrI). This interval represents the range over which we are 95% certain that the effect lies, given the data and the model.

Figure 4.4A shows the posterior distributions of the fixed effects for the analysis of the accuracy data. The data show an effect of group and condition: The estimated effect for group is of -24 % CrI: [-29, -18], indicating that IWA have more incorrect responses than controls. The effect of condition, -5 % CrI: [-9, -2] suggests that more incorrect responses are given in the OR condition. No indication for an interaction is seen, -1 % CrI: [-5, 3].

In listening times, large effects for group and condition were found: ORs yield longer listening times (effect of condition: 323 ms CrI: [227, 422]), and IWA are slower than controls (effect of group: 647 ms CrI: [309, 1003]). The interaction (-85 ms CrI: [-182, 9]) suggests that the effect of condition could be stronger for controls, but since the CrI overlaps with 0, strong conclusions cannot be drawn from this estimate.

Having summarized the inferences that can be made from the data, we now turn to a description of the two models, and the models' evaluation and comparisons.

4.4 The activation-based model

In cognitive psychology, response selection in simple choices is often modeled using accumulation of evidence (Heathcote & Love, 2012; Ratcliff, 1978). Evidence accumulation models assume that when facing a speeded decision, people accumulate noisy samples of information about the different choices that are available, until they have enough evidence to choose one of them (Forstmann, Ratcliff, & Wagenmakers, 2016).

Language processing can be seen as a similar process: When listening to a sentence, the comprehender samples evidence from the linguistic input that unfolds over time. Once the retrieval site is encountered, comprehenders have to retrieve an item from memory. Nicenboim and Vasishth (2018) argued that the retrieval process assumed in LV05 is conceptually similar to a race model (Rouder, Province, Morey, Gomez, & Heathcote, 2015; Usher & McClelland, 2001), in which each choice is represented with an accumulator of evidence. The speed of the process of sampling evidence in a race of accumulators can be equated to the activation in LV05: the item in memory with the faster rate of accumulation (equivalent to the higher activation in LV05) will be the item retrieved, and the rate of accumulation will determine the latency of the retrieval.

In the Caplan et al. (2015) data, the listening times at the relative clause verb and the second noun phrase serve as a measure of the speed of accumulation of evidence for the retrieval. Because there are two possible interpretations (subject or object relative clause), we assume that there are two accumulators racing against each other. For instance, consider again the object relative clause (14b), repeated here for convenience as (15):

(15) The girl who the mother chased hugged the boy.

When the comprehender reaches the verb *chased*, they need to retrieve a subject that matches the verb. If the comprehender understands the sentence correctly, they should have retrieved *mother* as the subject of the verb. An alternative possibility is that they accidentally misretrieve *girl* as the subject of the verb. Under these assumptions, the model has two accumulators: One accumulates evidence for the retrieval of the target (which corresponds to the correct OR interpretation in this example), and the other one accumulates evidence for the retrieval of the distractor *girl* (which corresponds to the incorrect SR interpretation in this example). The accumulator that finishes faster represents the interpretation chosen. We also assume that, when selecting one of the pictures during the picture-selection task, participants are choosing the interpretation that corresponds to the chunk retrieved from memory at *chased* (i.e., *mother* or *girl* in 15).

Implementation of the activation-based model

Following Nicenboim and Vasishth (2018), the activation-based model is implemented as a Bayesian lognormal race of accumulators. The Bayesian framework was chosen for two reasons. First, because modern probabilistic programming languages like Stan (Carpenter et al., 2017) make it possible to flexibly define any assumed generative process while including taking individual differences into account. Second, the Bayesian approach to parameter estimation allows the researcher to directly take the uncertainty of the estimates into account (Lee & Wagenmakers, 2014).

The model was implemented in Stan. For each trial i, the finishing times FT for the interpretation of a sentence as SR or OR are sampled from two lognormal distributions with scale σ, see Equation (4.3).[2] The noise component (σ) is assumed to be different for controls and IWA.[3] The accumulator with the faster (i.e., lower) FT will represent the winning interpretation, and its sampled value will become the estimated listening time for that particular trial i, as shown in Equation (4.4).

[2] What is meant here by interpretation of a sentence as SR is that the first noun phrase has been retrieved (i.e., the first noun phrase is interpreted as the agent) vs. the interpretation of a sentence as OR, where the second noun phrase is retrieved as the agent.

[3] We also fit a model with different variances for correct and incorrect responses, as introduced in Nicenboim and Vasishth (2018). However, the quantitative difference in predictive performance between the model with a single variance and the model with two variances was negligible. Both models show a comparable quantitative fit to the data. Here, we report the model with a single variance for correct and incorrect responses.

$$
\begin{aligned}
&\text{SR accumulator} \\
&\qquad FT_{SR_i} \sim lognormal(\mu_{SR}, \sigma) \\
&\text{OR accumulator} \\
&\qquad FT_{OR_i} \sim lognormal(\mu_{OR}, \sigma)
\end{aligned} \tag{4.3}
$$

$$
LT_i = min(FT_{SR_i}, FT_{OR_i}) \tag{4.4}
$$

The complete hierarchical model for the two accumulators is presented in Equation (4.5). The terms u and w are the by-participant and by-item adjustments to the fixed effects terms; these are the familiar varying intercepts and slopes in linear mixed models (Bates, Maechler, Bolker, & Walker, 2015, 1). All the parameters had regularizing priors, detailed in Chapter 3.

The level labeled *group* had contrast coding -1 for controls, and $+1$ for IWA; and the level labeled *relative clause type* (rc_{type}) was coded such that subject relatives were represented as -1 and object relatives as $+1$.

$$
\begin{aligned}
&\text{SR accumulator} \\
&\qquad \mu_{SR} = \alpha_1 + u_{\alpha_1} + w_{\alpha_1} + \\
&\qquad (\beta_1 + w_{\beta_1}) \cdot group + (\beta_3 + u_{\beta_3}) \cdot rc_{type} + \\
&\qquad \beta_5 \cdot group \cdot rc_{type} \\
&\text{OR accumulator} \\
&\qquad \mu_{OR} = \alpha_2 + u_{\alpha_2} + w_{\alpha_2} + \\
&\qquad (\beta_2 + w_{\beta_2}) \cdot group + (\beta_4 + u_{\beta_4}) \cdot rc_{type} + \\
&\qquad \beta_6 \cdot group \cdot rc_{type} \\
&\text{Noise parameter} \\
&\qquad \sigma = \sigma_0 + \beta_7 \cdot group
\end{aligned} \tag{4.5}
$$

The fixed effects β have the following interpretations:

- $\beta_1, \beta_3, \beta_5$ are the effects of group, RC type, and the group $\times$ RC type interaction respectively, in the accumulator for the subject relative interpretation.

- $\beta_2, \beta_4, \beta_6$ are the effects of group, RC type, and the group $\times$ RC type interaction, respectively, in the accumulator for the object relative interpretation.

- β_7 is the effect of group in the σ parameter.

Of interest in this model are the distributions of finishing times in the SR and OR accumulators, in the SR and OR conditions, and in the different population groups (controls vs. IWA). These are generated in ms once the posterior distributions of all the parameters in the model are estimated. The finishing times for each one of the accumulators in each condition and for each group are estimated taking into account the above-mentioned terms $\beta_{1,\ldots,7}$ and the adjustments by item and by participant listed in Equation (4.5).

Predictions

In the activation-based model the parameter σ and the finishing times of the accumulators have a theoretically meaningful interpretation. We expect these parameters to show different patterns across groups. The different σ reflect the assumption that for IWA, the rate of accumulation of evidence can be noisier. A larger estimated σ for IWA would be consistent with the *intermittent deficiencies* theory (Caplan et al., 2007), which claims that there are intermittent breakdowns in the parsing system of IWA. However, the effects of crucial interest are on the finishing times: When the mean finishing time of the incorrect interpretation is similar to the finishing time of the correct interpretation, misretrievals become more likely. We therefore expect that compared to controls, IWA should have more similar mean finishing times in the two accumulators; controls should have a bigger difference between the mean finishing times of the two accumulators. We also expect both accumulators to be slower for IWA than for controls because IWA may need more time than controls to retrieve items from memory and to build the dependency. Such a slowdown could be due to a *lexical access deficit* (Love et al., 2008) and/or to *slow syntax* (Burkhardt et al., 2008).

4.5 The direct-access model

The direct-access model (McElree, 2000) assumes that items (i.e., traces of words or phrases, such as *the girl*) in memory are accessed via a content-based, direct-access mechanism. That is, the cues set at the retrieval site enable direct access to matching items in memory. The retrieval process is subject to interference and decay: Increasing distance between the target and the retrieval site, or competing items in the sentence can lower the quality of the representation of the target item in memory. In the direct-access model, the probability of retaining a memory representation at the retrieval site is known as the *availability* of a given item. Crucially, proponents of the direct-access model argue that interference and decay have an impact on the availability of items in memory, but not on retrieval latencies. That is, whereas the probability of retrieving an item decreases as a function of the complexity of a sentence, complexity does not affect retrieval times. The direct-access model has been developed and tested within the speed-accuracy tradeoff paradigm (SAT) by McElree and colleagues (A. E. Martin & McElree, 2008, 2011; McElree et al., 2003), inter alia. They consistently found that the asymptote of the SAT function (which assesses successful retrieval of the target and/or quality of the retrieved representation) decreased as a function of sentence complexity. By contrast, the intercept and the rate of the SAT function (which assess processing speed) did not show a significant effect of complexity. Based on these findings, McElree and colleagues argue that interference and/or decay affect the probability of retrieving the target, but not the retrieval speed. In addition, it is assumed that low availability can cause a failure in parsing or the retrieval of a distractor item. On some trials, this initial failure could be followed by a reanalysis process (A. E. Martin & McElree, 2008; McElree, 1993; McElree et al., 2003; Van Dyke & McElree, 2011).

Implementation of the direct-access model

We follow Nicenboim and Vasishth (2018) by implementing the direct-access model as a two-component Bayesian mixture model. The key assumptions of the direct-access model are thus that retrieval cues enable direct access to the item's memory representation at the retrieval site, and that the retrieval of an item takes an average time t_{da}. Differences in availability can lead to an initial incorrect retrieval of the distractor item. McElree and colleagues assume that on a certain proportion of trials, after a failure in parsing, comprehenders could engage in a "costly reanalysis process" (A. E. Martin & McElree, 2008). We formalize this assumption with two main parameters: P_b, which is the probability of backtracking (what McElree and colleagues call *re-analysis*), and δ, which is the extra time needed for backtracking. This extra time is independent of the retrieval time t_{da}. Notice that these two parameters (P_b and δ) are not part of the SAT paradigm, and constitute an implementation of McElree and colleagues' assumption of reanalysis.

The parameter θ is the probability of correctly retrieving an item on the first retrieval attempt. This probability is allowed to vary across conditions, as it is assumed by McElree and colleagues (McElree et al., 2003) that sentence complexity can have an impact on the availability of the items, and therefore on their retrieval probability. If an initial misretrieval or failure in parsing occurs at the retrieval site, a backtracking process is initiated with probability P_b that, by assumption, always results in correct retrieval of the target (McElree, 1993).

There are four fixed-effects parameters that have to be estimated in this model. For the parameter θ we define varying intercepts by participants and by items, and varying slopes for the effect of relative clause type (by participants) and group type (by items). The parameter μ represents the estimated log mean listening times at the critical region. Since the direct-access model assumes that the retrieval time of an item takes on average t_{da} log ms and is not affected by sentence complexity, relative clause type was not included as a fixed effect for the parameter μ. However, we assume that IWA, given their impairment, could have a higher μ compared to controls and therefore add a main effect of group. That is, we assume that IWA may differ in the average time they need to process the critical region relative to controls. Notice that t_{da} is a latent variable that is part of μ, since we cannot directly compute t_{da} from the observed listening times. The probability of backtracking, P_b is also not assumed to vary across conditions, and thus only has an adjustment for group and a varying intercept by-participants because we assume that IWA could have a different P_b relative to controls. The parameter δ is the cost of backtracking, that is, the time (in log ms) that the backtracking process takes, and has an adjustment for group. The standard deviation σ also has a main effect of group.

$$LT \sim \begin{cases} lognormal(\mu, \sigma), & \text{retrieval succeeds, probability } \theta \\ lognormal(\mu + \delta, \sigma), & \text{retrieval fails initially, probability } 1 - \theta \end{cases} \tag{4.6}$$

As in the activation-based model, the terms u and w are the by-participant and by-item adjustments to the fixed effects terms. As with the activation-based model, all the parameters (which are on the logit scale for probabilities and on the log scale for listening times) have regularizing priors, detailed in Chapter 3. The level group had contrast coding -1 for controls, and $+1$ for IWA; and relative clause type rc_{type} was coded -1 for subject relative clauses and $+1$ for object relatives. The complete hierarchical model for all the parameters is shown in Equations (4.6) and (4.7). The mixture process is shown in Equation (4.6), and the parameters are defined in Equation (4.7).

$$
\begin{aligned}
\mu &= \mu_0 + u_{\mu 0} + w_{\mu 0} + \beta_1 \cdot group \\
\theta &= \alpha + u_\alpha + w_\alpha + (\beta_2 + u_{\beta_2}) \cdot rc_{type} + \\
&(\beta_3 + w_{\beta_3}) \cdot group + \beta_4 \cdot group \cdot rc_{type} \\
P_b &= \gamma + u_\gamma + \beta_5 \cdot group \\
\delta &= \delta_0 + \beta_6 \cdot group \\
\sigma &= \sigma_0 + \beta_7 \cdot group
\end{aligned}
\tag{4.7}
$$

The fixed effects β have the following interpretations:

- β_1 is the effect of group on the average time needed to listen to the critical region;
- $\beta_2, \beta_3, \beta_4$ are the effects of RC, group, and the group $\times$ RC interaction, respectively, on the probability of a first correct retrieval;
- β_5 and β_6 are the effect of group on the probability of backtracking and on the estimated backtracking time, respectively;
- β_7 is the effect of group on σ.

Consider the three possible scenarios according to the direct-access model.

Case (i): The target is retrieved through a direct-access mechanism based on the cues set at the retrieval site, with probability θ. In this case, listening times are assumed to be drawn from a lognormal distribution with mean μ and standard deviation σ: $LT \sim lognormal(\mu, \sigma)$.

Case (ii): The distractor is initially retrieved, but backtracking leads to the target being retrieved, with probability $(1-\theta) \cdot P_b$. Once θ (the probability of initial correct retrieval) has been estimated, $(1-\theta)$ yields the probability of an initial incorrect retrieval. The probability of backtracking is assumed to be independent of θ. Thus, multiplying P_b with $(1-\theta)$ yields the probability of correctly retrieving the target after an initial misretrieval and subsequent backtracking. In this case, the listening times are drawn from a lognormal distribution with mean $\mu + \delta$, which is the cost of backtracking, and standard deviation σ: $LT \sim lognormal(\mu + \delta, \sigma)$.

Case (iii): The distractor is initially retrieved and there is no backtracking, with probability $(1-\theta)\cdot(1-P_b)$. In this case, we multiply the probability that the first retrieval is incorrect with the probability that there is no backtracking. Here, the listening times are drawn from a lognormal distribution with mean μ and standard deviation σ: $LT \sim lognormal(\mu, \sigma)$, and a misretrieval is predicted.

Notice that incorrect answers without backtracking in case (iii) are expected to have similar listening times to correct answers without backtracking, case (i), whereas in case (ii), longer listening times should be observed due to the extra time needed for backtracking. As such, in this model, the distribution of listening times associated with correct responses is a mixture of initially retrieved targets (i), and initial misretrievals plus backtracking (ii).

Predictions

The parameters θ, μ, P_b, δ and σ have a group adjustment because they are expected to differ between controls and IWA. We present here a short theoretical explanation of the interpretation of these parameters.

We expect a lower estimate of the probability of correct initial retrieval, θ, for IWA, in object relative clauses. This would be in line with *resource reduction*. Complex sentences are assumed to require more processing resources, because additional linguistic operations need to be carried out and more material has to be kept in working memory (Caplan, 2012). This suggests that IWA should show a lower probability of initial correct retrieval in ORs relative to SRs. The different μ for controls and IWA reflect the assumption that IWA may need more time for parsing. This assumption can be linked to *slowed processing* theories, which would explain the slowdown in terms of lexical access (Love et al., 2008) or syntactic processing (Burkhardt et al., 2008). We expect IWA to have a lower probability of backtracking: If the model predicts IWA to backtrack, but not as often as controls, this could also be in line with the *resource reduction* hypothesis (Caplan, 2012). In unimpaired sentence comprehension, the DA model assumes that backtracking is a mechanism used on a certain proportion of trials when the initial interpretation of the sentence fails. If IWA show a lower probability of backtracking, this could mean that even though they can backtrack, they do not do it as often as controls because the mechanism is disrupted. Alternatively, the P_b parameter could also be linked to *intermittent deficiencies*, because the process of backtracking could be intermittently disrupted. In addition, we expect the cost of backtracking, δ to be higher for IWA. This would reflect delayed syntactic processing (Burkhardt et al., 2008). Finally, a larger σ would imply more noise in the retrieval mechanism for IWA. This would be consistent with the *intermittent deficiency* hypothesis (Caplan et al., 2007) that postulates that IWA suffer from intermittent reductions in the resources used in parsing.

4.6 Results

4.6.1 Results of the activation-based model

We used the *rstan* package (Stan Development Team, 2021a) to fit the models, with three chains, 6000 iterations and a warm-up of 3000.[4] The chains were plotted and visually inspected for convergence. An additional metric of convergence is the so-called Rhat statistic (the ratio of between-to-within chain variance); when the sampler has converged, the Rhat statistic is close to 1 (Gelman et al., 2013). We checked that Rhats were always near 1. Two tuning parameters, delta and the tree depth,[5] were adapted when necessary for achieving convergence. Following Gelman et al. (2013), we also made sure that the parameters of the model could be recovered using simulated data (see the online supplementary materials, available at https://osf.io/srfpm/).

The activation-based model assumes that for each trial, listening times are drawn from the two accumulators, and the accumulator with the fastest listening time wins the race. The two distributions of finishing times (that is, the finishing time of each one of the accumulators for each trial) can be plotted against each other, so as to assess the precise predictions of the model. For example, Figure 4.5 shows the distribution of finishing times for the correct and the incorrect interpretation for each of the two groups, and across the two conditions. Figures 4.5a and 4.5b display the accumulators for controls, while 4.5c and 4.5d stand for IWA's accumulators.

Figure 4.5a displays the distribution of finishing times associated with the accumulator for the correct interpretation (SR) in dark gray, and for the incorrect interpretation (here OR) in light gray, for controls. The distribution for SR is clearly faster: The mean of the finishing times for the SR accumulator is 1204 ms, whereas the mean finishing time for the OR accumulator is around 4000ms. In Figure 4.5b, finishing times for the correct interpretation (OR, in light gray) are faster on average (1655 ms) than the finishing times for the incorrect interpretation (SR, in dark gray, 4647 ms). Therefore, Figures 4.5a and 4.5b indicate that controls tend to choose the correct interpretation, since the distributions associated to the correct interpretations have faster finishing times.

[4] The code for both the activation-based and the direct-access models is available at https://bit.ly/3lda7Qj

[5] The adjustment of these tuning parameters (adapt_delta, max_treedepth) leads to the whole posterior distribution of the parameters being correctly explored by the Hamiltonian Monte Carlo algorithm used in Stan. See the Stan manual, or the short guide on warnings for more information (https://mc-stan.org/misc/warnings).

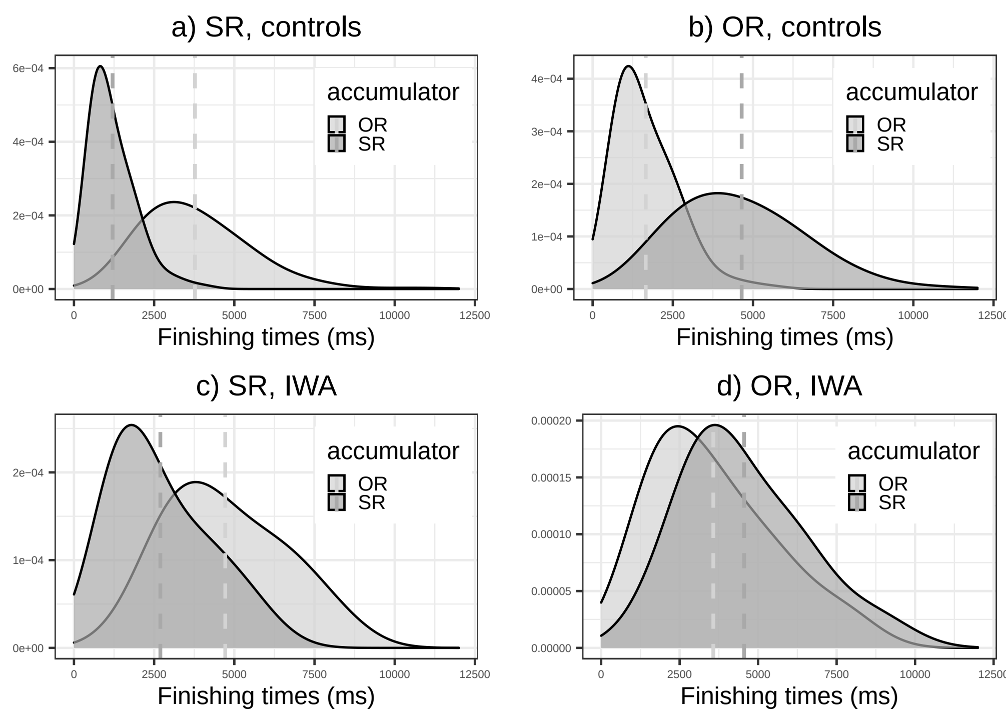

Figure 4.5: Accumulators of evidence. The figure presents the distribution of finishing times associated with each accumulator in the activation-based model, across groups and conditions. The x-axis stands for finishing times (in ms). The dashed lines represent the mean finishing time for the object relative clause interpretation (in light grey) and the subject relative clause intepretation (in dark grey).

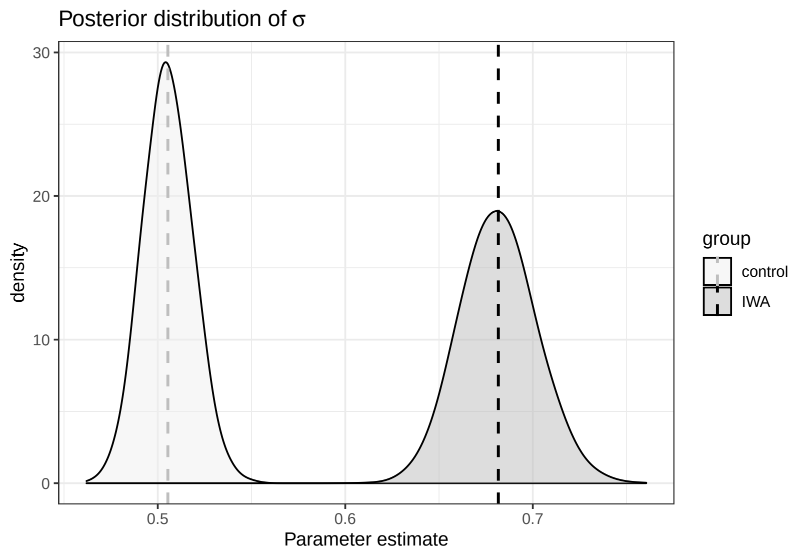

Figure 4.6: Posterior distribution of the σ parameter for both groups, in log scale. The dashed lines show the mean of the distributions.

Figure 4.5c shows that IWA also tend to choose the right interpretation in SRs. The mean of the accumulator for SR in the SR condition is 2694 ms, whereas the mean of the OR accumulator is 4717 ms. However, Figure 4.5d indicates that it is difficult for IWA to differentiate between the two interpretations in the OR condition (4.5d), where the two distributions show greater overlap. On average, the accumulator for the correct interpretation is faster: The estimated mean for the OR accumulator in the OR condition is 3573 ms, whereas the estimated mean for the SR accumulator in the OR condition is 4553 ms. But the overlap between the two distributions shows that the accumulator for the incorrect interpretation is sometimes as fast as the one for the correct interpretation. Therefore, the model predicts a difficulty for IWA in distinguishing between the correct and interpretation in ORs.

Figure 4.5 shows that the model exhibits the predicted patterns: The means for the finishing times across conditions are slower for IWA than for controls. For IWA, the mean finishing times of the accumulator in the OR condition are more similar than for controls. We also predicted IWA to have a higher σ because we assumed that their rate of accumulation could be noisier, and the model estimates reflect this prediction, as displayed in Figure 4.6.

Posterior predictive checks

In order to evaluate the performance of the model we compared the empirical data against the posterior predictive distributions estimated by the model (Gelman et al., 2013), a procedure that is known as posterior predictive checks (PPCs). We present

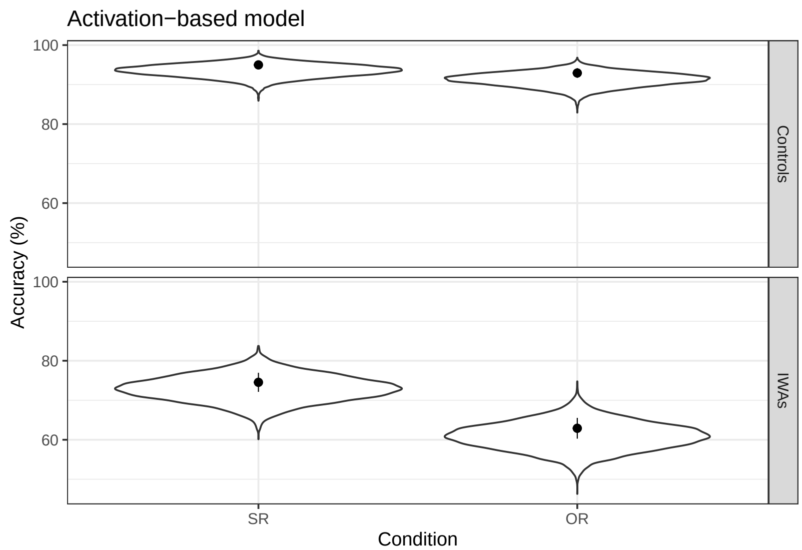

Figure 4.7: Violin plots depicting the PPCs for the activation-based model corresponding to the accuracy responses split by group and condition. The black dots represent the mean proportions of responses in the data and the corresponding error bars show 95% confidence intervals, and the violin plots display the posteriors predicted distributions from the model. Note that the controls' confidence intervals are not visible because variability is low in this group.

the PPCs graphically, with violin plots, where the dots represent the mean of the empirical data. This is a way to inspect whether the data could have been generated by the models: If the mean of the empirical data is predicted by the model, that is, if the dot lies within the violin plots, the model could have generated the data. If the model is unable to reflect the distribution of the data, that implies a bad fit.

Figure 4.7 shows the PPCs for the activation-based model in the picture-selection accuracies. In general, the activation-based model predicts the observed accuracies for both groups and conditions. Figure 4.8 shows the PPCs corresponding to the listening times. The model can correctly estimate the listening time (LT) distribution of the data across conditions and groups, although it tends to overestimate the LT for controls in incorrect responses.

4.6.2 Results of the direct-access model

The direct-access model was fit with three chains and 7000 iterations, and a warm-up of 3500. The chains were visually inspected, and we verified that all the Rhats were close to 1. Delta and the tree depth parameters were adapted when necessary and we made sure that the parameters of the model could be recovered using simulated data.

The DA model has three critical parameters: The probability of initial correct retrieval, θ, the probability of backtracking if the initial retrieval is not correct, P_b,

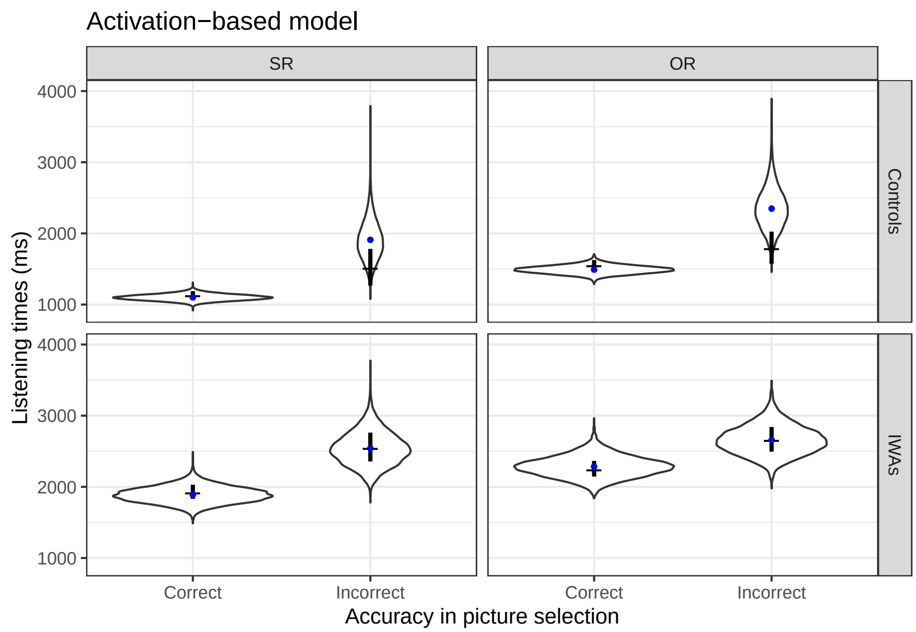

Figure 4.8: Violin plots depicting the PPCs of the activation-based model for listening times split by group and condition. The listening times correspond to the sum of the listening times for the verb of the subordinate clause plus the listening times of the second noun phrase. The horizontal bars represent the mean of the data and the vertical bars are the standard error of the mean. The dots represent the mean of the posterior distribution.

and δ, which is the time taken for backtracking. We turn now to assess the posterior distributions for these parameters across groups and conditions.

The posterior distribution of θ (Figure 4.9a) indicates that in SRs, controls initially retrieve the target 83% of the time, whereas IWA have a lower probability of initial correct retrieval, 69%. However, in ORs, the probability of initial correct retrieval is 41% for controls, and 53% for IWA. We discuss this surprising outcome below.

Regarding the probability of backtracking, the posterior distribution of the parameter P_b (Figure 4.9b) indicates that controls perform backtracking around 82% if they initially retrieve the distractor, whereas IWA backtrack 21% of the time. Notice that the parameters θ (Figure 4.9a) and P_b (Figure 4.9b) are interrelated, and should be interpreted together. The interpretation of both parameters shows that:

a) Controls initially carry out a retrieval that leads to the correct interpretation most of the time in SRs (83%), and 41% in ORs. If the first retrieval was incorrect, they backtrack and get the correct interpretation in 82% of the cases.

b) IWA are estimated to retrieve the correct interpretation without backtracking for SRs about 69% of the time and for ORs 53% of the time.

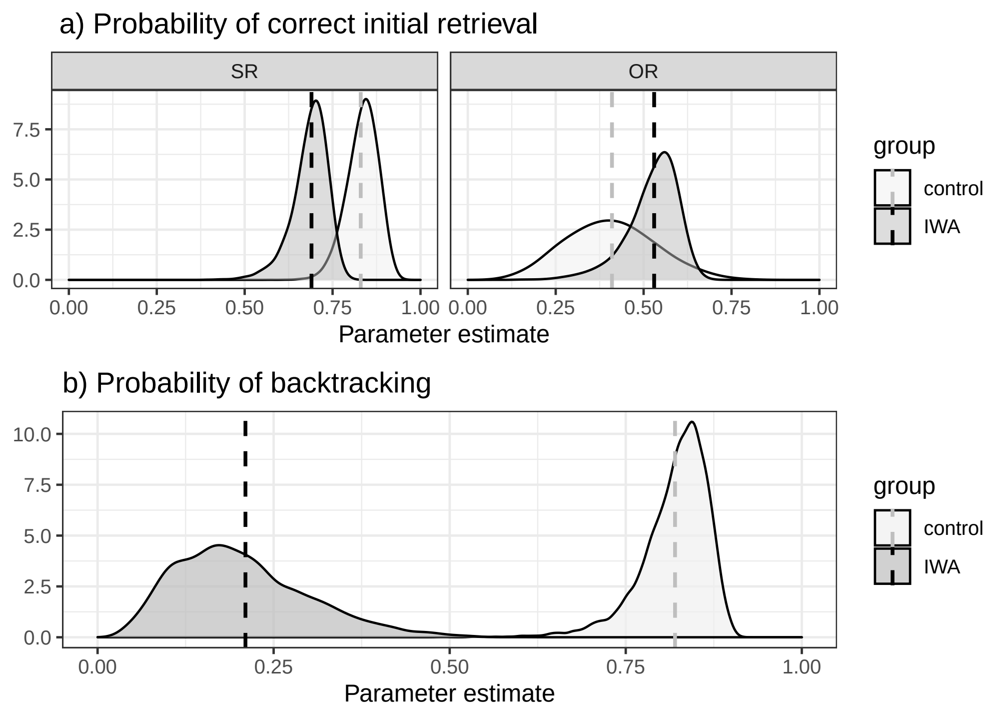

Figure 4.9: Posterior distribution for the probability of initial correct retrieval and backtracking in the direct-access model. This figure shows the estimated probability of initial correct retrieval across groups and conditions in the upper panel, and the estimated probability of backtracking across groups in the lower panel. The dashed lines stand for the means of their respective distributions.

> However, IWA backtrack only 21% after an incorrect first retrieval. Therefore, misretrievals are more likely for IWA than controls, especially in ORs.

Figure 4.10 shows the estimated time needed for backtracking. The posterior of δ shows that backtracking takes less time for controls, with a mean centered around 546 ms. By contrast, IWA's estimate for δ is higher, around 678 ms.

We predicted IWA to have a lower probability of backtracking relative to controls, and Figure 4.10 shows that the model confirms our prediction. We also predicted controls to have higher values for μ and σ. The model estimates are in line with these predictions (see Figure 4.11). However, the model's estimates contradict our prediction about θ: We had assumed that due to *resource reductions*, IWA should have a lower probability of initial correct retrieval in ORs. This surprising outcome in the direct-access model is an inherent shortcoming of the model, at least under the assumptions made here. We discuss alternative explanations in the discussion section.

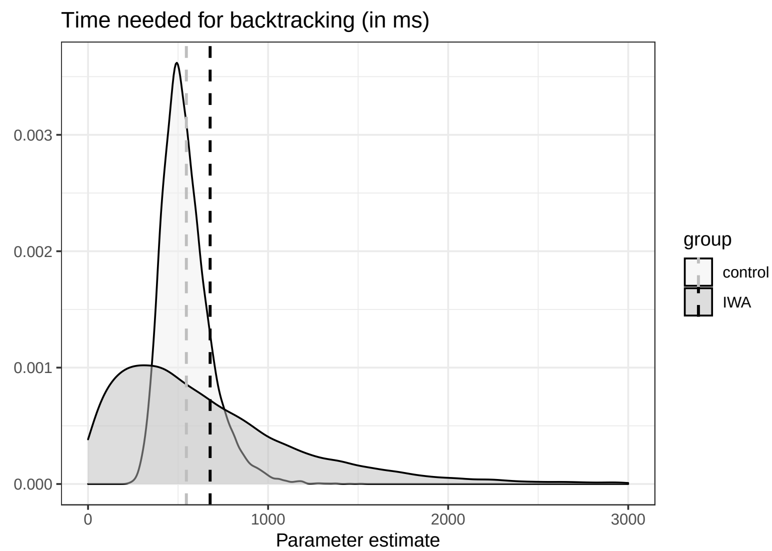

Figure 4.10: Posterior distributions of parameters representing the effect of backtracking (in miliseconds). The figure shows the posterior distribution of estimated time needed for backtracking across groups.

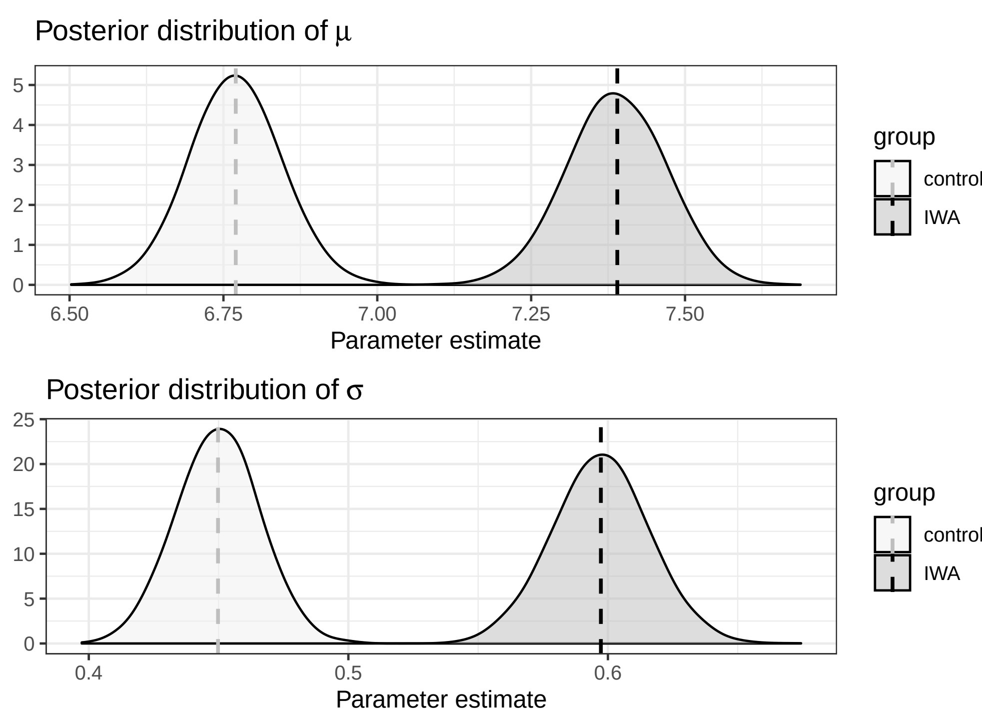

Figure 4.11: Parameter distributions of for the μ and the σ parameters across groups, on the log scale. The dashed lines stand for the means of their respective distributions.

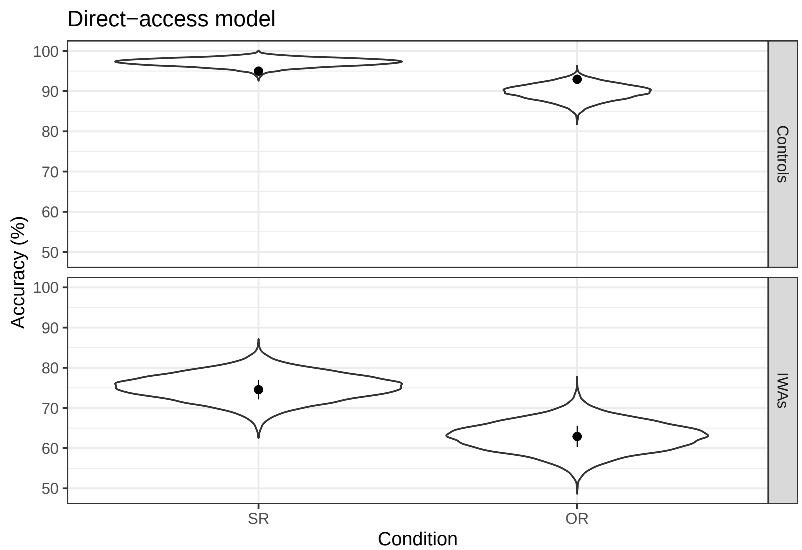

Figure 4.12: Violin plots depicting the PPCs for the DA model corresponding to the accuracy responses split by group and condition. The black dots represent the mean proportions of responses in the data, whereas the violin plots display the posteriors estimated by the model.

Posterior predictive checks

As with the activation-based model, we graphically compare the distribution of the empirical data with the estimated posteriors of the model. Figure 4.12 shows the PPCs corresponding to the picture-selection accuracies. In general, the model correctly predicts the qualitative pattern of the observed accuracies. Figure 4.13 shows that the model estimates the listening times across conditions and groups, but it tends to underestimate the listening times for incorrect responses, and overestimate the correct responses in the SRs condition for IWA.

4.7 Quantitative comparison of the activation model and the direct-access model

Although posterior predictive checks offer a visual way to assess the descriptive adequacy of the models, a more quantitative way of model assessment is required, in order to measure which model fits the data better. We compared the predictive accuracy of the models using 10-fold cross-validation (Vehtari et al., 2017). Cross-validation in the Bayesian framework allows for comparisons of models that assume different generative processes for the data, such as the two models in this study. 10-fold cross-validation involves splitting the dataset into 10 subsets of balanced data (balanced here means that each participant contributes approximately the same amount of data). One of the subsets is held out, and the model is fit to the nine remaining subsets. The pos-

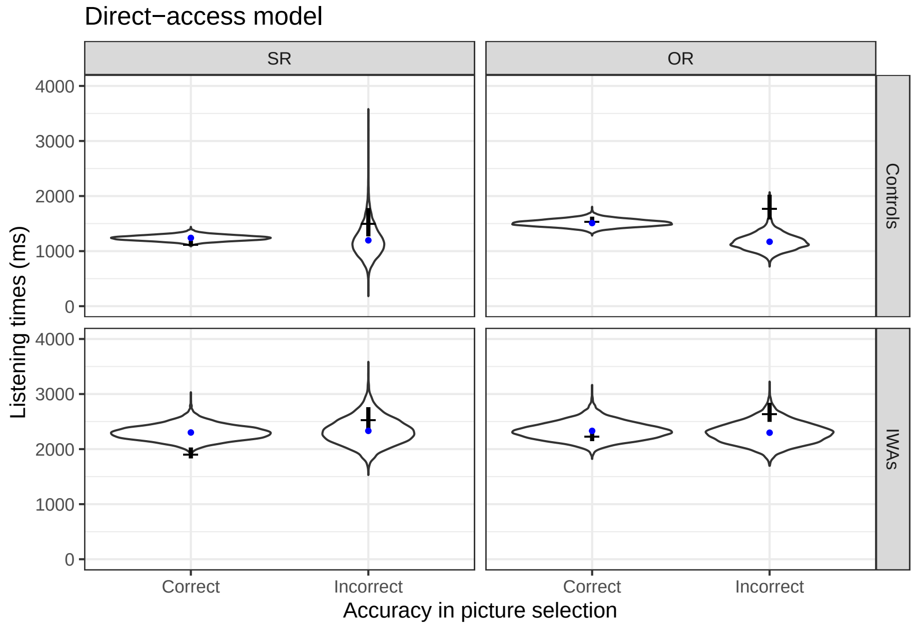

Figure 4.13: Violin plots depicting the PPCs of the direct-access model for listening times split by group and condition. The listening times correspond to the sum of the listening times for the verb of the subordinate clause plus the listening times of the second noun phrase. The horizontal bars represent the mean of the data and the vertical bars are the standard error of the mean. The blue dots represent the mean of the posterior distribution.

terior distributions of the parameters of this model are used to compute predictive accuracy on the subset of held-out data. This procedure is then repeated 10 times, one for each subset of held-out data. The difference between predicted and observed held-out data points is used to compute a measure of predictive accuracy: The *expected log point-wise predictive density*, or $\widehat{elpd}$. When comparing two models, the model with the higher $\widehat{elpd}$ value is the model that represents a better fit to the data. The standard deviation of the sampling distribution of $(\widehat{\Delta elpd})$, the difference in $\widehat{elpd}$, can also be computed, and has the standard frequentist interpretation: $(\widehat{\Delta elpd}) \pm$ $2\times$ SE) can be interpreted as a 95% confidence interval.

The $\widehat{elpd}$ values yielded a difference of 115 (SE = 69) in favor of the activation-based model. This suggests that the activation-based model shows a somewhat better fit for our data ($\widehat{elpd}_{act}$ = -12515, SE = 49 and $\widehat{elpd}_{DA}$ = -12630, SE = 52). However, the relatively large standard error means that the difference in the predictive performance of the models is not decisive. Table 4.1 details the difference in $\widehat{elpd}$ by condition and group, and their corresponding SE. Although the activation-based model consistently shows an advantage across conditions and groups, the standard errors indicate that the differences are not decisive.

In this section, the relative performance of the models was assessed. We turn now

Table 4.1: $\widehat{\Delta elpd}$ between the activation-based and the direct-access model across conditions and groups. A positive difference indicates an advantage for the activation-based model.

	$(\widehat{\Delta elpd})$	SE
SR, Controls	31	37
OR, Controls	42	36
SR, IWA	28	33
OR, IWA	14	33

to assess the relative importance that the individual parameters within each model have, in terms of explaining the data from IWA.

4.8 Model evaluation using Bayes factors

The estimates from the activation-based and the direct-access models show that IWA behave differently from controls. As discussed in the previous sections, given our linking assumptions, the different parameter estimates for the two groups can tell us whether the deficits that we link to the different parameters can explain IWA's data. For instance, the larger σ that IWA have in both models (relative to controls), indicates that *intermittent deficiencies* may be one of the causes of IWA's processing difficulties.

One question that arises is, to which extent is there evidence that these deficits are playing a role in IWA's sentence comprehension? By assumption, both models had group adjustments in all of the parameters. These adjustments reflect the difference between IWA and controls. However, if the group adjustment of a given parameter does not improve the model fit (i.e., the model would perform better if no difference was assumed between IWA and controls), this could mean that the processing deficit we are linking to this parameter may not be playing a role in impaired sentence comprehension. One way to assess whether the group adjustments improve the models' fit is to compute a series of Bayes factors.

The Bayes factor (BF) quantifies the evidence against or in favor of a null model (M0) that does not assume an effect of group (no β adjustment for the group factor), relative to a model that assumes a group effect (M1). The BF is a ratio of marginal likelihoods (as shown in Equation 4.8) and it indicates how likely it is that the data have been generated by one model relative to the other one. In Equation (4.8), the subscript in BF_{10} stands for the order of the models: Evidence of M1 over M0.

$$BF_{10} = \frac{P(Data|Model1)}{P(Data|Model0)} \tag{4.8}$$

The interpretation of BF is done in terms of relative odds. For instance, a BF_{10} of 5 means that the odds are $5:1$ in favor of M1. A BF closer to 1 is inconclusive, whereas a BF_{10} larger than 1 indicates evidence in favor of M1, and BF_{10} below 1 indicate evidence in favor of M0. The BF has a continuous scale (meaning the higher the BF_{10} the stronger the evidence for M1). There is no specific cut-off for the interpretation of the strength of the evidence in favor of a model over the other one, but guidelines have been proposed (Jeffreys, 1939/1998). In general, a BF_{10} larger than 100 is considered as strong evidence in favor of M1. Conversely, a BF_{10} of $1/100$ or smaller is considered as strong evidence in favor of M0.

BF and cross-validation are two different ways to perform model comparisons. Cross-validation is well suited for comparing models with different generative processes (such as the activation-based vs. the direct-access model), but cross-validation may be problematic with models that make very similar predictions. In this case, the estimated standard error might be biased (Sivula, Magnusson, & Vehtari, 2020). Since our model evaluation at the parameter level involves comparing nested models that are likely to make similar predictions, in this section we use Bayes factors instead of cross-validation. In what follows we perform a Bayes factor analysis for each parameter of the two models that has an adjustment for the group factor. For instance, for the σ parameter in both models, the M0 (null model) and M1 would be as shown in Equation (4.9).

$$\begin{aligned} M0_\sigma &: \sigma_0 \\ M1_\sigma &: \sigma_0 + \beta \cdot group \end{aligned} \tag{4.9}$$

Because BF is known to be sensitive to the choice of priors (Rouder, Haaf, & Vandekerckhove, 2018), we ran M1 with three different standard deviations for the prior of the β of interest (the adjustment for group) in order to show how the BF changes as a function of the prior standard deviation. The prior was always centered at 0 and the standard deviations were 0.1, 0.3, and 0.5. In addition, we included the following constraints:

(i) For the parameter μ in both models, and δ in DA, the group β in M1 was constrained to be positive. These parameters reflect the mean listening times and the time needed for backtracking, respectively. Therefore, according to theory, due to *slow syntax* and/or *delayed lexical access*, IWA should be slower than controls. Because the contrast coding is +1 for IWA and −1 for controls, a positive β would indicate that controls are faster than IWA, as shown in Equation (4.10).

$$\begin{aligned} Controls &: Intercept + (-1) \cdot \beta = Intercept - \beta \\ IWA &: Intercept + (+1) \cdot \beta = Intercept + \beta \end{aligned} \tag{4.10}$$

(ii) Similarly, for the parameter σ in both models, the group β was also constrained to be positive, since according to *intermittent deficiencies*, IWA should have more noise in the processing system.

(iii) We assumed that the probability of initial correct retrieval and the probability of backtracking could be linked to the *resource reduction* hypothesis. Therefore, IWA should show a lower θ and P_b estimate, and the group β was thus constrained to be negative. Since IWA are contrast coded +1, a negative β would imply a lower estimated probability for IWA.

(iv) In the activation-based model, a condition × group interaction is assumed on the μ parameter. The priors for the effect of this interaction should be vague because there is no prediction about the direction of the effect. One could assume that: a) IWA are more affected by the condition manipulation than controls, or b) IWA are less affected by the condition manipulation than controls, because IWA perform poorly in both conditions. Therefore, the β for the interaction did not have any constraint. And similarly, the β for the interaction in the θ parameter in DA was not constrained either.

A summary of the models that were run and their corresponding prior SD is shown in Table 2. All the BF were computed using the *bridgesampling* R package (Gronau, Singmann, & Wagenmakers, 2017) after running the models for 40,000 iterations. In addition, some of the models were run three times in order to confirm that the number of iterations was high enough to produce stable BF. Notice that for all parameters, M0 is the model that has no adjustment for the group effect. Three versions of M1 were run, each with a different prior sd for the group adjustment, as shown in Table 2. In the case of parameters with an interaction, nine versions of M1 were run, one for each possible combination of the prior SD of the two adjustments (the β for the group effect and the β for the interaction condition × group).

Table 4.2: Summary of the BF analysis for both models. This table shows the priors used, the theories that map to each parameter, and the BF results. ACT stands for the activation-based model, and DLA stands for delayed lexical access theory. The columns "Group SD" and "Inter. SD" show the different prior SD of the β adjustments to the effect of group and the interaction group $\times$ condition, respectively. In the "Group SD" column, a plus sign indicates that the β for the group adjustment was constrained to be positive, and a minus sign indicates that the β was constrained to be negative. No constraints were applied to the β of the interactions. The column "BF_{10}" summarizes the range of BF results for the priors shown in the table.

Model	Param.	Group SD	Inter. SD	Theory	BF_{10}
ACT	μ	0.1, 0.3, 0.5, +	0.1, 0.3, 0.5	Slow syntax, DLA	1/3 to 1
ACT	σ	0.1, 0.3, 0.5, +		Intermittent deficiencies	> 100
DA	μ	0.1, 0.3, 0.5, +		Slow syntax, DLA	> 100
DA	θ	0.1, 0.3, 0.5, −	0.1, 0.3, 0.5	Resource reduction	> 100
DA	P_b	0.1, 0.3, 0.5, −		Resource reduction	2 to > 100
				Intermittent deficiencies	
DA	δ	0.1, 0.3, 0.5, +		Slow syntax	1/3 to 1/11
DA	σ	0.1, 0.3, 0.5, +		Intermittent deficiencies	> 100

Results

Activation-based model

In the activation-based model there are two μ parameters, one for each accumulator of evidence, μ_{SR} and μ_{OR}. $M0_\mu$ does not include any adjustment for the effect of group or the interaction group $\times$ condition, for any of the two accumulators. $M1_\mu$ includes an adjustment for the effect of group and another adjustment for the interaction, for both accumulators. This is shown in more detail in in Equation (4.11).

$$
\begin{aligned}
&M0_\mu \\
&\qquad \mu_{SR} = \alpha_1 + u_{\alpha_1} + w_{\alpha_1} + (\beta_1 + u_{\beta_1}) \cdot rc_{type} \\
&\qquad \mu_{OR} = \alpha_2 + u_{\alpha_2} + w_{\alpha_2} + (\beta_2 + u_{\beta_2}) \cdot rc_{type} \\
&M1_\mu \\
&\qquad \mu_{SR} = \alpha_1 + u_{\alpha_1} + w_{\alpha_1} + (\beta_1 + w_{\beta_1}) \cdot group + \\
&\qquad (\beta_3 + u_{\beta_3}) \cdot rc_{type} + \beta_5 \cdot group \cdot rc_{type} \\
&\qquad \mu_{OR} = \alpha_2 + u_{\alpha_2} + w_{\alpha_2} + (\beta_2 + w_{\beta_2}) \cdot group + \\
&\qquad (\beta_4 + u_{\beta_4}) \cdot rc_{type} + \beta_6 \cdot group \cdot rc_{type}
\end{aligned}
\tag{4.11}
$$

The BF results are summarized in Table 2. The BF for μ in the activation-based model are either inconclusive or yield anecdotal evidence in favor the model that does not assume a difference between controls and IWA ($M0_\mu$).[6] In contrast, the BF results for σ yield strong evidence in favor of $M1_\sigma$: The model with a group adjustment for σ provides a better fit. This suggests that the group adjustment in σ could be sufficient to explain the differences between the two groups. Given our linking assumption, this means that the activation-based model estimates intermittent deficiencies to be the main source of processing deficits in IWA.

Direct-access model

In the direct-access model, the θ parameter also has a β for the interaction group $\times$ condition in addition to the β for the group effect. For the BF analysis, the $M0_\theta$ does not have any of these β, whereas the $M1_\theta$ has both, as shown in Equation (4.12).

$$\begin{aligned}
&M0_\theta \\
&\qquad \theta = \alpha + u_\alpha + w_\alpha + (\beta_2 + u_{\beta_2}) \cdot rc_{type} \\
&M1_\theta \\
&\qquad \theta = \alpha + u_\alpha + w_\alpha + (\beta_2 + u_{\beta_2}) \cdot rc_{type} + (\beta_3 + w_{\beta_3}) \cdot group + \beta_4 \cdot group \cdot rc_{type}
\end{aligned} \tag{4.12}$$

Nine versions of $M1_\theta$ models were run (see Table 2), such that all possible combinations of prior SD for both adjustments could be considered. All BF for θ yield strong evidence in favor of M1, the model that assumes that IWA have a lower probability of initial correct retrieval relative to controls (due to resource reductions). Irrespective of the prior SD, the BF for μ and σ yield strong evidence in favor of M1. The BF for P_b yields anecdotal to strong evidence in favor of M1 depending on the priors. In general, all of these parameters benefit from a group adjustment.

By contrast, the BF for δ yields some evidence in favor of M0, suggesting that the group adjustment is not needed. Recall that δ is the time needed for backtracking, and that estimated listening times for trials with backtracking are drawn from $(\mu + \delta, \sigma)$. The BF for δ could indicate that the group β is redundant because μ and σ (with their corresponding group adjustments) already explain the differences between controls and IWA. This means that IWA may not have an impairment in the mechanism of backtracking. That is, IWA perform backtracking less often than controls (as estimated in P_b), but when they do backtrack, the mechanism is not disrupted. These results suggest that the direct-access model accounts for *slow syntax* and/or *delayed lexical access* in μ (mean listening times), but not in δ (time needed for backtracking).

[6] A series of tables and plots showing the BF as a function of the priors for all of the parameters in both models is available in the online supplementary materials.

In conclusion, the Bayes factor analyses at the individual parameter level revealed that in the activation-based model, an increased noise value for IWA can explain the processing differences between IWA and controls, which speaks in favor of the *intermittent deficiencies* theory. The model could also be in line with *slow syntax* and/or *delayed lexical access*, but the BF for the parameter linked to these theories was inconclusive, so the role of these deficits in the activation-based model remains unclear. By contrast, the direct-access model is in line with a mixture of *slow syntax* and/or *delayed lexical access; resource reduction*, and *intermittent deficiencies*.

4.9 Discussion

In this study we presented a Bayesian implementation of two models of cue-based retrieval: The activation-based model, and the direct-access model. We linked the parameters of these models to major theories of processing deficits in sentence comprehension in aphasia, namely *slow syntax, delayed lexical access, resource reduction*, and *intermittent deficiencies*. The predictive performance of the two models was assessed with 10-fold cross-validation, and the quantitative and qualitative predictions of the models concerning data from IWA and controls has been discussed. A Bayes factor analysis was performed, in order to quantify the evidence that the models had with respect to the different processing deficits that were evaluated. In what follows we discuss some unexpected aspects of the direct-access model, we compare our findings to prior computational modeling work in the field of aphasia, and we point out some limitations of the present work as well as future directions.

4.9.1 Unexpected behavior of the direct-access model

The direct-access model estimates IWA to have a higher probability of initial correct retrieval in ORs relative to controls, which is surprising, since ORs are generally more difficult to process for IWA than for controls (Caramazza & Zurif, 1976). However, this prediction would be in line with studies showing that unimpaired controls have an agent-first preference: Unimpaired controls tend to interpret the first NP of a clause as the agent, which clashes with the actual thematic relations in some constructions (Hanne et al., 2015; Mack et al., 2016). For instance, in an eye-tracking experiment involving a sentence-picture matching task with active and passive sentences such as (16a) and (16b), Mack et al. (2016) found that unimpaired controls showed initial agent-first processing followed by a thematic reanalysis. That is, in passive sentences, controls tended to initially look at the image in which the first noun phrase was the agent. After hearing the region that contained the disambiguating morphological information (i.e., the verb: visiting/visited), controls started fixating the target picture. This implies that controls, after processing the morphological cues, had to reanalyze the initial agent-first interpretation. By contrast, in the study of Mack et al. (2016), IWA did not show signs of agent-first processing: They looked at the target and distractor pictures equally prior to the arrival of the disambiguating information.

(16) a. **Active:** The man was visiting the woman.
b. **Passive:** The man was visited by the woman.

Previous studies where controls showed an agent-first bias used eye-tracking and the visual world paradigm, but our modeling suggests that the agent-first bias could also be detected in a self-paced listening experiment. In our data, if unimpaired controls experienced an initial agent-first bias in ORs, they would initially parse the sentence as an SR. Consider sentence (17). Once they hear the disambiguating region (e.g., second noun phrase in sentence 17), they would have to backtrack on a high proportion of trials to end up with the right thematic interpretation. In this regard, the estimates for controls in ORs would be in line with an initial agent-first strategy. However, a replication of these estimates would be needed, ideally with visual-world eye-tracking data, as in Hanne et al. (2015) or Mack et al. (2016).

(17) **OR:** The girl who the mother chased hugged the boy.

Finally, a major issue for the DA model is the fact that the data show longer listening times for incorrect responses. This pattern contradicts the core assumptions of the model, because correct responses are expected, on average, to take longer due to the cost of backtracking.[7] Intuitively, IWA'S incorrect responses may be associated with longer listening times because after backtracking IWA may not be able to retain the retrieved representation. The *slow syntax* and the *resource reduction* hypotheses would be compatible with this view. However, the data show that the incorrect responses of unimpaired controls are also associated with longer listening times relative to correct responses. Therefore, the assumption that backtracking leads to the retrieval of the target (McElree, 1993) seems incompatible with our data.

4.9.2 Comparison with previous computational modeling work on aphasia

Taken together, the higher $\widehat{elpd}$ value in favor of the activation-based model, plus the fact that the direct-access model underestimates the listening times for incorrect

[7] Notice, however, that due to random noise the model estimates slower incorrect responses in some trials, as shown in the tails of the distribution for incorrect responses in Figure 14.

responses in ORs, suggest that the activation-based model is better at characterizing the processing of relative clauses in IWA and controls. The BF analysis for the activation-based model highlights the role that intermittent deficiencies may be playing in an activation-based mechanism of retrieval, but *slow syntax* and/or *delayed lexical access* should not be ruled out, since the BF for the parameters associated to these theories were rather inconclusive.

Our results are consistent with previous sentence processing modeling work on aphasia. For example, Patil et al. (2016) found that the LV05 model that included slowed processing (understood as a slowdown in the parsing mechanism) and intermittent deficiencies showed the best fit to data from IWA, relative to models that included only one of these deficits. It is also possible that IWA may exhibit different degrees of these deficits, as suggested by Mätzig et al. (2018), who modeled the accuracies of the Caplan et al. (2013) dataset estimating ACT-R parameters at the individual level. Interestingly, their modeling also revealed that *intermittent deficiencies* was the deficit that affected most of the IWA. Out of the 56 IWA, 53 showed a higher noise value (relative to the default noise value in ACT-R) in object relative clauses. Unfortunately, we do not have enough listening times data to get robust parameter estimates at the individual level, but our modeling suggests that on average, IWA are more subject to intermittent deficiencies than to *slow syntax* and/or *delayed lexical access*.

One caveat that applies to Patil et al. (2016), Mätzig et al. (2018), and our own work, is that the models cannot distinguish between *slow syntax* and *delayed lexical access*. In our implementation of the activation-based and the direct-access models, one possibility would be to include a shift parameter (Rouder, 2005) that accounts for lexical access, as implemented in Nicenboim and Vasishth (2018). Ideally, this parameter should have a group adjustment (to assess whether there is a delay in lexical access in IWA on average, taking the estimate for controls as reference), and an individual adjustment, to assess to which extent each individual is affected by this deficit. Unfortunately, such parameter could not be fit due to data sparsity.

Another issue to consider is that our modeling is limited to sentence comprehension. There is important modeling work in the aphasia literature that focuses on lexical processing (Evans, Hula, & Starns, 2019; Mirman, Yee, Blumstein, & Magnuson, 2011), the interface between lexical access and word production (Dell, Lawler, Harris, & Gordon, 2004), and word production (Walker, Hickok, & Fridriksson, 2018), among others. Ideally, a model of impairments in IWA should account for both aphasic comprehension and production, and disentangle the difficulties that arise from lexical and syntactic processes. However, as we show in this study, there is no single parameter that can account for aphasic impairments, and it is very unlikely that a computational model, even with a larger number of parameters, could account for all the particularities of aphasic performance, which is variable in nature. Nevertheless, we believe that more computational modeling is needed in the field of aphasia, in order to better understand the underlying nature of language impairments in IWA. Computational models require researchers to formalize hypotheses and assumptions, which is essential for theory development (Guest & Martin, 2021).

4.9.3 Some limitations of the present work and future directions

An important limitation of the present work is that even though the Caplan et al. (2015) dataset on IWA and age-matched controls is the largest currently in existence, the data are still relatively sparse compared to standard datasets used for similar model comparisons in psycholinguistics, both in terms of the number of of items (10) and participants (33 IWA and 46 controls). For example, Nicenboim and Vasishth (2018) compared the predictive performance of the activation-based and direct-access models from reading time data from some 180 participants. It would be useful to revisit these model comparisons with larger datasets in the future. Another important step will be to test the two models against new experimental designs and with different experimental paradigms. This would allow for a more comprehensive evaluation of the differences between the models, as well as an assessment of their predictive ability when modeling interference effects in different tasks, languages, and conditions. We are currently compiling a comprehensive database containing several tasks and conditions of data from IWA and unimpaired controls in German (Pregla, Lissón, et al., 2021). In future work, we intend to use this database to further evaluate the models discussed here.

4.10 Conclusion

We compared the predictive performance of two competing models of cue-based retrieval using data from individuals with aphasia and age-matched controls. We tested whether the two models—the activation-based model and the direct-access model—could account for experimental data from both individuals with aphasia and controls. This is the first study where competing models of cue-based retrieval have been tested against data from impaired populations. We also investigated the relative importance of the various parameters in both models using Bayes factors. The Bayes factors analyses show that in the activation-based model, *intermittent deficiencies* (Caplan et al., 2015) best explains the behavioral data from IWA, although *slow syntax* (Burkhardt et al., 2008), and *delayed lexical access* (Ferrill et al., 2012) may also play a role. In the direct-access model, the behavior of IWA is best explained in terms of a combination of *slow syntax*, *delayed lexical access*, *resource reduction* (Caplan, 2012), and *intermittent deficiencies*. The model comparisons show that both models have a similar performance for out-of-sample predictions (assessed with 10-fold cross-validation), with a slight advantage for the activation-based model.

In closing, we have presented the first-ever computational evaluation of different models of dependency completion, using the largest-available database from individuals with aphasia and unimpaired controls that currently exists. Our work lays out a systematic workflow that can be used to quantitatively compare the predictions of competing models of language processing.

Chapter 5

The modified direct-access model of retrieval

The contents of this chapter are published in:

Lissón, P., Pregla, D., Paape, D., Burchert, F., Stadie, N., & Vasishth, S. (2021). 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). ACL.

5.1 Introduction

Several researchers have argued that sentence comprehension is mediated by a cue-based, content-addressable retrieval mechanism that allows fast and direct access to memory items (e.g. McElree, 2000, 2006; McElree et al., 2003; Van Dyke & Johns, 2012; Van Dyke & McElree, 2006, 2011). Cue-based retrieval theory assumes that word representations are retrieved from working memory via their syntactic and semantic features. One model of cue-based retrieval is the direct-access model developed by McElree and colleagues (A. E. Martin & McElree, 2008; McElree, 2000; McElree et al., 2003). The direct-access model assumes that retrieval cues allow parallel access to candidate items in memory, as opposed to a serial search mechanism. Due to the parallelism assumption, the speed of retrieval is predicted to be constant across items (aside from individual differences and stochastic noise in the retrieval process).

Factors such as increased distance between the target and the retrieval point and the presence of distractor items can lower the probability of retrieving the correct dependent (also known as *availability*). Low availability of the target dependent can lead to failures in parsing or to misretrievals of competitor items. When such errors occur, a backtracking process can be initiated, which by assumption leads to the correct retrieval of the target (McElree, 1993). The backtracking process requires additional time that is independent of the retrieval time. For instance, according to the direct-access model, in the sentences below, (18a) should have shorter processing times than (18b) on average, because in (18b) some trials require costly backtracking due to lower availability of the target item *boy*.

(18) a. The boy who tickled the girl greeted the teacher.

b. The boy who the girl tickled greeted the teacher.

As shown in Chapter 4, the direct-access model can be adapted to explain impaired sentence comprehension in IWA. However, there is one crucial aspect of the direct-access model that is at odds with the aphasia literature, specifically with the finding that IWA have longer processing times for incorrect than for correct responses (e.g., Hanne et al., 2015; Pregla, Lissón, et al., 2021). The direct-access model assumes that some percentage of correct interpretations are only obtained after costly backtracking, and thus predicts that the average processing time for incorrect responses should be *faster* than for correct responses. To address this issue, we implement a modified version of the direct-access model that is specifically relevant for sentence processing in IWA. In this model, backtracking can lead to correct retrieval of the target, as in the base model, but can also result in misretrieval and parsing failure.

5.1.1 Sentence comprehension in aphasia

In the aphasia literature, there are several theories that aim to explain the source of these impairments in language comprehension. One possibility is that IWA carry out syntactic operations at a slower-than-normal pace, which could cause failures in parsing. This is the *slow syntax* theory (Burkhardt et al., 2008). By contrast, Ferrill et al. (2012) claim that the underlying cause of slowed sentence processing in IWA is *delayed lexical access*, which cannot keep up with structure building. Another theory, *resource reduction*, assumes that IWA experience a reduction in the resources used for parsing (Caplan, 2012), such as working memory. Finally, Caplan et al. (2013) claim that IWA suffer from *intermittent deficiencies* in their parsing system that lead to parsing failures. The results in Chapter 4 and previous computational modeling work have shown that these theories may be complementary (Patil et al., 2016), and that IWA may experience a combination of all of these deficits (Mätzig et al., 2018).

Assuming that a direct-access mechanism of retrieval subserves sentence comprehension, this mechanism could interact with one or more of the proposed processing deficits in IWA. One way to assess whether these deficits are plausible under a direct-access model is the computational modeling of experimental data. In Chapter 4, the direct-access model was tested against self-paced listening data from individuals with aphasia, and the model was found to be in line with multiple theories of processing deficits in aphasia. Despite this encouraging result, the model could not fit slow incorrect responses, due to its assumptions about backtracking and its consequences.

In what follows, we present an implementation of the original direct-access model and the modified version with backtracking failures. We fit the two models to data from individuals with aphasia and compare their quantitative performances. In order to assess the role of the different proposed deficits of IWA in sentence comprehension, we also map the models' parameters onto theories of processing deficits in aphasia.

5.2 Data

The data that we model come from a self-paced listening task in German (Pregla, Lissón, et al., 2021). 50 control participants and 21 IWA completed the experiment. Sentences were presented auditorily, word by word. Participants paced the presentation themselves, choosing to hear the next word by pressing a computer key. The time between key presses (here called listening time) was recorded. At the end of the sentence, two images (target and foil) were presented, and participants had to select which image matched the meaning of the sentence they had just heard. Accuracies for the picture-selection task were also recorded. To assess test-retest reliability, each subject completed the task twice, with a break of two months in between. Our modeling is based on the pooled data of both sessions.

5.2.1 Items

We investigate interference effects in a linguistic construction that is understudied in IWA: Control constructions. In control constructions, the subject of an infinitival clause is not overly specified, but understood to be coreferential with one of the overt noun phrases in the matrix clause of the same sentence (e.g, *Brian promises Martha to take out the trash* → Brian takes out the trash). In linguistic theory, it is assumed that a a phonologically empty element (PRO) occupies the subject position of *take out* (Chomsky, 1981). PRO is co-indexed with a noun phrase in the matrix clause that acts as its antecedent. The verb in the matrix clause specifies, according to its semantic and syntactic properties, which noun phrase in the matrix clause triggers the interpretation of PRO in the subclause.

In sentence (19a) below, the verb *verspricht* (promises) is lexically specified as a **subject-control** verb, and the subject noun phrase of the main clause, *Peter*, is chosen as the antecedent of PRO. By contrast, in (19b), the **object-control** verb *erlaubt* (allows) specifies that the object noun phrase of the main clause, *Lisa*, is the antecedent of PRO.

(19) a. **Subject control**
Peter$_i$ verspricht nun Lisa$_j$, PRO$_i$ das kleine Lamm zu streicheln und zu kraulen.

'Peter now promises Lisa to pet and to ruffle the little lamb'

b. **Object control**
Peter$_i$ erlaubt nun Lisa$_j$, PRO$_j$ das kleine Lamm zu streicheln und zu kraulen.

'Peter now allows Lisa to pet and to ruffle the little lamb'

Cue-based retrieval theory assumes that control clauses require completion of the PRO dependency through memory access to the correct noun phrase. The direct-access model would predict (19b) to be easier to process than (19a), because the target (*Lisa*) is linearly closer to the retrieval site at PRO, and thus more available. Therefore, at PRO, the probability of retrieval of the target should be higher in (19b) relative to (19a). In line with this prediction, unimpaired subjects show a processing advantage for object control over subject control (Kwon & Sturt, 2016). Similarly, IWA exhibit more difficulties understanding subject control conditions in acting-out tasks (Caplan & Hildebrandt, 1988; Caplan, Hildebrandt, & Makris, 1996). However, the object control advantage in IWA has not been previously tested using online methods.

Our experimental items were 20 sentences (10 per condition) similar to (19a) and (19b). The corresponding pictures for the picture-selection task are shown in Figure (5.1). The top picture is the target picture for (19a), whereas the bottom picture is the target for (19b). We assume that trials where the foil picture has been selected (i.e., the picture that shows the distractor noun as the agent of the action) correspond to a misretrieval.

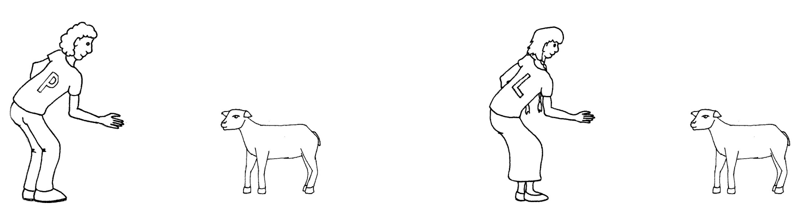

Figure 5.1: Example pictures used in the picture-selection task.

5.2.2 Dependent variables

The dependent variables used for modeling were the listening times (henceforth, LT) at the retrieval site (PRO) and the accuracy of the picture-selection task. Given that PRO is phonologically empty, we assumed that the retrieval process takes place at some point between the second and the third noun phrase (*Lisa* and *das kleine Lamm* in (19a)). We therefore summed the listening times of these regions within each trial. In order to evaluate the slowed lexical access hypothesis (Ferrill et al., 2012), we also used data from an auditory lexical decision task that participants performed in addition to the experiment. This task was based on LEMO 2.0 (Stadie, Cholewa, & De Bleser, 2013). Participants had to decide whether an auditorily presented item was a word or a neologism, and the response times were recorded. For each participant, we computed the mean response times for correct responses. These were then centered and scaled within groups and used as continuous predictors in the models. We will refer to the scaled lexical decision task reaction times as the LDT predictor.

5.3 Direct-access model

The implementation of the direct-access model follows is very similar to the implementation presented in Chapter 4. The model assumes that listening times for correct responses come from a mixture distribution, given that there are trials with backtracking, where an additional processing cost δ is added, and trials without backtracking, where no such cost is added. By contrast, incorrect responses never involve backtracking, and the average listening time should be the same as for correct responses without backtracking. A graphical representation of the model is displayed in Figure (5.2). The three possible cases are as follows:

(a) Retrieval of the target succeeds at first attempt, with probability θ: $LT \sim lognormal(\mu, \sigma)$

(b) Retrieval fails at first attempt, backtracking is initiated, with probability $(1-\theta) \cdot P_b$: $LT \sim lognormal(\mu + \delta, \sigma)$

(c) Retrieval fails, no backtracking, and a misretrieval occurs, with probability $(1-\theta) \cdot (1-P_b)$: $LT \sim lognormal(\mu, \sigma)$

The model includes both fixed and random effects in order to account for sentence complexity, group differences, and individual variability. The hierarchical structure is shown in Equation (5.1). All parameters have an adjustment by group (IWA versus control), because we expect IWA to have different parameter estimates from control participants. Since DA assumes that retrieval times are not affected by sentence complexity, the average listening times (μ) do not have an adjustment for condition. By contrast, the probability of retrieval of the target, θ, includes a condition adjustment. This parameter can be thought of as indexing memory availability.

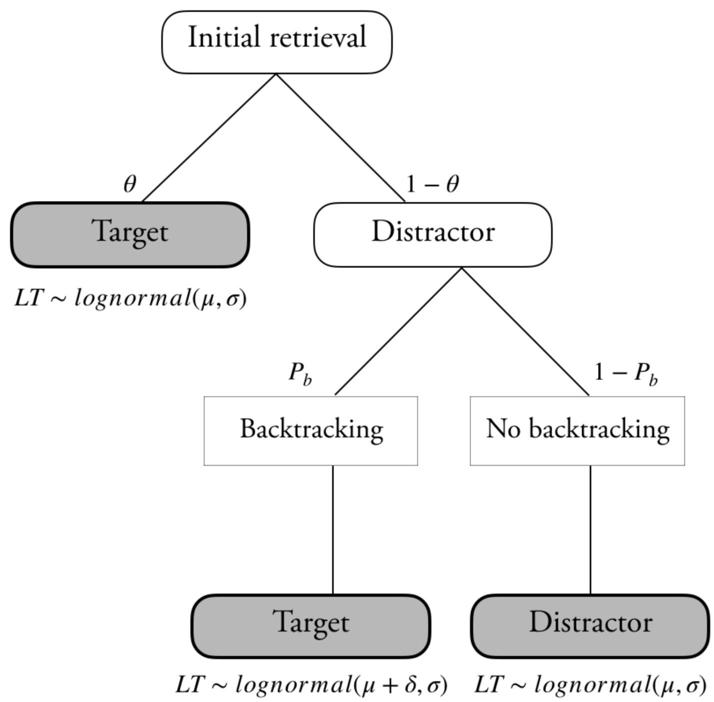

Figure 5.2: Graphical representation of the direct-access model.

The probability of backtracking P_b, the cost of backtracking δ, and σ do not depend on sentence complexity, but may vary between IWA and controls. The hierarchical structure is embedded within the parameters when possible (we report the maximal hierarchical structure that could be fit). In Equation (5.1), the terms u and w are the by-participant and by-item adjustments to the fixed effects, respectively. These are assumed to come from two multivariate normal distributions.

The priors used for all model parameters have been detailed in Chapter 6.9. The model was implemented in the probabilistic programming language Stan (Stan Development Team, 2021a), and fit via the rstan package (Carpenter et al., 2017) in R (R Core Team, 2020). The model was fit with 3 chains and 8,000 iterations, half of which were used as warm-up.

$$
\begin{aligned}
\mu &= \mu_0 + u_{\mu 0} + w_{\mu 0} + \beta_1 \cdot group \\
\theta &= \alpha + u_\alpha + w_\alpha + \beta_2 \cdot LDT + \\
&\beta_3 \cdot LDT \cdot group + \\
&(\beta_4 + w_{\beta_4}) \cdot group \\
&(\beta_5 + u_{\beta_5}) \cdot condition + \\
&\beta_6 \cdot group \cdot condition \\
P_b &= \gamma + u_\gamma + \beta_7 \cdot group \\
\delta &= \delta_0 + \beta_8 \cdot group \\
\sigma &= \sigma_0 + \beta_9 \cdot group
\end{aligned}
\tag{5.1}
$$

5.3.1 Predictions

Based on the theories of processing deficits in aphasia discussed in Section (5.1.1), and on the findings in Chapter 4, we make the following predictions:

1. IWA's μ and δ values should be higher than controls'. This would be in line with slow syntax, assuming that both the initial retrieval and the backtracked retrieval are accompanied by appropriate structure-building processes.

2. The probability of initial retrieval of the target θ should be lower for IWA relative to controls, across conditions.

3. Object control conditions should have a larger θ, relative to subject control. In addition, IWA should have a bigger interference effect, i.e., the difference in θ between the two conditions should be larger in IWA than in controls. This pattern would be expected under the resource reduction theory, which states that IWA should have greater difficulties in more complex sentences.

4. Slower lexical decision (LDT) should be associated with a decrease in θ across groups. Strong support for delayed lexical access would come from an interaction between LDT and group, such that an increase in LDT predicts a greater decrease in θ for IWA than for controls: Slow lexical access could cause parsing problems for controls, but if delayed lexical access is the main cause of difficulty in IWAs, parsing failures should occur

more often in this group for individuals whose lexical access is particularly slow.

5. The probability of backtracking should be lower for IWA, which would be in line with resource reduction.

6. Finally, the dispersion parameter σ of the listening-time distribution should be larger for IWA, which would indicate that IWA have more noise in their parsing system. This would be in line with intermittent deficiencies, since more noise could be due to more breakdowns in parsing.

These predictions build on the previous work by Lissón et al. (2021), but other options for the mapping between parameters and theories of comprehension deficits in aphasia are possible, see Mätzig et al. (2018), Patil et al. (2016).

5.3.2 Results

We begin by assessing the posterior distribution of the probability of retrieval of the target, θ, shown in Figure (5.3).

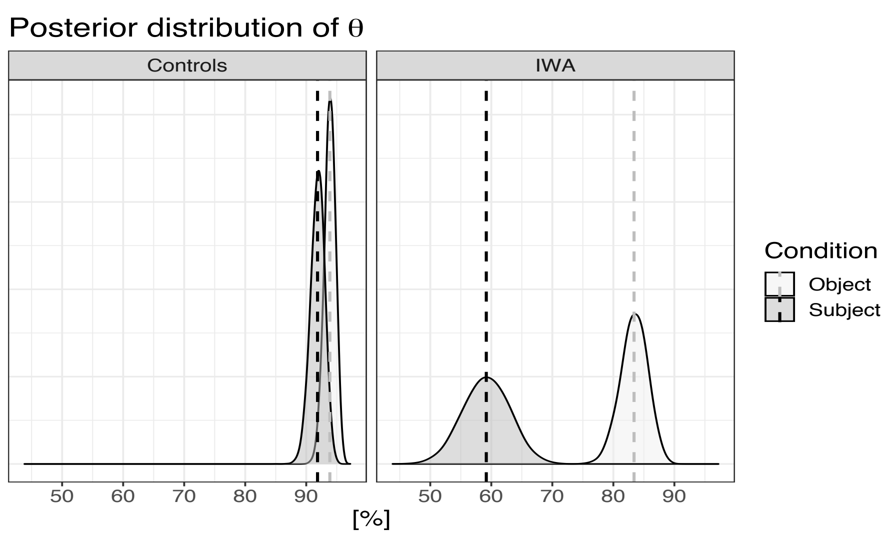

Figure 5.3: Posterior distribution of θ across conditions and groups.

Controls are estimated to retrieve the target at the first retrieval attempt in both conditions in more than 90% of trials. The mean of the subject-control condition is slightly lower than the mean for the object-control condition. By contrast, IWA display a greater effect of interference: In object-control sentences, where the antecedent is close to PRO, IWA are estimated to correctly retrieve the target at the first attempt 85% of the time, compared to 60% for subject-control. An increase in LDT leads to a decrease in θ of -6% CrI: $[-11\%, -2\%]$, but there was no interaction with group

$\times$ LDT (-2% CrI: $[-6\%, 2\%]$). The credible intervals for the remaining parameters are shown in Table (5.1).

Table 5.1: Parameter credible intervals, DA model.

Par.	Control participants	IWA
μ	[1668 ms, 1901 ms]	[2508 ms, 3073 ms]
δ	[1084 ms, 1385 ms]	[2897 ms, 6836 ms]
P_b	[63%, 78%]	[3%, 10%]
σ	[0.15, 0.16]	[0.27, 0.3]

As expected under the slow syntax theory, IWA's mean listening times (μ) and the time needed for backtracking (δ) are higher than controls'. Similarly, σ is also higher for IWA, as predicted by intermittent deficiencies. Finally, the probability of backtracking is much lower for IWA than for controls. Assuming that backtracking uses general parsing resources, this estimate is in line with resource reduction.

5.3.3 Posterior predictive checks

One way to assess the behavior of the model is to check the posterior distribution of data generated by the model against the empirical data. If the mean of the empirical data falls within the range of predicted values of the model, the model could have generated the empirical data. By contrast, if the empirical data are outside of the range of the generated values, this indicates a suboptimal fit. Figure (5.4) shows the posterior predictive distributions of the direct-access model across groups and conditions. Overall, correct responses are modeled reasonably well, except in the object-control condition for IWA. The model also underestimates the listening times for incorrect responses, except for IWA in the subject-control condition. In all other design cells, incorrect responses are slower than correct responses, contrary to the model's assumption that slow backtracking responses are always correct.

5.4 Modified direct-access model

Based on the original DA model's suboptimal fit, we propose a modified version (MDA). In this version, the distribution of listening times for both correct and incorrect responses is a mixture of directly accessed and backtracked retrievals. The MDA model assumes that backtracking can fail. In terms of implementation, the main difference between the models is a newly-introduced parameter θ_b, which is the probability of correct retrieval after backtracking. Figure (5.5) displays a graphical representation of this new model: After backtracking, the target is retrieved with probability θ_b, and a misretrieval occurs with probability $1 - \theta_b$. The hierarchical

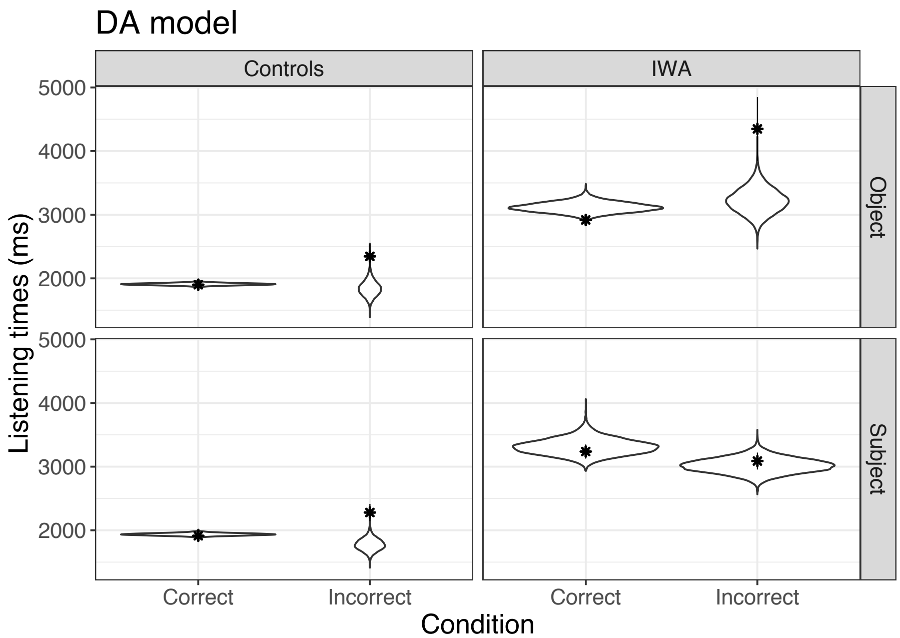

Figure 5.4: Posterior predictive checks of the direct-access model split by accuracy, group, and condition. The violin plots indicate the distribution of listening times generated by the model. The black stars stand for the mean of the empirical data.

structure is the same as in the DA original model, except for θ_b, whose adjustments are shown in Equation (5.2).

$$\theta_b = \alpha_b + u_{\alpha_b} + \beta \cdot group \tag{5.2}$$

The model was run with 10,000 iterations, half of which were used as warm-up.

5.4.1 Predictions

All predictions are carried over from the base DA model. In addition, the probability of retrieval of the target after backtracking θ_b should be lower for IWA than for controls. This would indicate that IWA are more likely to experience parsing failure or misretrieval even after backtracking.

5.4.2 Results

We begin by assessing the probability of first correct retrieval, θ. The posterior distribution across groups and conditions is shown in Figure (5.6). The estimates are quite similar to the ones in the original DA model: Controls have a very high probability of initial correct retrieval across conditions, and IWA display a greater interference effect.

As in the base model, IWA have a low probability of backtracking in this model (7%

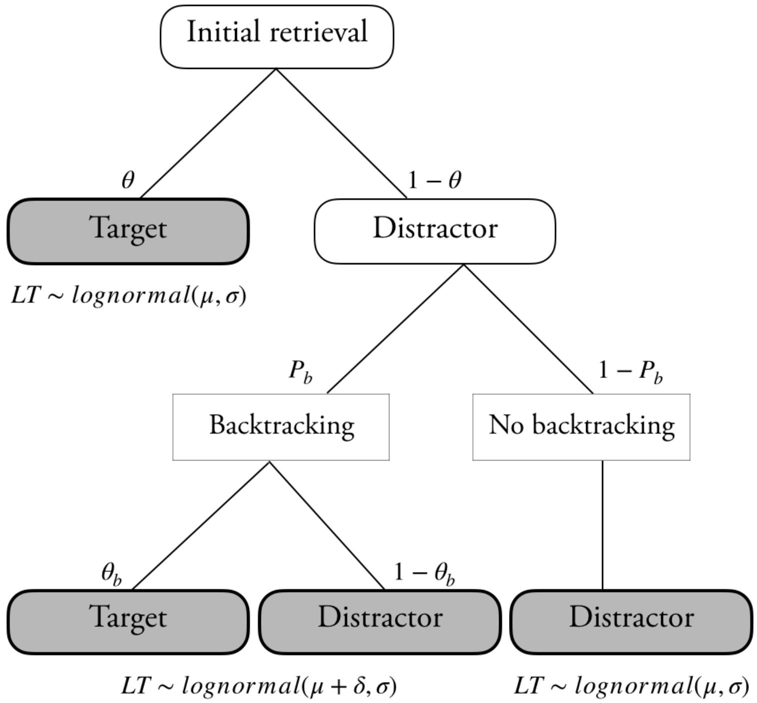

Figure 5.5: Graphical representation of the modified direct-access model.

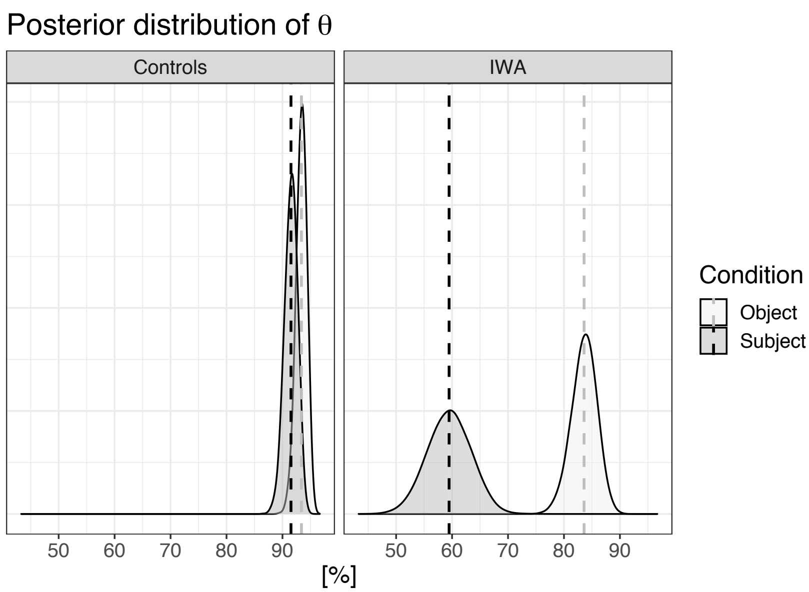

Figure 5.6: Posterior distribution of θ across conditions and groups.

CrI: [4%, 12%]) relative to controls (80%, CrI: [72%, 86%]). The probability of correct retrieval after backtracking, θ_b, determines the amount of slow incorrect responses. The posterior distribution of θ_b is shown in Figure (5.7). After backtracking, controls are estimated to retrieve the target 90% of the time, compared to around 70% for

IWA.

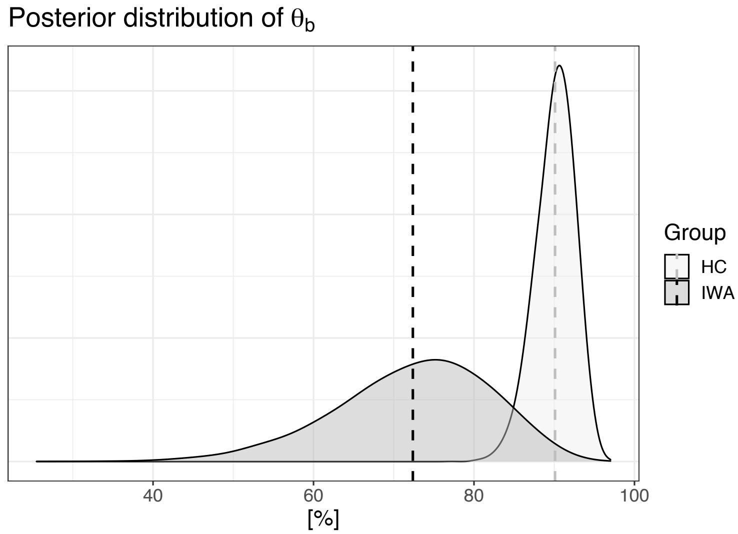

Figure 5.7: Posterior distribution of θ_b across conditions and groups.

The rest of estimates are also similar to the ones in the original DA model: IWA's μ is higher than controls' (2751 ms, CrI: [2477, 3046] versus 1770 ms, CrI: [1654 ms, 1890 ms]). The cost of backtracking, δ, is very high for IWA (6394 ms CrI: [4235, 9468]) relative to controls (1238 ms, CrI: [1103 ms, 1387 ms]). Finally, σ is also higher for IWA (0.27 CrI: [0.25, 0.28]) than for controls (0.15 CrI: [0.14, 0.15]).

5.4.3 Posterior predictive checks

The posterior predictive checks for the modified direct-access model are shown in Figure (5.8). Like the base model, the MDA mostly correctly estimates listening times for correct responses across the board. The fits for incorrect responses seem to have improved, except for object-control in IWA, where the predicted listening times are still faster than the observed listening times.

5.5 Model comparisons

In order to quantitatively compare the performance of the models, we computed Bayes factors. We chose Bayes factors over other alternatives (e.g. cross-validation), because the two models seem to predict similar distributions, and Bayes factors are especially suited for nested models, or models that make very similar predictions. The hypothesis being tested is whether there is a non-zero parameter θ_b that indexes the probability of successful backtracking, assumed by the MDA model, or whether backtracking is always successful, as assumed by the base DA model.

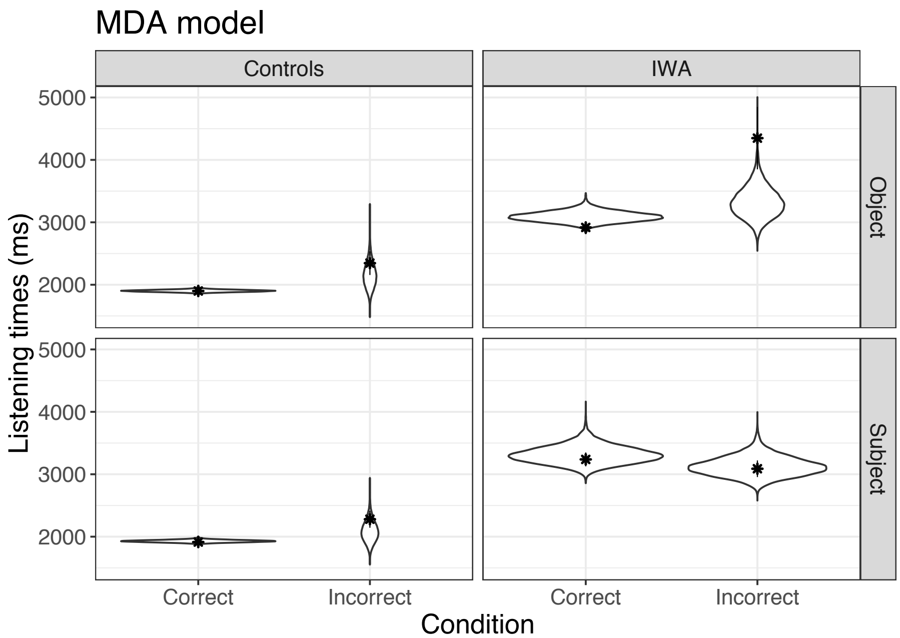

Figure 5.8: Posterior predictive checks of the modified direct-access model split by accuracy, group, and condition. The violin plots indicate the distribution of listening times generated by the model. The black stars stand for the mean of the empirical data.

In order to perform the comparison, the models were run for 40,000 iterations, of which 3,000 were used for warm-up. Bayes factors were computed using the *bridge-sampling* package (Gronau et al., 2017) in R. The Bayes factor of DA over MDA was estimated to be 2. This result is inconclusive, and indicates that the models provide similar quantitative fit to the data.

5.6 Discussion and conclusion

In the present paper, we implemented and tested two versions of the direct-access model of cue-based retrieval and evaluated their predictive performance on data from individuals with aphasia and control participants. Specifically, we modeled interference in an under-studied linguistic construction, namely control structures.

Both the base model and the modified model are in line with a combination of processing deficits in IWA: slow syntax, resource reduction, and intermittent deficiencies. Neither of the two models showed support for delayed lexical access as a source of retrieval difficulty specifically for IWA. Although a delay in LDT was connected to a decrease in the probability of correct retrieval, the effect of LDT was similar for IWA and control participants. In general, our results are consistent with other studies showing that a combination of processing deficits may be the source of impairments in sentence comprehension in IWA (Caplan et al., 2015; Mätzig et al., 2018).

Unlike the base direct-access model, our modified DA model (MDA) assumes that backtracking can fail, resulting in slow, incorrect retrievals. However, this added assumption does not result in a decisive advantage in fit for the MDA model, as shown by the posterior predictive checks and the Bayes factor analysis. This result is unexpected, and leads us to think that the MDA model may be overparametrized. In MDA, all of the main parameters include a group adjustment. As a consequence, for instance, the mean listening times, μ, are estimated to be higher for IWA than for controls. The cost of backtracking, which is only added to μ if backtracking is performed, accounts for slower responses. However, because IWA's μ is estimated to be higher than controls' μ, the model may not need to rely on backtracking in order to account for slow responses in IWA. This could be the reason why the probability of backtracking for IWA is very low (7%) relative to controls (80%). In addition, IWA's θ_b has to be estimated from the 7% of trials that include backtracking. Given the size of the IWA group (21 participants), and the small amount of trials that include backtracking, perhaps the model cannot correctly estimate the θ_b parameter. This could be investigated in several ways. One possibility would be to remove the group adjustments from μ, P_b, δ, and θ_b one at the time, and see which of these models shows a better quantitative fit for the data (see Chapter 4). Another possibility would be to evaluate how these parameters interact with and without group adjustments (e.g., do P_b and/or δ for IWA increase if there is no group adjustment in μ?). We will address these questions in future work.

The present chapter contributes to the aphasia literature by proposing a modification of the direct-access model that can account for incorrect slow responses. Despite our inconclusive results, we believe that the modified direct-access model offers a more appropriate set of assumptions for individuals with aphasia than the direct-access model. The modified-direct access model can account for slow incorrect responses, which are frequently found in studies on sentence processing in IWA (e.g., Hanne et al., 2015; Lissón et al., 2021; Pregla, Lissón, et al., 2021). It remains to be seen, by testing the new modified direct-access model against more data from individuals with aphasia, whether there is a difference in predictive performance between the two models.

Chapter 6

Similarity-based interference in sentence comprehension in aphasia

The contents of this chapter are submitted for publication to the *Journal of Memory and Language*:

Lissón, P., Paape, D., Pregla, D., Burchert, F., Stadie, N., & Vasishth, S. (2021). Similarity-based interference in sentence comprehension in aphasia: A computational evaluation of two models of cue-based retrieval. (Submitted to Journal of Memory and Language.)

6.1 Introduction

Similarity-based interference is a key assumption in the cue-based retrieval theory. When a retrieval is triggered, processing difficulty is predicted if multiple items in memory match the same retrieval cues. This phenomenon occurs because the two items in memory become difficult to distinguish from each other, and this results in a processing slow-down, or an occasional misretrieval of the distractor item. Similarity-based interference has been attested in multiple linguistic constructions across different languages (e.g. Engelmann et al., 2019; Jäger et al., 2017; Jäger et al., 2020; A. E. Martin et al., 2012; Van Dyke, 2007; Van Dyke & Johns, 2012; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011; Vasishth et al., 2008; Vasishth & Engelmann, 2021). Both the Lewis and Vasishth (2005) model and McElree's direct-acccess model (2002), predict similarity-based interference effects. However, the two models assume that similarity-based interference has underlying different generative processes.

In LV05, 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 speed. 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 consequence of the reduced activation of items in memory leads to difficulty in retrieving the correct item; this is a key property of similarity-based interference.

In the direct access-model, it is assumed 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 a parsing failure. When a misretrieval occurs, in a certain proportion of trials, a backtracking process is initiated. Backtracking, also known as reanalysis, is implicitly assumed to lead to the retrieval of the target (McElree, 1993), and requires some extra processing time that is independent of the retrieval time (A. E. Martin & McElree, 2008). Therefore, in the direct-access model, similarity-based interference can lead to slow, correct responses (due to the cost of backtracking), or to fast, incorrect responses.

However, as explained in Chapter 5, the assumption that backtracking leads to the retrieval of the target constrains the DA model: Due to the added backtracking time in some trials, DA assumes that correct retrievals are, on average, slower than misretrievals. Yet, it is known from different cognitive tasks that "slow errors" can occur in addition to "fast errors" (e.g., van Maanen, Katsimpokis, and van Campen, 2019). In addition, the DA 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, as seen in the previous chapters. The reason is that individuals with aphasia often have slower latencies in incorrect trials relative to correct trials (see Adelt, Stadie, Lassotta, Adani, & Burchert, 2017; Hanne et al., 2015; Pregla, Lissón, et al., 2021). In Chapter 5, we implemented a modified version of the direct-access model in which backtracking can fail, resulting in a misretrieval. The modified direct-access model and the original direct-access model were tested against self-paced listening data from individuals with aphasia (IWA) and control participants in German (Pregla, Lissón, et al., 2021). The models were compared using Bayes factors, and the result was inconclusive. We concluded that more studies testing the modified direct-access model in IWA are needed, in order to evaluate the fit of the modified-direct access model to data from IWA and controls. In the present chapter, we implement the same modified direct-access model, in which backtracking can fail, resulting in slowed reaction times and a misretrieval. In the modified direct-access model, both correct and incorrect responses are therefore a mixture of directly-accessed and backtracked retrievals. In the present work we focus on evaluating the modified direct-access model, but for the sake of completeness, we also report model comparisons between the modified and the original direct-access model, i.e., a model in which backtracking can only lead to the retrieval of the target.

In this chapter, we model interference effects in IWA and control participants using a subset of the data from Pregla, Vasishth, et al. (2021). 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 been found to have comprehension difficulties processing them (Adelt et al., 2017; Burchert, De Bleser, & Sonntag, 2003; Caplan et al., 2015; Choy & Thompson, 2010; Pregla, Lissón, et al., 2021). Because we map prominent theories of processing deficits in aphasia to the different parameters of each model, we expect to gain valuable insights into the nature of comprehension deficits in IWA by investigating how the two competing models of retrieval fit these constructions. In addition, we simultaneously model visual-world eye-tracking data, as well as reaction times and accuracies from a sentence-picture matching task. This is a novel approach that allows for the simultaneous assessment of IWA's online and offline comprehension within the same model. Our study is, to our knowledge, the first to compare two different models of cue-based retrieval using the visual-world paradigm and behavioral data from both IWA and unimpaired controls.[1]

We seek to answer the following questions: 1) Which model offers a better account of interference effects in IWA and control participants across these structures? 2) What do the parameters of each model tell us about the source of processing deficits and about interference in IWA?

We begin by briefly 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.

6.1.1 Theories of processing deficits in aphasia

Caplan et al. (2015) discuss the different theories that aim to explain why non-adjacent dependencies are challenging for IWA. For instance, Burkhardt et al. (2003) and Burkhardt et al. (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 et al. (2012), Love et al. (2008) posit that *delayed lexical access* causes the slowdown in the formation of syntactic dependen-

[1] Patil et al. (2016) modeled visual-world eye-tracking data from 7 IWA and 8 controls with different versions of the LV05 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.

cies. Love et al. 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 a extragrammatical heuristic may be used instead, which can lead to comprehension errors.

Caplan and colleagues argue that IWA may have an impairment in the resources needed for parsing, such as memory capacity (Caplan, 2012; Caplan et al., 2007). 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 in computational models of retrieval. In our modeling, which follows the same mapping as Chapter 4 and Chapter 5, intermittent deficiencies are implemented as increased stochastic noise in memory activations/availabilities. A higher noise value in IWA would mean more mistretrievals and presumably more parsing failures 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.

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. We use data from a visual-world experiment that tested the comprehension of pronouns and relative clauses in German (Pregla, Vasishth, et al., 2021). We now introduce each linguistic construction in turn.

6.2 Experiment 1: Pronoun resolution

Consider sentence (20a). 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]. 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), so that the cue [+subj] is set at the pronoun. The item whose features fully match these retrieval cues is the licensed antecedent, and therefore *Peter* is the target of the retrieval.

(20) a. **Mismatch.**
Peter$^{+subj}_{+masc}$ verspricht nun Lisa$^{-subj}_{\mathbf{-masc}}$, dass er$\{^{subj}_{masc}\}$ das kleine Lamm streichelt und krault.

Peter promises now Lisa, that he the small lamb pets and ruffles.

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

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

Peter promises now Thomas, that he the small lamb pets and ruffles.

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

Across the two sentences, the object nouns, *Lisa* in (20a) and *Thomas* in (20b), 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 (20a) as mismatch conditions, because the target noun (Peter) and the distractor noun (Lisa) do not share the same gender, and sentences like (20b), as match conditions.

A processing advantage in gender mismatch conditions in unimpaired populations has been observed in English by Badecker and Straub (2002) and Runner and Head (2014), but not by Chow, Lewis, and Phillips (2014). Laurinavichyute, Jäger, Akinina, Roß, and Dragoy (2017) reported mixed results for German. In the aphasia literature, Choy and Thompson (2010) and Engel et al. (2018) found that IWA had difficulties in pronoun resolution, but these studies did not target the gender mismatch configurations that Pregla, Vasishth, et al. (2021) 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. Based on cue-based retrieval theory, we predict a processing advantage in mismatch conditions relative to match conditions, and 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 from their implementation in the models.

6.3 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., Fedorenko, Gibson, & Rohde, 2006; Gordon et al., 2006; Grodner & Gibson, 2005; Staub, 2010; Staub, Dillon, & Clifton Jr, 2017) and for IWA (e.g., Burchert et al., 2003; Caplan et al., 2013, 2015; Caplan et al., 2007; Caramazza & Zurif, 1976; Dickey & Thompson, 2009; Pregla, Lissón, et al., 2021). The subject-object asymmetry in IWA and controls has been computationally modeled in the cue-based retrieval framework (Lissón et al., 2021; Mätzig et al., 2018; Vasishth et al., 2019) 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 individuals with aphasia and unimpaired controls.

Consider the sentences in (21). 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. In (21), the arrow points towards the target of the retrieval that we model, i.e., the agent. A + or a − preceding the number cue of the target and distractor indicates a match or a mismatch with the retrieval cue, which is shown at the retrieval site, the verb.

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

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

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

'Here is the donkey who bathes the tiger.'

b. **OR, match**

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

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

'Here is the donkey who the tiger bathes.'

c. **SR, mismatch**

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

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

'Here is the donkey who bathes the tigers.'

d. **OR, mismatch**

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

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

'Here is the donkey who the tigers bathe.'

In (21a) and (21b), both noun phrases in the sentence are singular (*der Esel, der/den Tiger*). Accordingly, both noun phrases have the retrieval cue [singular]; this is expected to cause similarity-based interference during retrieval. By contrast, in (21c) and (21d), the second noun phrase is plural (*die Tiger*), which should result in easier identification of the correct noun phrase, based on the retrieval cue from the verb ([singular] or [plural]). We will refer to sentences like (21c) and (21d) as mismatch conditions, because target and distractor do not share the same number, and sentences like (21a) and (21b), as match conditions.

According to cue-based retrieval, there should be no processing difference between subject and object relatives at the final verb in the sentences in (21), because the subject and object need to be retrieved at the verb in both types of relatives. However, both types of relative clauses should be easier to process in mismatch configurations compared to match conditions. Thus, (21c) and (21d) should be easier to process than (21a) and (21b), 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 (21c) vs. (21a) and (21d) vs. (21b), because in (21c) and (21d) 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. Friederici, Steinhauer, Mecklinger, and Meyer (1998) tested the comprehension of these sentences in unimpaired controls using ERP. They found a P600 in ORs relative to SRs at the relativizer. No significant difference in accuracies between the two conditions was reported. Studies investigating case-unambiguous subject vs. object relatives in IWA in German indicate that object relatives are generally more difficult to process for IWA. For example, Burchert et al. (2003) report that in case-unambiguous relatives the accuracies for OR are lower, relative to SR, in 4 out of the 7 IWA tested. Similarly, in a visual-world experiment, Adelt et al. (2017) also tested the comprehension of case-unambiguous relative clauses in IWA and control participants in German. In order to compare the accuracy estimates from the present paper with those of Adelt et al. (2017), we carried out a Bayesian logistic regression analysis of the accuracy data reported in Adelt et al. (2017). The analysis is available in the online supplementary materials. In the case-unambiguous relative clauses, IWA have lower accuracies than controls (-26% credible interval (CrI): [-42%, -13%]) and both groups have lower accuracies in object relatives (-13 % CrI: [-23%, -4%]). There is no indication of a group $\times$ RC type interaction (3 % CrI: [-5%, 11%]). Although the items in Adelt et al. (2017) and in Burchert et al. (2003) also included relative clauses with plural noun phrases, these relative clauses were case ambiguous. Therefore, these studies did not investigate whether a unique number cue facilitates processing. This is the empirical question that we address in this experiment.

The main goals in Experiment 2 are to compare the performance of the activation-based and the modified direct access model by modeling number interference in relative clauses, and to evaluate the different theories of processing deficits in aphasia from their implementation in the models.

6.4 Methods

The data that we model come from the experiments carried out by Pregla, Vasishth, et al. (2021). The participants, procedure, and materials described here summarize the methods in Pregla, Vasishth, et al. (2021). By contrast, the dependent variables and contrast coding described here are specific to the present work.

Participants and procedure. Twenty-one IWA (mean age = 60.2 years, SD = 11.4) and fifty control participants (mean age = 47.7 years, SD = 19.6), 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 one year after onset of the aphasia). 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 seconds. 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 two months. 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, 10 items per condition (match versus mismatch) were included, as in example (20). Example pictures accompanying the pronoun sentences are shown in Figure (6.1). The pronoun items always used subject-control verbs, so that the target of the retrieval was always the first noun phrase.[2] Control verbs were selected from the ZAS Database of Clause-Embedding Predicates (Stiebels et al., 2018).

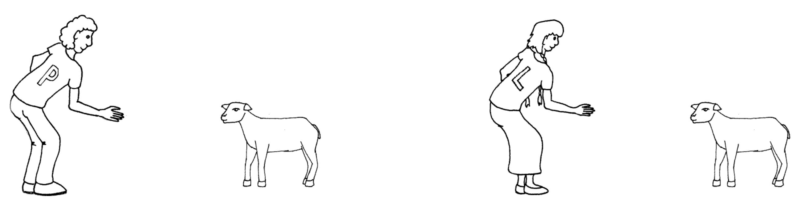

Figure 6.1: Example pictures used in the picture-selection task in Experiment 1. For the sentences in example (20) the left picture is the target and the right picture is the foil.

In Experiment 2, 20 items per condition (subject and object relatives) were included. The noun phrase of the matrix clause (henceforth NP1) was always 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 phrases were always bisyllabic animal names. Example pictures of the match relative clause conditions are given in Figure 6.2, and of the mismatch relative clause conditions in 6.3.

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 from a lexical decision task, based on LEMO 2.0 (Stadie et al., 2013). This

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

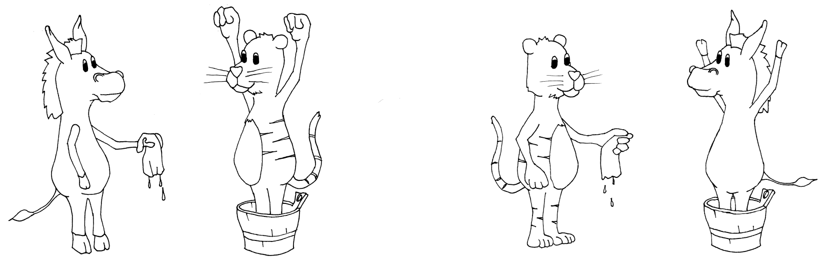

Figure 6.2: Example pictures used in the picture-selection task in Experiment 2 for the match conditions. The left picture is the target for SR conditions, and the right picture is the target for OR.

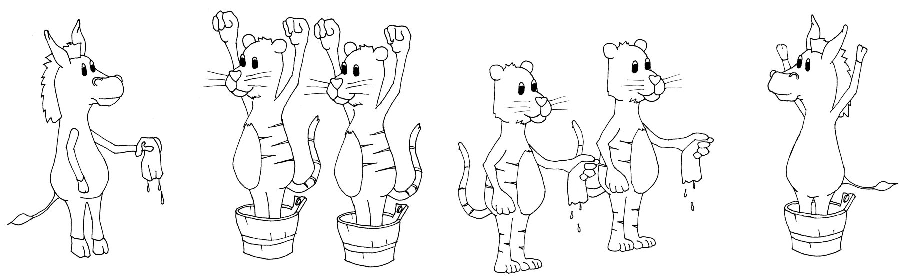

Figure 6.3: Example pictures used in the picture-selection task in Experiment 2 for the mismatch conditions. The left picture is the target for SR conditions, and the right picture is the target for OR.

yielded a single measure (lexical decision time, LDT) for each participant that we use as a continuous predictor in the models. We centered and scaled the LDT predictor within groups. An LDT $\times$ 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 a retrieval is assumed to take place (see below). 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 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 each experiment, two different versions of the activation-based model and the modified direct-access model were implemented.[3] The models were implemented in Stan (Carpenter et al., 2017) and fitted in R (R Core Team, 2020) via the rstan package (Stan Development Team, 2021a). The packages brms (Bürkner, 2017) and bayesplot (Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019, 2) were used for examining and plotting the posterior distributions of the parameters. For each model, three chains each, with at least 6,000 iterations each were run. Each chain included at least 3,000 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 (Nicenboim et al., 2021; Schad et al., 2021; Sorensen et al., 2016).

6.4.1 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 relatives 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. 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 pronoun resolution, the retrieval target must be re-accessed at this point, as it is the subject of the verb. In relative clauses, the retrieval target must be re-accessed too. 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

[3]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.

region. As the picture-selection task takes place immediately after hearing the verb region, we assume that accuracies and RT at this task should show the interference effects predicted by the cue-based retrieval theory.

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.

6.5 The activation-based model

The activation-based model can be implemented as a lognormal race of accumulators (Nicenboim & Vasishth, 2018): 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 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 model 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: 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 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. Across both constructions, 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 faster and more likely.

2. The mean finishing time of the target accumulator should be faster for the mismatch conditions relative to the match conditions. The mean of the distractor accumulator should be similar or slower in mismatch conditions relative to match conditions. That is, retrieval should be faster for mismatch than for match conditions, as interference slows retrieval.

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

6.6 Modified direct-access model

We implement the modified direct-access model as a hierarchical mixture model in the Bayesian framework, following Lissón, Pregla, et al. (2021). 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 MDA as a mixture model allows us to take into account the probability of backtracking as a latent variable, and to fit RT-based measures. We assume that both correct and incorrect responses can be generated from one of two distributions: Responses without backtracking are generated from a given distribution with parameters μ and σ; and responses with backtracking are generated from another distribution with parameters $\mu\star = \mu + \delta$ and σ, where δ is the time needed for backtracking. We begin this section by explaining the conceptual link between our parameters and the original direct-access model, which was originally evaluated using the speed-accuracy tradeoff paradigm (McElree, 2000, 2006).

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. Memory availability has been found to decrease as a function of sentence complexity (A. E. Martin & McElree, 2011; McElree et al., 2003). 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. 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 low availability can lead to a misretrieval or a failure in parsing. In a certain proportion of trials with a failed initial retrieval, a process of backtracking (or reanalysis) is triggered, which leads to correct retrieval of the target. This process requires some extra time (A. E. Martin & McElree, 2008; McElree et al., 2003). Our modified direct-access model assumes that backtracking can also 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 stands for the probability of backtracking and 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.[4]

The key difference between the activation-based model and the modified-direct model is that the latter assumes retrieval times to be unaffected by interference. 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 be slowed in IWA compared to controls. The noise parameter, σ, also has an adjustment by group, as IWA may have more

[4]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.

variable retrieval times.

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 $LogNormal(\mu, \sigma)$.

(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 $LogNormal(\mu, \sigma)$.

Notice that the probability of successful retrieval of the target, θ, and the probability of backtracking, P_b are assumed to be independent.

Correct and incorrect responses following backtracking (b), are expected to be slower, on average, than correct retrievals (a) and misretrievals without backtracking (c), due to the cost of backtracking δ. The RT 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 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 estimate of 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 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, 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 (Adelt et al., 2017; Burchert et al., 2003) showing that OR are more difficult to process than SR, 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.

6.7 Modeling Results

6.7.1 Experiment 1

A graphical summary of the data in Experiment 1 is shown in Figure 6.4.

Activation-based model

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 6.5 shows the distribution of 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 (6.5a vs. 6.5b and 6.5c vs. 6.5d). 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 distributions within each plot. Whereas the distributions lie far apart from each other in controls (6.5a and 6.5c), in IWA, the distributions overlap, especially in the mismatch condition (6.5d). 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 $\times$ condition interaction (μ_{NP1}: 199 ms, CrI: [81, 322] ms and μ_{NP2}: 1424 ms, CrI: [400, 2720] ms).

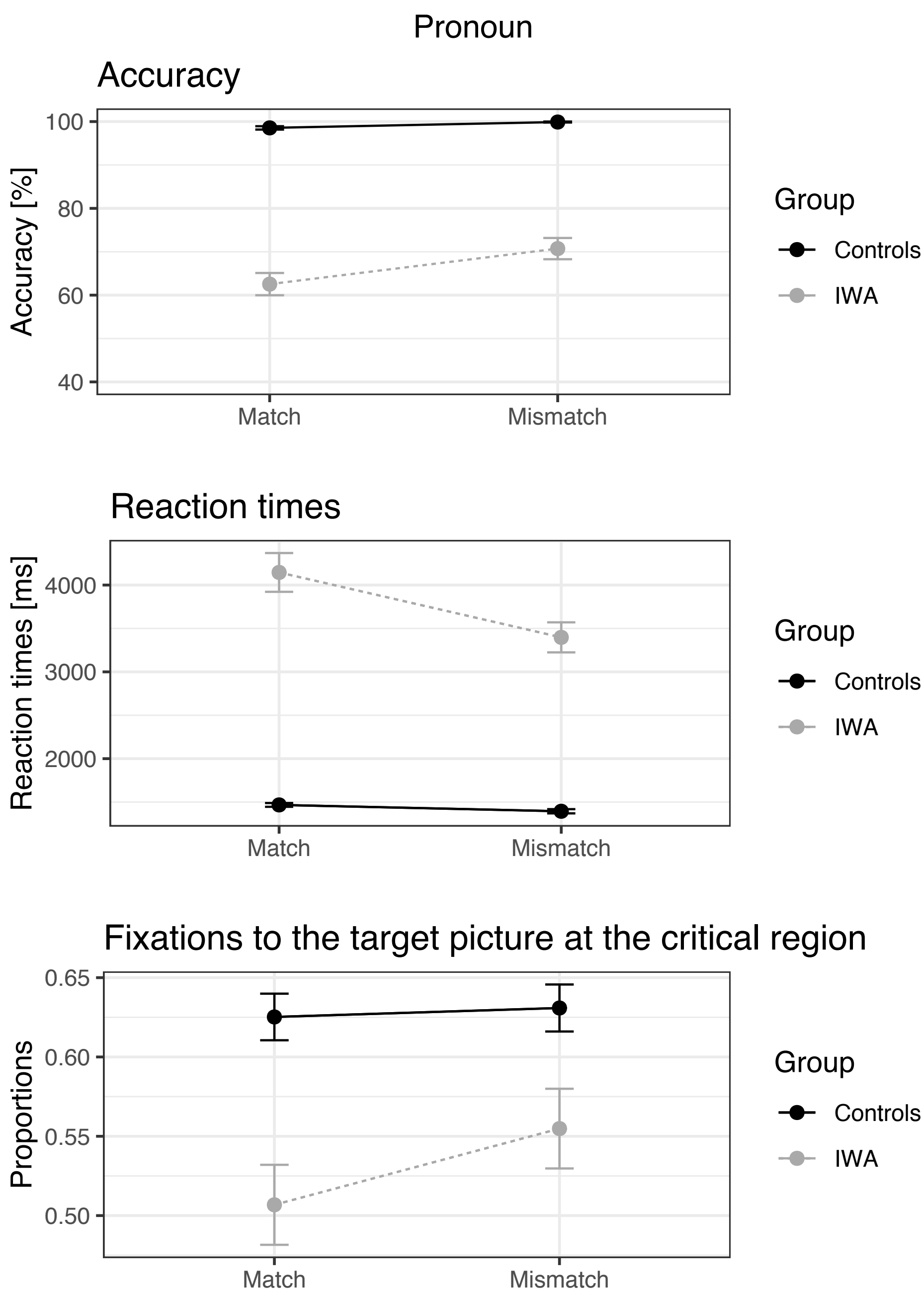

Figure 6.4: Descriptive statistics for the pronoun sub-experiment. The dots stand for the means, and the error bars show the standard error of the means.

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 Table 6.1.

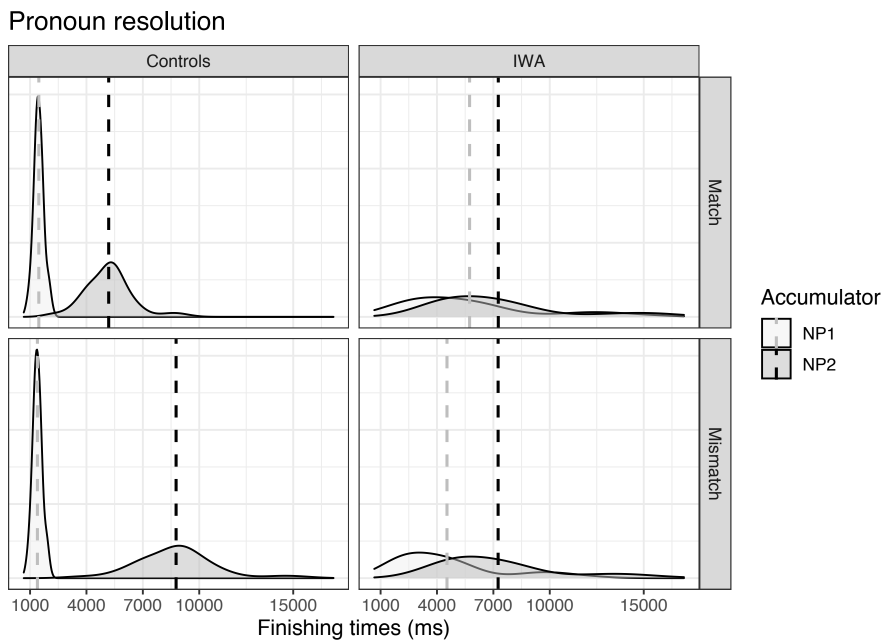

Figure 6.5: Distribution of the accumulators of evidence across groups and conditions for the pronoun conditions.

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: An increase in fixations on the target picture leads to an increase in RT 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 6.6 shows that the

Table 6.1: 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

probability of retrieval of the target is very high for controls: The mean of the distribution lies above 95% in both conditions (CrI mismatch: [96, 98]%, CrI match: [98, 99]%). By contrast, IWA show lower retrieval probabilities overall. This can be also seen in Figure 6.6, where IWA's mean estimate for mismatch is 55% CrI: [47, 62]%, whereas the estimate for match is 72% CrI: [66, 77]%. The group × condition interaction is inconclusive (2 CrI: [-1, 5]%).

A unit increase in LDT leads to θ: -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 some 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 probability of retrieval of the target after backtracking, θ_b, is shown in Figure 6.7. 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

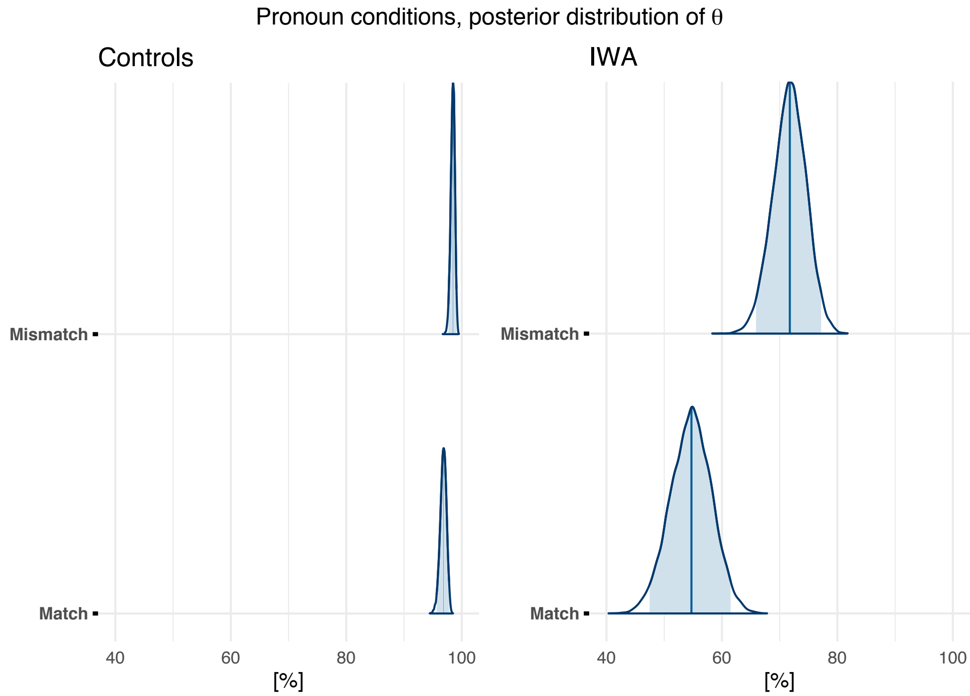

Figure 6.6: Posterior distribution of the probability of initial retrieval of the target, θ for the two groups, in the pronoun conditions. The vertical lines stand for the means of the distributions, and the shaded areas indicate the 95% credible interval.

ms, CrI: [1202,1447] ms); and IWA also have a larger estimate of noise (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.

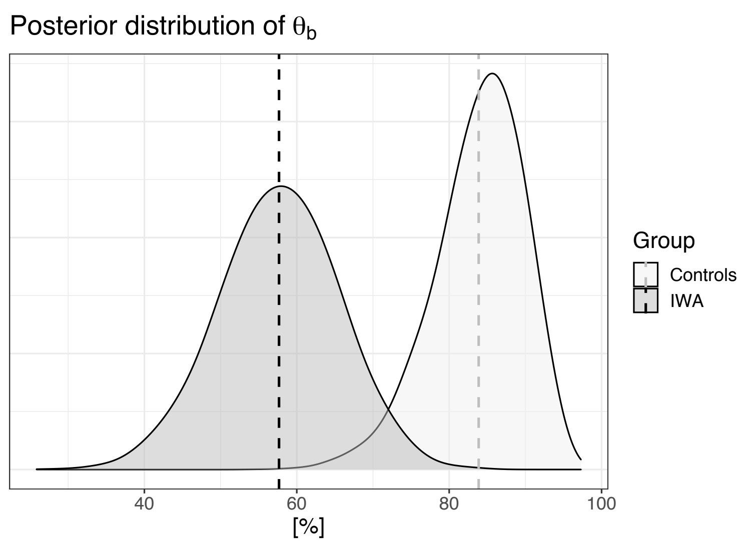

Figure 6.7: Posterior distribution of the probability of retrieval of the target after backtracking, θ_b, for the two groups, in the pronoun conditions. The dashed lines stand for the means of the distributions.

6.7.2 Experiment 2

A graphical summary of the data in Experiment 2 is provided in Figure 6.8.

Activation-based model.

In the relative-clause items, 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.

Figure 6.9 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 slightly higher in mismatch vs. 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. This suggests that in subject relatives, IWA retrieve the target more often than the distractor; yet IWA retrieve the distractor more often than controls.

Figure 6.10 shows the distribution of finishing times of the two accumulators in object relative clauses. IWA have higher finishing times than controls across conditions, and 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 Figure 6.10, 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 Figure 6.9 and Figure 6.10 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.

The model estimates for the fixed effects and interactions on μ_{NP1} and μ_{NP2} are shown in Table 6.2. No indication of an effect was found for condition or the condition $\times$ group interaction, but there was a RC type $\times$ condition interaction on μ_{NP2} (631

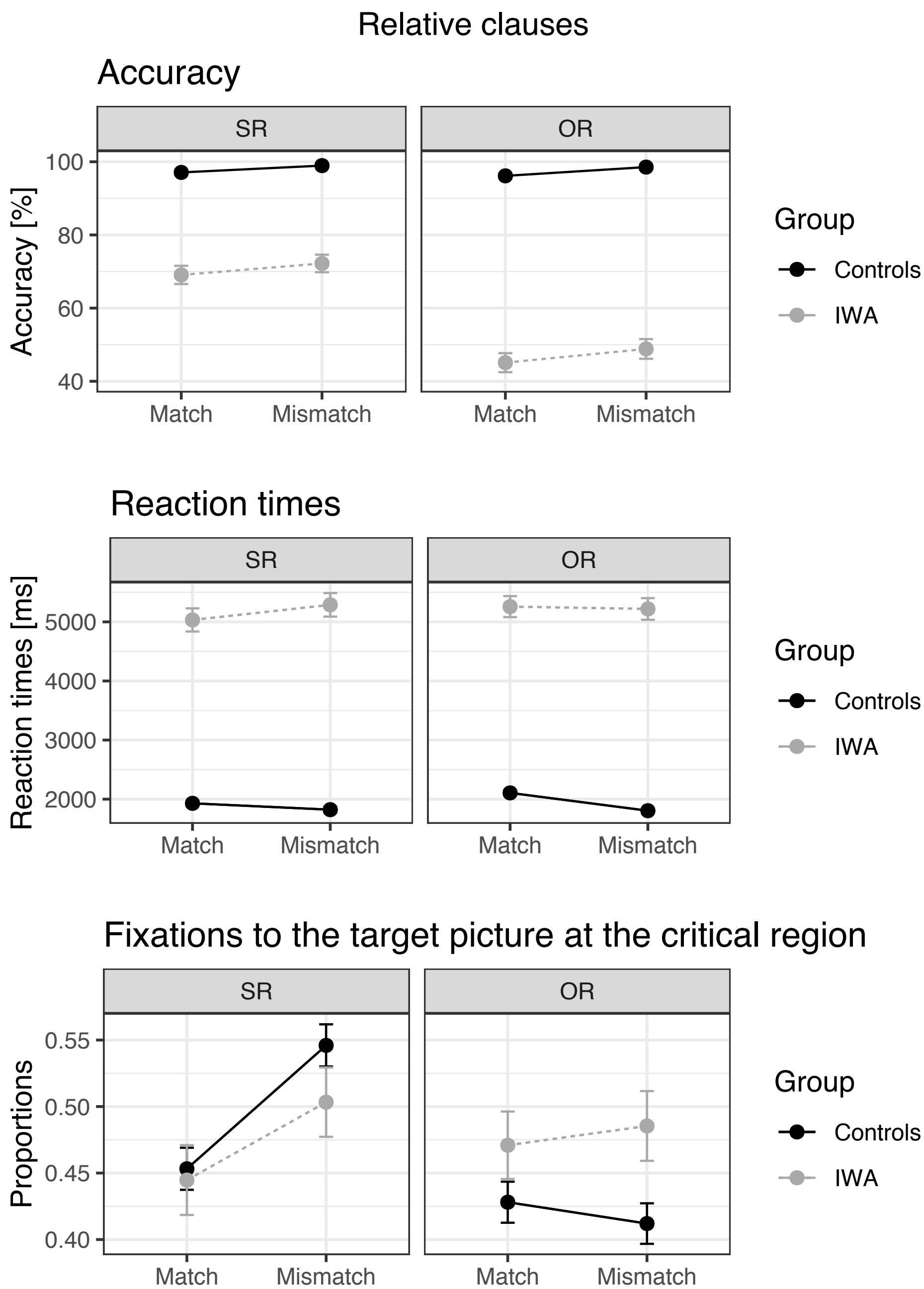

Figure 6.8: Descriptive statistics for the relative clauses sub-experiment. The dots stand for the means, and the error bars show the standard error of the means.

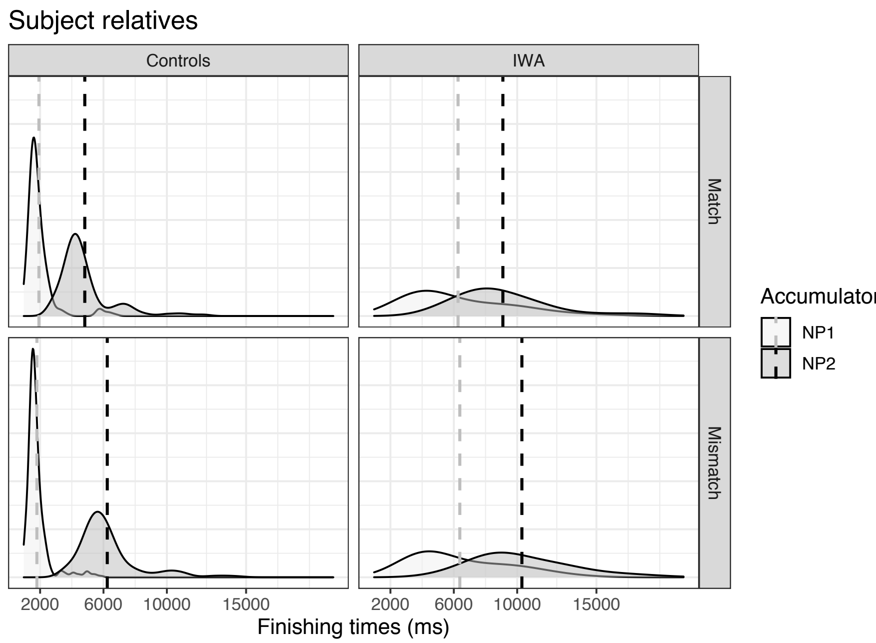

Figure 6.9: Distribution of the accumulators of evidence across groups and conditions for subject relative clauses. The dashed lines indicate the means of the distributions.

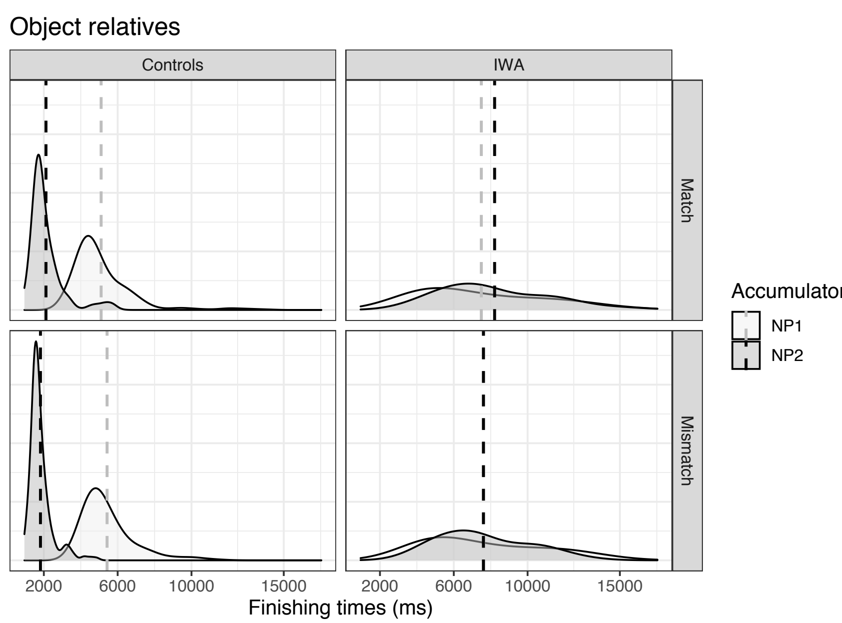

Figure 6.10: Distribution of the accumulators of evidence across groups and conditions for object relative clauses. The dashed lines indicate the means of the distributions. Note that in the mismatch condition, for IWA, the mean of the two distributions overlap.

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 $\times$ group $\times$ 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. Since the NP2 accumulator wins in half of the OR mismatch trials on average (see Figure 6.10), the number mismatch helps IWA to interpret OR correctly in about half of the OR mismatch trials.

Table 6.2: 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	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

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 $\times$ condition interaction or a fixations $\times$ group $\times$ 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 Figure 6.11. While controls have a slightly lower θ in OR relative to SR in the match conditions, both SR and OR have a similar θ in mismatch conditions, around 95%. This indicates that the number 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 what Figure 6.11 shows. The effect of condition (-10% CrI: [-13, -6]%) suggests that match conditions elicit a lower θ across the board, but a condition $\times$ group interaction (7% CrI: [3, 10]%) suggests that the effect of match is stronger for controls than for IWA. There was no indication of a RC type $\times$ condition interaction (-1% CrI: [-5, 2]%), nor of a RC type $\times$ condition $\times$ group interaction (1% CrI: [-3, 4]%). The effect of RC type (-13%, CrI: [-18, -9]%) and the RC type $\times$ 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 $\times$ LDT interaction (3%, CrI: [-2, 8]%). A unit increase in fixations led to -2%, CrI: [-5, 1]%) on θ, and the interaction group $\times$ 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 for controls it is 87% CrI: [83, 91]%. The posterior distribution of θ_b is shown in Figure 6.12: 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

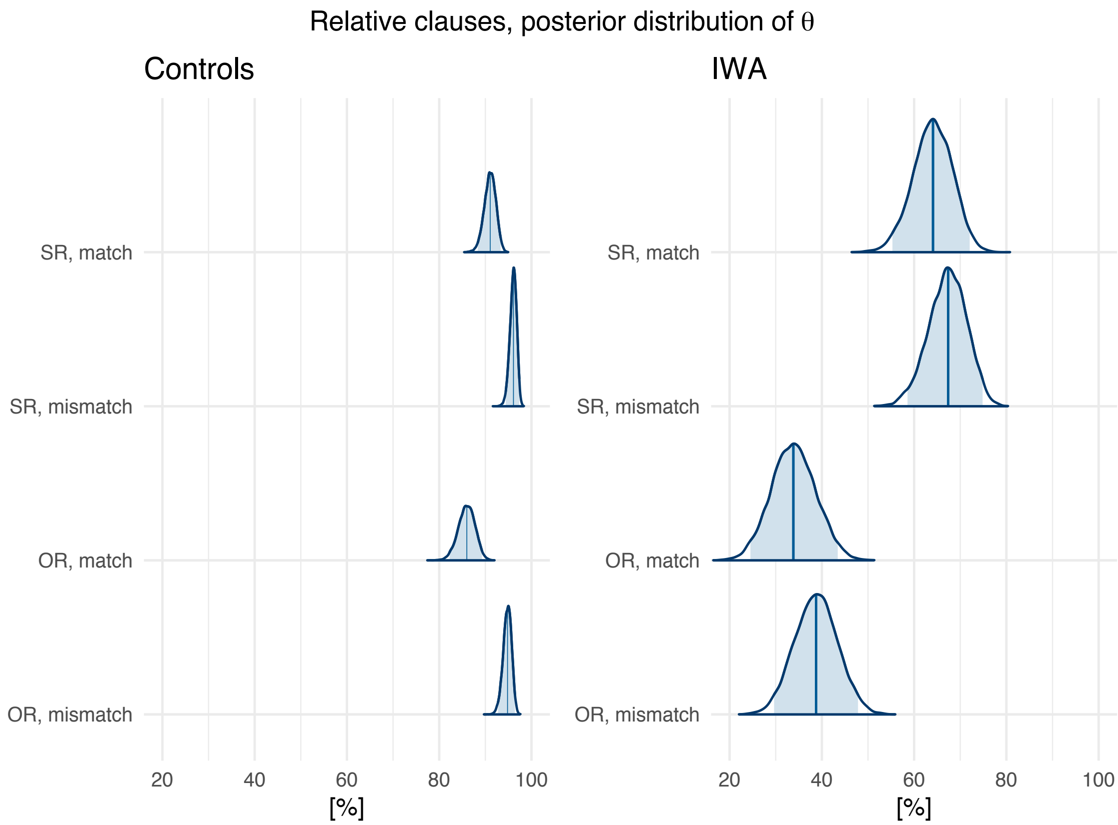

Figure 6.11: Posterior distribution of the probability of retrieval of the target (θ) across groups and conditions in relative clauses. The vertical lines stand for the means of the distributions, and the shaded areas represent the 95% credible interval.

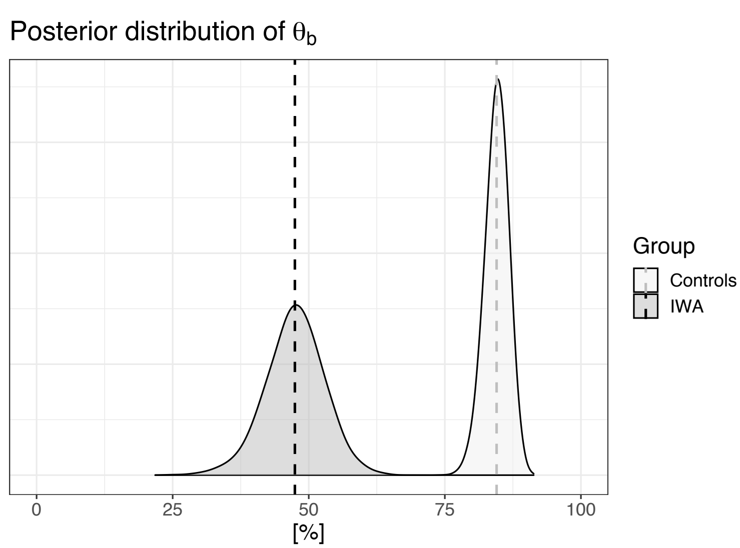

Figure 6.12: Posterior distribution of the probability of retrieval of the target after backtracking (θ_b) across groups in the relative clauses experiment.

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]).

6.8 Discussion

In the two models, the location and scale parameters of the log-normal distribution from which RT are sampled (μ and σ) were consistently higher for IWA than for controls. We linked these parameters to the slow syntax and intermittent deficiencies, 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.

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/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 show that IWA experience a subject-object asymmetry, i.e., IWA have more difficulties processing object relatives.

The conclusions for the rest of our predictions are more complex, since the results differ across the two experiments. For instance, a group $\times$ 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. Speculatively, it could be that a retrieval triggered by a pronoun is more strongly affected by delayed lexical access than one triggered by a verb: Our pronouns were directly coreferential with a previous noun phrase (*Peter* ... *er*), while the association between a verb and its arguments is more indirect.

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.

6.8.1 Modified direct-access model

We expected a lower probability θ of retrieval for the target in the match versus 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.

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 $\times$ LDT interaction lends some support to this theory in the pronoun resolution subexperiment, but not in relative clauses. Therefore, the effect of delayed lexical access in DA remains inconclusive. 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.

Finally, 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. 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.

6.9 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 fits the data better overall? In order to answer this question, we performed 10-fold cross-validation (Vehtari et al., 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 goodness of 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

Table 6.3: Differences in $\widehat{elpd}$ between the two models and their corresponding SE. Positive differences indicate an advantage for the activation-based model, whereas negative differences indicate an advantage for the modified direct-access model.

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

into 10 balanced subsets containing about the same amount of data per subject.[5] 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 6.3. In pronoun resolution, the MDA model has a predictive advantage, but since the SE of $\widehat{\Delta elpd}$ is large, the result is not conclusive. In relative clauses, the activation-based model has a clear advantage over MDA. In order to find out which group and/or condition is driving the advantage, we can compute $\widehat{\Delta elpd}$ separately for each group and each condition. The results are shown in Figure 6.13. It appears that the advantage of the activation-based model over MDA mostly comes from fitting the data of control participants, especially in the OR, mismatch conditions ($\widehat{\Delta elpd}$ 163, SE 66).

We also evaluated the predictive performance of the original direct-access model, i.e., a model in which backtracking can only lead to the retrieval of the target. 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,

[5] 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.

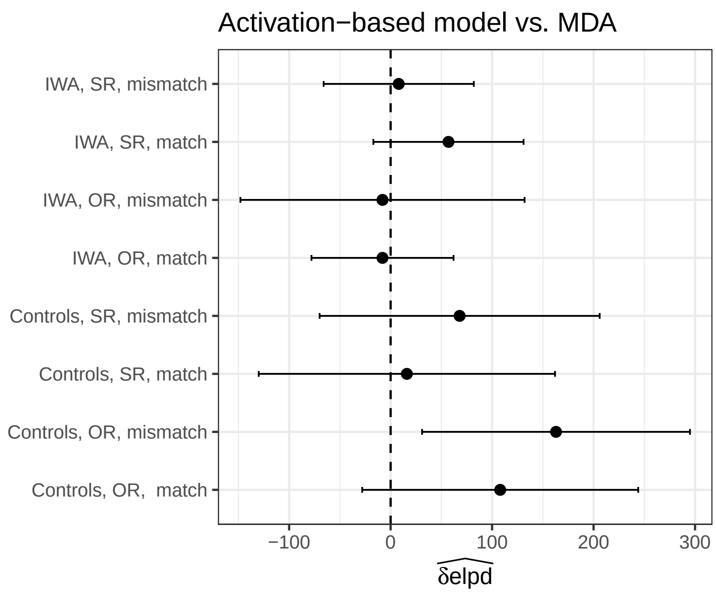

Figure 6.13: Graphical representation of the $\widehat{\Delta elpd}$ between the activation-based and the modified direct-access model across groups and conditions. The dot stands for the $\widehat{\Delta elpd}$ and the bars indicate to the 95% confidence interval. Positive values indicate an advantage for the activation-based model, and negative values indicate an advantage for the modified direct-access model.

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.

6.10 General discussion

This is the first-ever computational investigation of competing models of similarity-based interference in German language comprehension in individuals with aphasia 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 in 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 if 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 (Adelt et al., 2017; Hanne et al., 2015; Lissón, Pregla, et al., 2021; Pregla, Lissón, et al., 2021). The predictive performance of the two models was compared against data from a visual-world experiment (Pregla, Lissón, 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. 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 regards to the underlying processing deficits in aphasia, both models are in line with slow syntax (Burkhardt et al., 2008; Burkhardt et al., 2003) and intermittent deficiencies (Caplan et al., 2013). Resource reduction (Caplan, 2012; Caplan et al., 2007), 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 both models, that delayed lexical access (Ferrill et al., 2012; Love et al., 2008) is a source of processing deficits in IWA: The predicted relationship between individual lexical decision latency and participant group was not found in all 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 both models, we can conclude that in pronoun resolution, IWA are more sensitive to gender interference than control participants. 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 some possible explanations of the subject-object asymmetry in IWA and the comparatively small effect of number mismatch in relative clauses.

6.10.1 Subject-object asymmetry

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., Adelt et al., 2017; Burchert & De Bleser, 2004; Burchert et al., 2003; Hanne et al., 2011; Pregla, Lissón, et al., 2021). Canonicity effects refer to the fact that sentences with a non-canonical order (e.g., object-subject-verb in German) are more difficult to process than sentences with a canonical order. Canonicity effects could be orthogonal to the memory retrieval process, and could be caused by frequency, as canonical sentences are more frequent than non-canonical ones. However, neither the activation-based model, nor the (modified) direct-access model can account for frequency effects in sentence structures. An interesting test for both models would be to implement frequency structure effects and to test the models' predictions against data from IWA.

In contrast to English, where subject and object relative clauses are distinguished by word order, successful comprehension of German case-unambiguous relative clauses requires comprehenders to use grammatical case as a cue to correctly identify the agent and the theme of the verb; both in subject and in object relatives. Given that controls' performance was at ceiling in both relative clause types (see Figure 6.14), we conclude that controls were able to use case-marking cues accurately. This is consistent with the the results in the ERP study by Friederici et al. (1998). Our data also shows that IWA have difficulties understanding relative clauses, especially object relatives. These results are in line with previous studies testing the comprehension of subject vs. object relatives in German in IWA (Adelt et al., 2017; Burchert et al., 2003). However, cue-based retrieval cannot explain this subject-object asymmetry: 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. It is thus not clear why processing object relatives in German should be more difficult from a purely retrieval-based point of view. One way to explain the subject-object asymmetry would be to assume that IWA have difficulties deploying case features in relative clauses.

One possible explanation for the subject-object asymmetry in IWA is that IWA may be more sensitive to case attraction. Case attraction is based on the well-known phenomenon of number attraction and the feature percolation account (e.g., Eberhard, 1997; Nicol, Forster, & Veres, 1997). In relative clauses, case attraction posits that processing is facilitated when a pronoun and its antecedent have the same case (Bader & Meng, 1999). If the head noun and the relativizer mismatch in case, the case feature of the head noun could percolate down to the head noun, overriding its

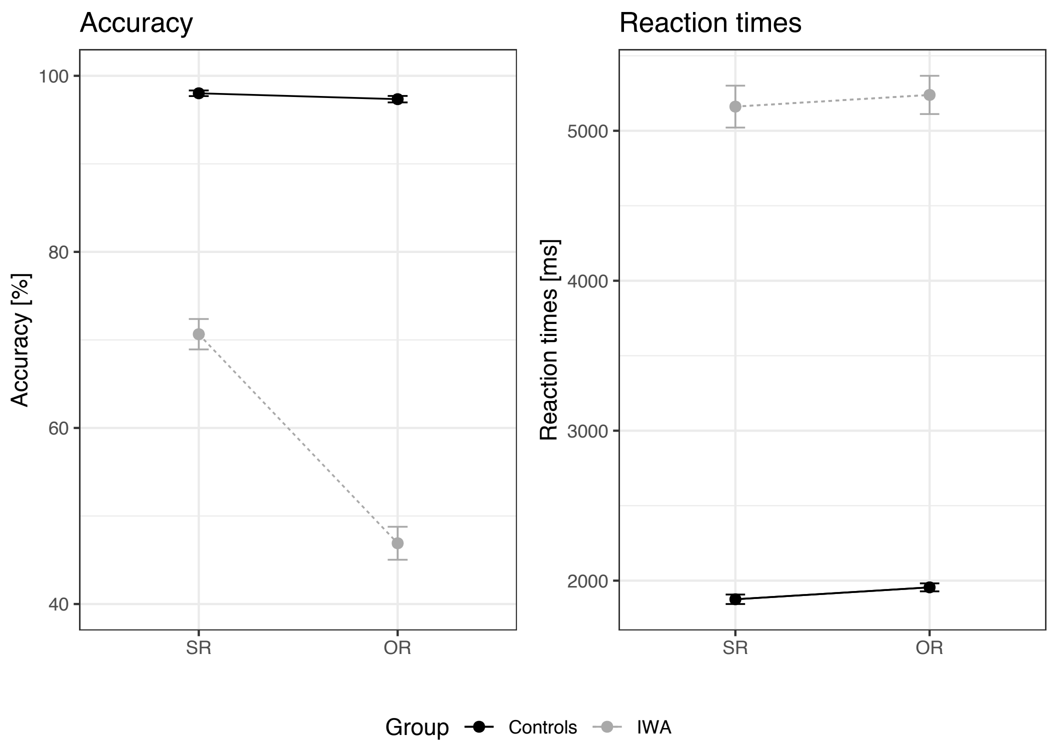

Figure 6.14: Descriptive statistics for the relative clauses experiment. The dots stand for the means, and the error bars show the standard error of the means.

original case, and creating a misinterpretation. Once the verb is reached, at the end of the sentence, and the subject needs to be retrieved, the misinterpretation in the case feature of the relativizer could lead to a costly reanalysis process. Bader and Meng (1999) and Bader, Meng, and Bayer (2000) claim that case attraction occurs with dative-marked pronouns. In their account, the feature [+dative] percolates upwards to the head noun. Czypionka, Dörre, and Bayer (2018) refer to this configuration as inverse case attraction. Other studies testing unimpaired populations in German have shown that case attraction can also happen with nominative/accusative cases, and that feature percolation can also happen downwards, from the head noun to the relativizer (Logačev and Vasishth, 2012, experiment 2; Schlesewsky, 1996, reported as experiment 1 in Fanselow, Schlesewsky, Cavar, and Kliegl, 1999). However, these studies tested complex center-embedded relative clauses that are very different from our items.

In our data, assuming that IWA have a larger effect of case attraction could explain the subject-object asymmetry. Consider again sentence (21b), repeated here as (22). In relative clauses with masculine head nouns, such as in our items, the case attraction theory would predict that subject relatives are easier to process because both the head noun and the relativizer have the [+nominative] feature.

+nom

(22) Hier ist [der Esel]$^{+nom}_{+sing}$, den$^{+acc}_{+sing}$ [der Tiger]$^{+nom}_{+sing}$ gerade badet$^{nom}_{sing}$.
Here is the$_{\text{NOM}}$ donkey who$_{\text{ACC}}$ the$_{\text{NOM}}$ tiger now bathes.
'Here is the donkey who the tiger bathes.'

The first noun phrase, *der Esel*, has the features [+nominative, +singular]. The relativizer, *den*, has the features [+accusative, +singular]. However, due to case attraction, the relativizer could end up encoded in memory with features [+nominative, +singular]. The second noun phrase, *der Tiger*, also has features [+nominative, +singular]. At the verb *badet* (bathes), when the retrieval of the subject is triggered with cues [nominative, singular], both noun phrases would share the retrieval cues [+nominative, +singular]. In both the activation-based and the modified direct-access model, this configuration would lead to misretrievals of the distractor. In the modified direct-access model, after this initial misretrieval, in a certain proportion of the trials, IWA could initiate a backtracking process, which would fail more than half of the time (as estimated by θ_b). Therefore, in combination with a cue-based retrieval mechanism, in IWA, case attraction would predict more misretrievals and slower processing times in object relatives compared to subject relatives. Control participants may experience case attraction less often than IWA, and if an initial misretrieval occurs, they can backtrack and correctly retrieve the target with a very high probability. Thus, case attraction could also explain the pattern in the data of control participants, which show slightly lower accuracies and slightly higher RT in object compared to subject relatives, as shown in Figure 6.14. The role of feature percolation in the activation-based model has been recently investigated by Yadav, Smith, and Vasishth (2021). Their results support the view that in unimpaired populations, feature percolation feeds retrieval processes, suggesting that a hybrid model may explain similarity-based interference better. It remains to be seen whether such hybrid model would provide a better fit for data from individuals with aphasia.

Alternatively, the subject-object asymmetry could also be explained by the noisy channel or rational inference account of sentence comprehension (Gibson, Bergen, & Piantadosi, 2013). This account claims that sentence comprehension involves a Bayesian computation of the probabilities of possible intended sentences given a perceived sentence that may have undergone distortions, for instance because of noise in the processing system. The impairments in IWA are assumed to increase the likelihood of sentence distortions due to an increased amount of noise (Gibson, Sandberg, Fedorenko, Bergen, & Kiran, 2016; Warren, Dickey, & Liburd, 2017), which could lead IWA to rely more on their prior probability of a sentence interpretation. Prior sentence expectations are based on construction frequency and world knowledge. Essentially, because the incoming perceptual data are noisy, IWA are predicted to use the prior to infer that the speaker probably intended the more frequent subject-relative structure.

6.10.2 Number mismatch

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 6.8). The MDA model estimates no effect of number mismatch for IWA. Although the activation-based model estimates that number mismatch between the NPs slightly increases the probability of retrieving the target in object relatives, the target is only retrieved about half the time in these conditions. It thus seems that for IWA, the effect of RC type is greater than the effect of number interference.

One possible explanation for the differential effects of number marking and RC type is that IWA may not weight case and number cues equally. The cue-weighting proposal, implemented by Engelmann (2016) in the framework of the LV05 model, claims that depending on the linguistic structure, some retrieval cues may be weighted higher than others (see also Engelmann et al., 2019; Vasishth et al., 2019). This assumption is motivated by research in individual differences in the memory literature, which has shows that some individuals learn to use certain retrieval cues more efficiently than others (e.g., Danker, Fincham, & Anderson, 2011). In sentence processing, it has been argued that structural and non-structural cues could be weighted differently (e.g., Cunnings & Sturt, 2014; Dillon et al., 2013; Parker & Phillips, 2017). Differences in cue weighting could be integrated in both the activation-based and the direct-access model, as both models rely on retrieval cues. Speculatively, it could be that IWA weight case and number cues differently, and that the retrieval in relative clauses is mostly guided by case cues. Studies investigating processing of number and case in IWA in German provide mixed results. 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, and Nicole (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 in 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. By contrast, our modeling shows that when both case and number cues are needed to interpret a sentence, in IWA, the effect of case overrides the effect of number.

Recently, a Bayesian version of the Lewis and Vasishth (2005) model has been developed that includes an estimation of cue-weighting at the individual level (Yadav, Paape, Smith, Dillon, & Vasishth, 2021). An interesting future direction would be to investigate whether differences in cue-weighting between individual IWA and control participants could account for our relative clause data. Similarly, it would be interesting to develop a hybrid model that accounts for both case attraction and cue weighting, along the lines of the work by Yadav, Smith, and Vasishth (2021).

6.10.3 Comparison with previous work

In Chapter 4, we investigated English relative clause processing in IWA vs controls using a large-scale dataset from Caplan et al. (2015). We 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 et al., 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, and as soon as they hear the relevant morphological cues (e.g., the relativizer in unambiguous German relative clauses), they start looking at the target picture. In Chapter 4, where self-paced listening data was modeled, 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 more often the distractor 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 the data in the present chapter, 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, Chapter 4 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 SE of the $\widehat{\delta elpd}$. In the present study, we show that in relative clauses in German, the activation-based model furnishes 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 (i.e., trials with correct responses to the comprehension question), RT were on average higher than the RT in incorrect trials. This pattern in Nicenboim's data is crucial, because higher RT 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 shows the opposite pattern. Correct responses are, on average, slower than incorrect ones, 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 (Lissón et al., 2021; Mätzig et al., 2018; Patil et al., 2016). 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.

6.10.4 Limitations and future directions

One important limitation of this work is the amount of data used for computational modeling. The data that we modeled in this paper (Pregla, Lissón, 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. Unfortunately, collecting online data from impaired populations is extremely 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., Adelt et al., 2017; Burchert & De Bleser, 2004; Burkhardt et al., 2003; Choy & Thompson, 2010; Dickey & Thompson, 2009; Engel et al., 2018; Hanne et al., 2015; Hanne et al., 2011; Love et al., 2008; Mack, Ji, & Thompson, 2013; Mack et al., 2016). 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 fixations. 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 (Mätzig et al., 2018; Patil et al., 2016; Pregla, Lissón, 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, Paape, 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.

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 leftcorner 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.

6.11 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 versus 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 suggests that IWA exhibit a subject-object asymmetry, whereas controls process subject and object relatives similarly. The effect of number interference in IWA seems to be overridden by the subject-object asymmetry, 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 a similar quantitative performance for the pronoun structures, the activation-based model outperforms the modified direct-access model in relative clauses.

Chapter 7

Conclusions

This dissertation had two main goals. The first goal was to computationally investigate the source of processing deficits in sentence comprehension in aphasia. The second goal was to compare the performance of two competing models of cue-based retrieval in impaired and unimpaired sentence comprehension.

Two models of cue-based retrieval have been implemented in the Bayesian framework. These models build on the original implementations by Nicenboim and Vasishth (2018). A series of prominent theories of processing deficits, from the aphasia literature, have been mapped onto the different parameters of the models. The theories of processing deficits that have been investigated here (detailed in Chapter 2), include delayed lexical access (Ferrill et al., 2012), slow syntax (Burkhardt et al., 2003), resource reduction (Caplan, 2012), and intermittent deficiencies (Caplan et al., 2007).

Chapter 4 shows that both models of cue-based retrieval perform similarly at predicting the performance of IWA and controls in self-paced listening data from English relative clauses. However, the activation-based model had a slight advantage in predictive performance, because the direct-access model could not account for the incorrect, slow responses from IWA. In the activation-based model, the most likely sources of processing deficits are slow syntax and/or delayed lexical access, and intermittent deficiencies. In the direct-access model, all of the previous deficits plus resource reduction could be playing a role. This chapter reveals that both models can account for data from IWA and control participants, although the direct-access model has some limitations. This chapter confirms the conclusions in previous computational work in aphasia: IWA experience a combination of several processing deficits that cause sentence comprehension difficulties (Mätzig et al., 2018; Patil et al., 2016).

In Chapter 5, a modified version of the direct-access model has been developed. The implementation of this augmented model addresses the major caveat that Chapter 4 had revealed; namely, that the original direct-access model cannot account for slow incorrect trials. The original and the modified models were compared using self-paced listening data from items testing the comprehension of control structures in German. Although the generative process of the modified direct-access model allows for slow incorrect responses, model comparisons using Bayes factors resulted inconclusive. Both the original and the modified direct-access model presented a similar fit to the data tested. The results of both models are in line with delayed lexical access, slow syntax, intermittent deficiencies, and resource reduction. In addition, the results revealed that IWA experience a more pronounced object-subject asymmetry in control structures, relative to control participants.

Chapter 6 compared the activation-based model against the modified direct-access model developed in chapter 5. The data modeled in this chapter, which came from a visual-world experiment, contain two linguistic constructions in German: Pronoun resolution and relative clauses. Reaction times and accuracies at the picture selection task were modeled, and the proportions of looks to the target were taken as a predictor in the models. This chapter shows that the activation-based model outperforms both the original and the modified direct-access models in relative clauses. In pronoun resolution, the three models have a similar performance. Importantly, the relative clauses data modeled in this chapter contain double amount of items (20) compared to the data modelled in previous chapters. This indicates that data sparsity may be the reason behind the inconclusive results in the previous chapters.

Taking together the conclusions from the three modeling chapters, our results suggest that (a) an activation-based mechanism of cue-based retrieval may underlie sentence comprehension in individuals with aphasia and unimpaired controls; (b) within this activation-based mechanism, two processing deficits play an important role: Slow syntax, and intermittent deficiencies. In other words, a generally slower accrual of activation, and noise in the parsing system are the best way in which an activation-based model can account for data from IWA. The role of delayed lexical access remains inconclusive, and more studies are needed in order to understand the extent to which this deficit may also play a role in an activation-based mechanism of cue-based retrieval.

With regards to the (modified) direct-access model, more research is needed, especially with a larger sample size, in order to better understand the behavior of the model. Our results suggest that the activation-based model outperforms the direct-access model, but it remains to be seen whether this could change with a larger sample size. Although model comparisons between the original and the modified direct-access model are inconclusive, we argue that the modified direct-access model should be taken as the default model for IWA, based on theoretical grounds. However, the risk of using this model with sparse data is that the model may end up being over-parametrized. The modified direct-access model is even more complex than the direct-access model because it contains one more latent variable (the probability θ_b), which may be difficult to estimate with sparse data. Unfortunately, gathering data from IWA is a time-consuming and extremely difficult task. From the modeling perspective, one could, in principle, add as much data as possible within a single model (e.g., the 13 conditions tested in Caplan et al., 2015 instead of focusing on a single pair of conditions). However, in practice, this is very difficult to achieve. For example, in chapter 6 it was initially planned to model pronoun resolution and relative clauses within the same model. However, more conditions require more fixed and random effects, which generates a more complex model structure. Unfortunately, such complex model structures lead to convergence issues with the (modified) direct-access model.

Overall, this work has shown that theory can be advanced by computational modeling. As an example, consider the following. The direct-access model was developed in the speech accuracy trade-off paradigm, a complex experimental paradigm in which IWA cannot be tested. By specifying and formalizing the assumptions of the direct-access model in different experimental paradigms, the present work has shown that that the original assumptions in the direct-access model are at odds with data from IWA. A modified version of the direct-access model has been proposed as an alternative model, more suitable for IWA's performance. This theoretical development had not been possible without computational modeling, which shows that formal modeling leads to specific and more transparent theorizing (Guest & Martin, 2021). Similarly, the systematic modeling approach developed here opens the door to testing more competing theories of processing deficits in aphasia. One obvious caveat of this systematic modeling approach is that one can only infer theoretically meaningful insights through the lenses of the model. A cognitive model is an oversimplification of the human cognitive system, and the limits of basing theory development on computational modeling are in the specific assumptions of the computational model itself. Yet, computational models are a useful tool to check whether verbally-stated theories about a theorized system match the realizations of this system, i.e., the experimental data (Farrell & Lewandowsky, 2018, p.21). We have tried to be explicit regarding the modeling assumptions in the hope that the direct mapping between theoretical constructs and model parameters is transparent. Future avenues of research should broaden the theoretical scope of this work by adding, for example, a joint mechanism between retrieval interference and encoding interference, as suggested in Chapter 6, which has never been tested with data from IWA.

In conclusion, the work in this dissertation contributes to the aphasia literature by showing that the activation-based model of sentence processing, originally designed for unimpaired populations, can account for sentence processing in IWA by assuming a slower accrual of activation, and noise in the processing system. These two deficits, which we map to the slow syntax and intermittent deficiencies theories in the aphasia literature, disrupt the retrieval process, causing delays in sentence processing, and misinterpretations. The results in this dissertation have also shown that the assumptions of the original direct-access model are at odds with the general performance of IWA reported in the sentence processing literature in aphasia. A modified version of the direct-access model that is theoretically more suitable for modeling sentence comprehension in aphasia has been developed.

References

Adelt, A., Burchert, F., Adani, F., & Nicole, S. (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, November). Feature dissimilarities in the processing of German relative clauses in aphasia. *Journal of Neurolinguistics*, *44*, 17–37.

Alexander, M. P., & Hillis, A. E. (2008). Aphasia. *Handbook of clinical neurology*, *88*, 287–309.

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.

Badecker, W., & Caramazza, A. (1985). On considerations of method and theory governing the use of clinical categories in neurolinguistics and cognitive neuropsychology: The case against agrammatism. *Cognition*, *20*(2), 97–125.

Badecker, W., & Caramazza, A. (1986). A final brief in the case against agrammatism: The role of theory in the selection of data. *Cognition*, *24*(3), 277–282.

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

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

Bader, M., Meng, M., & Bayer, J. (2000). Case and reanalysis. *Journal of Psycholinguistic Research*, *29*(1), 37–52.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Berndt, R. S., & Caramazza, A. (1980). A redefinition of the syndrome of Broca's aphasia: Implications for a neuropsychological model of language. *Applied Psycholinguistics*, *1*(3), 225–278.

Berndt, R. S., Mitchum, C. C., & Haendiges, A. N. (1996). Comprehension of reversible sentences in "agrammatism": A meta-analysis. *Cognition*, *58*(3), 289–308.

Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. (ArXiv preprint 1701.02434)

Broca, P. (1863). Localisation des fonctions cérébrales: Siége de langage articulé. *Bulletins de la Société d'Anthropologie de Paris*, *4*, 200–208.

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.

Burkhardt, P., Avrutin, S., Piñango, M. M., & Ruigendijk, E. (2008). Slower-than-normal syntactic processing in agrammatic Broca's aphasia: Evidence from Dutch. *Journal of Neurolinguistics*. The Left Periphery of Sentences, *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.

Burzio, L. (1986). *Italian syntax: A government-binding approach*. Dordrecht: Reidel.

Caplan, D. (2001). Aphasia. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 581–585).

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

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*(2), 117–175.

Caplan, D., DeDe, G., & Michaud, J. (2006). Task-independent and task-specific syntactic deficits in aphasic comprehension. *Aphasiology*, *20*(9), 893–920.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension*.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain*, *119*(3), 933–949.

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. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94.

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. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, *21*(1), 9–20.

Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language*, *76*(2), 158–184.

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).

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

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

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT press.

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.

Danker, J. F., Fincham, J. M., & Anderson, J. R. (2011). The neural correlates of competition during memory retrieval are modulated by attention to the cues. *Neuropsychologia*, *49*(9), 2427–2438.

De Bleser, R., Burchert, F., Holzinger, P., & Weidlich, C. (2012). Agrammatism at the sentence level: The role of morphology and prosody. In C. K. Thompson & R. Bastiannse (Eds.), *Perspectives on Agrammatism* (pp. 120–135). London: Psychology Press.

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, *21*(2-4), 125–145.

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.

Dronkers, N. F., Wilkins, D. P., Van Valin Jr, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1-2), 145–177.

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.

Eling, P., & Whitaker, H. (2009). History of aphasia: From brain to language. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of clinical neurology* (Vol. 95, pp. 571–582).

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.

Evans, W. S., Hula, W. D., & Starns, J. J. (2019). Speed-accuracy trade-offs and adaptation deficits in aphasia: Finding the "sweet spot" between overly cautious and incautious responding. *American Journal of Speech-Language Pathology*, *28*(1S), 259–277.

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.

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. New York: Cambridge University Press.

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.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641–666.

Friederici, A. D., & Kilborn, K. (1989). Temporal constraints on language processing: Syntactic priming in Broca's aphasia. *Journal of Cognitive Neuroscience*, *1*(3), 262–272.

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 and Hall/CRC.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, *2000*, 95–126.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Gibson, E., Sandberg, C., Fedorenko, E., Bergen, L., & Kiran, S. (2016). A rational inference approach to aphasic language comprehension. *Aphasiology*, *30*(11), 1341–1360.

Gigley, H. (1983). HOPE–AI and the dynamic process of language behavior. *Cognition & Brain Theory*, *6*(1), 39–88.

Gigley, H. (1988). Process synchronization, lexical ambiguity resolution, and aphasia. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution* (pp. 229–267).

Goodglass, H., Blumstein, S. E., Gleason, J. B., Hyde, M. R., Green, E., & Statlender, S. (1979). The effect of syntactic encoding on sentence comprehension in aphasia. *Brain and Language*, *7*(2), 201–209.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Bdae-3: Boston diagnostic aphasia examination*. Philadelphia, PA: Lippincott Williams & Wilkins.

Goodglass, H., & Wingfield, A. (1998). The changing relationship between anatomic and cognitive explanation in the neuropsychology of language. *Journal of Psycholinguistic Research*, *27*(2), 147–165.

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.

Grillo, N. (2009). Generalized minimality: Feature impoverishment and comprehension deficits in agrammatism. *Lingua*, *119*(10), 1426–1443.

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

Grodzinsky, Y. (1995). A restrictive theory of agrammatic comprehension. *Brain and Language*, *50*(1), 27–51.

Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry*, *24*(1), 69–101.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). Bridgesampling: An R package for estimating normalizing constants. *arXiv preprint arXiv:1710.08162*.

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, *16*(4), 789–802.

Gutman, R., DeDe, G., Caplan, D., & Liu, J. S. (2011). Rasch model and its extensions for analysis of aphasic deficits in syntactic comprehension. *Journal of the American Statistical Association*, *106*(496), 1304–1316.

Gutman, R., DeDe, G., Michaud, J., Liu, J. S., & Caplan, D. (2010). Rasch models of aphasic performance on syntactic comprehension tests. *Cognitive Neuropsychology*, *27*(3), 230–244.

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language*, *59*(1), 76–120.

Haarmann, H. J., & Kolk, H. H. J. (1991a). A computer model of the temporal course of agrammatic sentence understanding: The effects of variation in severity and sentence complexity. *Cognitive Science*, *15*(1), 49–87.

Haarmann, H. J., & Kolk, H. H. J. (1991b). Syntactic priming in Broca's aphasics: Evidence for slow activation. *Aphasiology*, *5*(3), 247–263.

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., Burchert, F., & Vasishth, S. (2016). On the nature of the subject–object asymmetry in wh-question comprehension in aphasia: Evidence from eye tracking. *Aphasiology*, *30*(4), 435–462.

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.

Harley, T. A. (2013). *The psychology of language: From data to theory*. London: Psychology press.

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, *3*, 292.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(47), 1593–1623.

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.

Jeffreys, H. (1939/1998). *The theory of probability*. Oxford University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science*, *1*(3), 273–290.

Kolk, H. H. J., & Van Grunsven, M. M. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology*, *2*(4), 347–384.

Kwon, N., & Sturt, P. (2016). Processing control information in a nominal control construction: An eye-tracking study. *Journal of Psycholinguistic Research*, *45*(4), 779–793.

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.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Leimkuhler, B., & Reich, S. (2004). *Simulating Hamiltonian dynamics*. Cambridge University Press.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

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

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., Paape, D., Pregla, D., Burchert, F., Stadie, N., & Vasishth, S. (2021). Similarity-based interference in sentence comprehension in aphasia: A computational evaluation of two models of cue-based retrieval. (Submitted to Journal of Memory and Language.)

Lissón, P., Pregla, D., Nicenboim, B., Paape, D., van het Nederend, M. L., Burchert, F., ... Vasishth, S. (2021). 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. (2021). 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). ACL.

Logačev, P., & Vasishth, S. (2012). Case matching and conflicting bindings interference. In *Case, word order and prominence.* (Vol. 40, pp. 187–216). Dordrecht: Springer.

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.

Luzzatti, C., Toraldo, A., Guasti, M. T., Ghirardi, G., Lorenzi, L., & Guarnaschelli, C. (2001). Comprehension of reversible active and passive sentences in agrammatism. *Aphasiology*, *15*(5), 419–441.

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.

Marin, O. S., Saffran, E. M., & Schwartz, M. F. (1976). Dissociations of language in aphasia: Implications for normal function. *Annals of the New York Academy of Sciences*, *280*(1), 868–884.

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.

Martin, R. C. (2006). The neuropsychology of sentence processing: Where do we stand? *Cognitive Neuropsychology*, *23*(1), 74–95.

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.

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.

Meyer, A. M., Mack, J. E., & Thompson, C. K. (2012). Tracking passive sentence comprehension in agrammatic aphasia. *Journal of Neurolinguistics*, *25*(1), 31–43.

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

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53–68.

Müller, S. (2002). *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German.* Studies in Constraint-Based Lexicalism. Stanford: Center for the Study of Language and Information.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (Vol. 2, *11*, pp. 116–162). Chapman & Hall / CRC.

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, November). 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, June). 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.

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).

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

Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. In *Proceedings of the Berkeley Linguistics Society* (Vol. 4, pp. 157–190).

Piñango, M. M., & Burkhardt, P. (2005). Pronominal interpretation and the syntax-discourse interface: Real-time comprehension and neurological properties. In A. Branco, T. McEnery, & R. Mitkov (Eds.), *Anaphora processing: Linguistic, cognitive and computational models* (pp. 221–238).

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Prather, P. (1994). The time course of lexical activation in fluent and nonfluent aphasia. In D. Hillert (Ed.), *Linguistics and cognitive neuroscience: Theoretical and empirical studies on language disorders* (pp. 128–144). Wiesbaden: VS Verlag für Sozialwissenschaften.

Prather, P., Zurif, E., Love, T., & Brownell, H. (1997). Speed of lexical activation in nonfluent Broca's aphasia and fluent Wernicke's aphasia. *Brain and Language*, *59*(3), 391–411.

Prather, P., Zurif, E., Stern, C., & Rosen, T. J. (1992). Slowed lexical access in nonfluent aphasia: A case study. *Brain and Language*, *43*(2), 336–348.

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. (2021). A visual world study of sentence processing in aphasia in German: The resource reduction hypothesis revisited. (Submitted to Cognitive Neuropsychology)

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Rouder, J. N. (2005). Are unshifted distributional models appropriate for response time? *Psychometrika*, *70*(2), 377–381.

Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, *25*(1), 102–113.

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*(2), 491–513.

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.

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

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.

Sivula, T., Magnusson, M., & Vehtari, A. (2020). Unbiased estimator for the variance of the leave-one-out cross-validation estimator for a Bayesian normal model with fixed variance. *arXiv preprint arXiv:2008.10859*.

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-Verlag.

Stan Development Team. (2021a). RStan: The R interface to Stan (Version 2.21.2).

Stan Development Team. (2021b). Stan reference manual, version 2.27.

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.

Stern, C., Prather, P., Swinney, D., & Zurif, E. (1991). The time course of automatic lexical access and aging. *Brain and Language*, *40*(3), 359–372.

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*. Mannheim: Institut für Deutsche Sprache.

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

Swinney, D., Zurif, E., Prather, P., & Love, T. (1996). Neurological distribution of processing resources underlying language comprehension. *Journal of Cognitive Neuroscience*, *8*(2), 174–184.

Thompson, C. K., & Choy, J. J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research*, *38*(3), 255–283.

Thompson, C. K., Choy, J. J., Holland, A., & Cole, R. (2010). Sentactics: Computer-automated treatment of underlying forms. *Aphasiology*, *24*(10), 1242–1266.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

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.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass*, *6*(4), 193–211.

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, October 1). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263.

van het Nederend, M. L. (2018). *Comparing competing models of retrieval processes: A bayesian approach to sentence processing* (Unpublished MSc thesis. Universiteit Utrecht, The Netherlands).

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. (2021). *Sentence comprehension as a cognitive process: A computational approach*. Cambridge, UK: Cambridge University Press.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 141–161.

Vasishth, S., Nicenboim, B., Chopin, N., & Ryder, R. (2017). Bayesian hierarchical finite mixture models of reading times: A case study. (PsyArXiv preprint psyarxiv.com/a4hs9)

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.

Walker, G. M., Hickok, G., & Fridriksson, J. (2018). A cognitive psychometric model for assessment of picture naming abilities in aphasia. *Psychological Assessment*, *30*(6), 809–826.

Warren, T., Dickey, M. W., & Liburd, T. L. (2017). A rational inference approach to group and individual-level sentence comprehension performance in aphasia. *Cortex*, *92*, 19–31.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis*. Breslau: Cohn & Weigert.

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. (Submitted to Open Mind)

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*.

Zurif, E., Swinney, D., Prather, P., & Love, T. (1994). Functional localization in the brain with respect to syntactic processing. *Journal of Psycholinguistic Research*, *23*(6), 487–497.

Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language*, *45*(3), 448–464.

Appendix

- The code and data for chapter 4 are available at https://osf.io/kdjqz/
- The code and data for chapter 5 are available at https://osf.io/spjer/
- The code and data for chapter 6 are available at https://osf.io/2aetr/