Similarity-based Interference and Faulty Encoding Accounts of Sentence Processing

Anna Laurinavichyute

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.

The Faculty of Human Sciences

University of Potsdam

2020

Supervisors:

Prof. Dr. Shravan Vasishth and Dr. Titus von der Malsburg

Submitted: August 5, 2020
Defended: June 8, 2021

Reviewers:
Prof. Dr. Shravan Vasishth, University of Potsdam, Germany
Prof. Dr. Dan Parker, College of William & Mary, USA

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-50966
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-509669

Similarity-based Interference and Faulty Encoding Accounts of Sentence Processing

Anna Laurinavichyute

Abstract

The goal of this dissertation is to empirically evaluate the predictions of two classes of models applied to language processing: the similarity-based interference models (Lewis & Vasishth, 2005; McElree, 2000) and the group of smaller-scale accounts that we will refer to as faulty encoding accounts (Bock & Eberhard, 1993b; Eberhard et al., 2005). Both types of accounts make predictions with regard to processing the same class of structures: sentences containing a non-subject (interfering) noun in addition to a subject noun and a verb. Both accounts make the same predictions for processing ungrammatical sentences with a number-mismatching interfering noun, and this prediction finds consistent support in the data. However, the similarity-based interference accounts predict similar effects not only for morphosyntactic, but also for the semantic level of language organization. We verified this prediction in three single-trial online experiments, where we found consistent support for the predictions of the similarity-based interference account. In addition, we report computational simulations further supporting the similarity-based interference accounts. The combined evidence suggests that the faulty encoding accounts are not required to explain comprehension of ill-formed sentences.

For the processing of grammatical sentences, the accounts make conflicting predictions, and neither the slowdown predicted by the similarity-based interference account, nor the complementary slowdown predicted by the faulty encoding accounts were systematically observed. The majority of studies found no difference between the compared configurations. We tested one possible explanation for the lack of predicted difference, namely, that both slowdowns are present simultaneously and thus conceal each other. We decreased the amount of similarity-based interference: if the effects were concealing each other, decreasing one of them should allow the other to surface. Surprisingly, throughout three larger-sample single-trial online

experiments, we consistently found the slowdown predicted by the faulty encoding accounts, but no effects consistent with the presence of inhibitory interference.

The overall pattern of the results observed across all the experiments reported in this dissertation is consistent with previous findings: predictions of the interference accounts for the processing of ungrammatical sentences receive consistent support, but the predictions for the processing of grammatical sentences are not always met. Recent proposals by Nicenboim et al. (2016) and Mertzen et al. (2020) suggest that interference might arise only in people with high working memory capacity or under deep processing mode. Following these proposals, we tested whether interference effects might depend on the depth of processing: we manipulated the complexity of the training materials preceding the grammatical experimental sentences while making no changes to the experimental materials themselves. We found that the slowdown predicted by the faulty encoding accounts disappears in the deep processing mode, but the effects consistent with the predictions of the similarity-based interference account do not arise.

Independently of whether similarity-based interference arises under deep processing mode or not, our results suggest that the faulty encoding accounts cannot be dismissed since they make unique predictions with regard to processing grammatical sentences, which are supported by data. At the same time, the support is not unequivocal: the slowdowns are present only in the superficial processing mode, which is not predicted by the faulty encoding accounts. Our results might therefore favor a much simpler system that superficially tracks number features and is distracted by every plural feature.

Acknowledgements

First and foremost, I want to thank Shravan Vasishth for being the model of an inquisitive researcher who never stops to push the boundaries of his own knowledge no matter what. I can only wish to be as fierce in broadening my horizons throughout life as you are in broadening yours, but I will certainly strive to meet this high standard. Thank you for the freedom you granted me to pursue my own scientific interests, for valuable feedback, and for constantly providing new intellectual challenges.

I also want to express my sincere gratitude to Titus von der Malsburg for his support and mentoring, but most of all, for bringing even more fun into science. Thank you for all the numerous improvements you initiated, and for your kind patience.

I would like to thank all the present and past members of the Vasishth lab for creating a very stimulating and friendly environment, for sharing their ideas and for their valuable feedback. I am especially grateful to Lena Jäger for her mentorship, to Kate M. Stone and Garrett Smith for their continuous help with creating experimental materials and for feedback on the previous versions of some chapters; and to Dorothea Pregla for her thoughtful criticism of the modeling reported in Chapter 3.

I would also like to express my gratitude to all the wonderful people who I worked with on projects that might have delayed the completion of this dissertation, but were too exciting to pass by: Daniela Mertzen for leading an amazing research project that I am proud to be a part of; Reinhold Kliegl for introducing me to reading research and for constantly demonstrating that a rigorous scientist can be the most kind and friendly person; Olga Dragoy for believing in me and trusting me, and the whole team of the Center for Language and Brain at HSE for being there for me; Irina Sekerina for her support, mentoring, help and advice; Olga Parshina and Svetlana

Dorofeeva for inviting me to collaborate on very interesting projects studying reading skills, and Anastasiya Lopukhina for being not only a supportive colleague, but more importantly, a great friend.

I am grateful to DAAD, which granted me a three-year scholarship for working on this dissertation and covered my travel costs for several conferences. I would also like to express my gratitude to Strukturbereich Kognitionswissenschaften of Potsdam University that financed many of the experiments reported in this dissertation.

And finally, I want to thank Aleksei for agreeing to move to Potsdam with me and for making sure I have the time to do all the work that this dissertation required.

Contents

Chapter 1

Introduction

While theoretical linguistics focuses on the properties of human language that allow us to exchange messages, it abstracts away from the fact that messages are produced and perceived by humans, who are known to be error-prone. This becomes the territory of psycholinguistics. One of the important questions addressed by psycholinguistics is why misinterpretations arise. On the one hand, we know from experience that in the overwhelming majority of cases people understand language successfully, otherwise communication would not be possible. On the other hand, we also know that comprehension suffers from occasional errors: not every syntactic structure is assembled faithfully, not every message is perceived in the way intended by the speaker.

Importantly, while some types of errors are observed relatively often, other potentially possible errors almost never arise, and when they do, it might be attributed to an attentional glimpse. In contrast, many types of errors have long been noticed to occur systematically, such as, for example, misinterpretation of the so-called garden-path sentences ("The complex houses married and single soldiers and their families") or erroneous agreement between the subject and the verb in number ("The computer installed in the Russian antiballistic missiles are…", adapted from Bock and Miller, 1991).

The systematic nature of these errors led researchers to believe that they arise due to some glitch within the mechanism responsible for language comprehension — parser — and point at the weak spots within this mechanism. In other words, systematic comprehension errors are a natural consequence of how the parser works, just as optical illusions are a natural consequence of how the visual system works. Consequently, researchers see these linguistic illusions as an opportunity to get a glimpse into the mechanics of human parser system, and try to infer the inner workings of the parser based on the kinds of mistakes it occasionally makes, just as some properties of visual perception were decoded thanks to the optical illusions.

The same logic applies to the situations in which people do not necessarily make mistakes, but experience measurable difficulties, such as processing of garden-path sentences (Clifton & Ferreira, 1989; Frazier, 1987; Rayner et al., 1983) or double center embeddings (Gibson, 2000; Miller & Chomsky, 1963; Noam et al., 1957). If processing of a certain structure is difficult, then it must make the parser approach some kind of limit that it abides by. One of the most frequently postulated constraints on the parser is the limitation in memory available for processing, which we will discuss in more detail later.

Consequently, the general accounts of language processing that aim for cognitive plausibility need to not only model production or comprehension of well-formed sentences (and be constrained enough to not produce ill-formed structures), they also need to capture the processing difficulties that people experience and make the errors that people systematically make. For example, the first paper presenting what is currently the most well-known general-purpose model of language comprehension — the Lewis and Vasishth model — has demonstrated that the model not only successfully processes well-formed sentences, but also predicts processing difficulties exactly in those syntactic configurations that people struggle to parse, such as garden-path sentences and sentences with double center embeddings (Lewis & Vasishth, 2005). Numerous other models and processing principles had been proposed in the last 50 years to account for the particular systematic errors that comprehenders make or for the difficulties they experience (Minimal Attachment by Frazier and Rayner, 1982; Late Closure by Frazier, 1978; the sausage machine by Frazier and Fodor, 1978; the feature percolation account by Vigliocco et al., 1995; the marking and morphing model by Eberhard et al., 2005; the good-enough processing account by Ferreira and Patson, 2007; the local coherence account by Tabor et al., 2004; etc.). This proliferation of models reflects the rapid development of scientific thought, but at the same time, creates a certain overabundance problem.

In an ideal world, one parsimonious model of sentence processing that covers a range of known effects would be more preferable than a set of narrowly-focused accounts, each specializing in one specific effect, but making no (or false) predictions outside its domain of application. Currently, there are several large-scale models of sentence processing each of which accounts for a range of well-known effects: the Lewis and Vasishth model (Engelmann et al., 2019; Lewis & Vasishth, 2005; Nicenboim & Vasishth, 2018), the self-organized sentence processing model (Smith et al., 2018; Tabor & Hutchins, 2004), the expectation-based surprisal account (Hale, 2001; Levy, 2008a), the noisy channel model (Levy, 2008b, 2011), and its recent extension as a lossy-context surprisal (Futrell et al., 2020; Futrell & Levy, 2017). All these models do not limit their domain of application and strive to account for every existing type of linguistic structure as well as for cross-linguistic variation. At the same time, there is a greater number of smaller-scale theoretical accounts, each aiming to cover one known type of comprehension difficulties or errors. The predictions of general-purpose language processing models naturally overlap with at least some predictions of the smaller-scale accounts. This raises the question whether at least some of the smaller-scale accounts might be redundant — cannot all known effects be covered by the general-purpose models? Or do we need the smaller-scale accounts because some predictions they make are unique?

The main goal of this dissertation is to empirically evaluate the predictions of the general-purpose models assuming similarity-based interference (Lewis & Vasishth, 2005; McElree, 2000) and the predictions of the group of smaller-scale accounts that we will refer to as faulty encoding accounts (Bock & Eberhard, 1993b; Eberhard et al., 2005). Both the similarity-based interference and the faulty encoding accounts make predictions with regard to processing the same general class of structures: sentences containing some non-subject noun in addition to a subject noun and a verb (all other constituents being optional), see (1) and (2).

(1) **Ill-formed sentences:**

a. The drawer with the knife apparently cuts ...

b. The drawer with the handle apparently cuts ...

c. The drawer with the knife apparently open ...

d. The drawer with the knives apparently open …

(2) **Well-formed sentences:**

a. The admirer of the singer apparently thinks

b. The admirer of the singers apparently thinks

c. The admirer of the play apparently thinks

d. The admirer of the plays apparently thinks

… the show was a big success.

In both examples (1) and (2), to complete a subject-verb dependency, the parser needs to establish a relationship between the verb and a non-adjacent subject noun that was encountered earlier during parsing. In every sentence, there is more than one noun, although only one is a syntactically licensed subject. In some versions of the sentences, these additional nouns (we will refer to them as *interfering nouns*) might seem to be a good match for the verb because they share some features with the verb. In (2-c) it is the grammatical number, in (1-a) it is the semantic fit, and in (2-a) it is both. In other versions of the sentences, the interfering nouns do not match the verb that well.

For some sentence configurations, the predictions of the similarity-based interference and the faulty encoding accounts coincide, for others, they contradict each other, for others still, the models make complementary predictions that can be true at the same time. In the next section, we will review the processing mechanisms assumed by the two groups of accounts and the predictions they make with regard to particular sentence configurations.

1.1 Similarity-based interference accounts

Similarity-based interference is an umbrella term for the mechanisms postulated by different memory accounts to lead to forgetting (M. C. Anderson & Neely, 1996; Nairne, 2002b). In particular, forgetting is assumed to result from our inability to retrieve a particular item from memory when it is similar to other items held in memory, and not from the decay of the target item activation, as previously believed (Brown, 1958). The underlying processes leading to poorer recall can be different: corruption of the item representation during encoding or maintenance in memory when similar items are already memorized (encoding interference, proposed in Lewandowsky et al., 2008; Oberauer and Kliegl, 2006), or errors that arise during retrieval from memory (retrieval interference, proposed in J. R. Anderson and Lebiere, 1998; M. C. Anderson and Neely, 1996; Oberauer and Kliegl, 2006).

Why is the concept of similarity-based interference relevant to language processing? We assume that the processing of ongoing linguistic input must rely on memory in order to build syntactic and semantic relationships between words in real time. Speech proceeds one word at a time, and once a word is uttered, there is no opportunity to hear it again. Yet people are surprisingly efficient at extracting meaning under these circumstances[1]. Based on that observation, researchers conclude that human parser must rely on memory in order to store and extract the constituents when needed, which, in turn, allows the parser to build relationships between words and constituents that had been processed some time ago. The second assumption linking language processing and memory is that parsing must be subserved by the same memory mechanisms that are used in other tasks. That is, memory resources required for parsing are not domain-specific, and language processing must adhere to the general restrictions imposed by human memory, and must be fallible in those cases where human memory generally is. In particular, the general mechanisms that affect recall, such as similarity-based interference, must affect recall required during parsing as well. This assumption is shared, in particular, by the models of sentence comprehension introduced by Lewis and Vasishth (2005) and McElree (2000).

The Lewis and Vasishth model relies on the cognitive architecture ACT-R (J. R. Anderson, 1996, 2014). The central assumption of the model is that only one item is immediately available for processing: words and more complex constituents are activated when they are encountered, but then their activation decays. Consequently, when a dependency between two constituents must be built, only the one that is currently being processed is available, the other must be reactivated and retrieved from content-addressable memory in order to complete the dependency. Retrieval from memory is central to parsing: it determines the structure that will be built;

[1] Although during reading, we can get back to the parts of text or sentence that we had already read, people do not generally do it; constant rereading is rare and signals general comprehension difficulties (Vasishth et al., 2013).

in addition, while other operations involved in processing have constant latencies, retrieval latencies can vary, and it is the retrieval latency that ultimately determines the differences in processing times between structures. We will now review how retrieval works in more detail.

In the Lewis and Vasishth model, each word and constituent is represented as a bundle of features. When an already processed word is required in order to build a syntactic dependency, the parser initializes retrieval from memory by identifying a set of features that the word must have, and sets retrieval cues for these features (such as `+MASCULINE, PLURAL`). Note that retrieval cues do not necessarily target all the features of the to-be retrieved word. If the retrieval site is a verb unmarked for number, then number cannot be specified as a retrieval cue since no information is available, although the word that needs to be retrieved may have a value of number feature.

When retrieval cues are set, each cue propagates activation among all items that have a feature matching the cue. It is this spreading activation that largely determines the outcome of the retrieval operation. To be retrieved, a constituent must have the highest activation among other items in memory, and this activation should also exceed the retrieval activation threshold. The activation of each item depends on its base-level activation (corresponding to the recency of its use), spreading activation, and random noise. In an ideal situation, only one item — the target of retrieval — matches all the retrieval cues and gets all the spreading activation. It will have the highest activation among other items in memory and will be retrieved. Importantly, the activation of an item that is selected for retrieval determines also the speed of its retrieval: the higher the activation, the greater is the retrieval speed.

However, the target of retrieval is often not the only item that has features matching retrieval cues. If that is the case, spreading activation is divided equally among all the constituents that match retrieval cues. For example, if several nouns have been processed by the time when search for a particular noun is initiated, the activation that the `+NOUN` cue spreads is divided equally among all the nouns. This situation is referred to as *cue overload*. According to the Lewis and Vasishth model, it is precisely the cue overload that is responsible for similarity-based interference. When some of the spreading activation goes to the non-target items, the target item gets less activation than it could have received, and is retrieved more slowly than if it were the only item matching all the retrieval cues. This slowdown is referred to as *inhibitory interference*. In extreme cases, the target item may end up having lower activation that some other item in memory (due to random fluctuations in activation, for example), and not get selected for retrieval.

Another account assuming similarity-based interference in language processing is the direct access model proposed by McElree (Martin & McElree, 2011; McElree, 2000; McElree et al., 2003). In contrast to the Lewis and Vasishth model, it was not formally specified and makes no quantifiable predictions with regard to processing times. The central assumption of the model is that the latency of each retrieval is constant (modulo noise) and does not depend on the activation of the target item in memory. But due to cue overload, retrieval may fail. In that case, additional retrieval attempt will be initiated, which will affect processing times. Although the proposed mechanism was recently formalized and pit against the predictions of the Lewis and Vasishth model by Nicenboim and Vasishth (2018), we will not evaluate the detailed predictions of the direct access model in the following sections because its predictions for processing ill-formed sentences are still unclear.

To summarize, similarity-based interference accounts assume that language processing is subserved by domain-general memory and must adhere to the limitations imposed on human memory. In particular, language processing must be fallible to similarity-based interference: when several words that are held in memory share some linguistic property, forming a dependency between the currently processed word and one of the words held in memory should be more difficult and fail more often than if the targeted word had a unique feature markup. Both the Lewis and Vasishth model and the direct access model by McElree specifically assume that similarity-based interference arises at the stage of retrieval from memory. Surprisingly, the experimental evidence supporting this assumption is rather scarce: the seminal paper by Van Dyke and McElree (2006) reported effects compatible with retrieval, but not encoding interference; however, a recent replication by the same authors (Van Dyke et al., 2014) as well as a large-scale replication attempt (Mertzen et al., 2020) failed to find retrieval interference effects across three languages.

Chapter 2 of this dissertation (published as Laurinavichyute et al., 2017) aims to answer two questions: whether similarity-based interference influences the processing of reflexive-antecedent dependency, and if it does, whether interference arises specifically during retrieval from memory. We report three experiments targeting similarity-based interference in processing grammatical sentences with gender-marked personal pronouns compared to gender-unmarked reflexives (Experiment 1, German), and gender-unmarked reflexives compared to gender-marked reflexives (Experiments 2A and 2B, Russian). Across three experiments, we found no main inhibitory effect of gender match between the antecedent and the interfering noun. However, in Experiments 2A and 2B participants had longer reading times at the gender-unmarked, but not gender-marked reflexive in conditions where the gender of the antecedent and the interfering noun coincided (only participants with high comprehension question response accuracy in case of Experiment 2A). These results are incompatible with retrieval interference: if interference arises during retrieval, we should not observe any inhibitory interference effects in processing gender-unmarked reflexives since gender is not available as a retrieval cue. Our results are instead compatible with encoding interference, which might, however, observed only in a subset of readers, those who answered most comprehension questions correctly. Whether high accuracy of our participants results from higher working memory capacity, greater attention to the task, or higher motivation to perform the task, it seems to be correlated with similarity-based interference. For a more extended discussion of possible processing strategies that might link interference to any of the properties mentioned above, the reader can refer to Mertzen et al., 2020; Nicenboim et al., 2016; Swets et al., 2008; Von der Malsburg and Vasishth, 2013.

To summarize, across two experiments (and only in a subset of participants in Experiment 2A) we found slowdowns in reading times consistent with the predicted inhibitory interference effect, but crucially, interference could have arisen only during encoding to memory, not retrieval from memory. And although the Lewis and Vasishth model specifically assumes that similarity-based interference arises during retrieval, in the remainder of this chapter we will refer to similarity-based interference in general rather than retrieval interference even when discussing the predictions of the Lewis and Vasishth model.

1.2 Faulty encoding accounts

While the distinctive feature of the similarity-based interference accounts is that they draw inspiration from memory research, and, in case of the Lewis and Vasishth model, aim to explain the processing of all kinds of syntactic (and discourse, see Brasoveanu and Dotlačil, 2019) structures, the group of faulty encoding accounts differs along both dimensions: these accounts rely on syntactic theory and aim to cover a limited set of syntactic configurations. They assume that difficulties and errors in processing stem not from limitations imposed by memory, but rather from the normal morphosyntactic processing gone astray.

The faulty encoding accounts were originally designed to explain agreement attraction errors in production. Agreement attraction is the term that originally referred to the relatively frequently occurring type of error in spontaneous speech, such as "We speculate that the difference between the studies stem from...". The crucial factor contributing to the emergence of attraction errors is the presence of an interfering noun, the co-called attractor (*studies* in the example above), that takes over morphosyntactic control of the verb. A parallel effect was found in comprehension: in ungrammatical sentences where the attractor has the same number marking as the verb, reading times on the verb are faster than in control sentences where the number marking of the attractor does not match that on the verb. In addition, ungrammatical sentences with attraction errors are more often judged as grammatical or acceptable than ungrammatical sentences that do not contain an interfering noun matching the verb in number (Hammerly et al., 2019; Patson & Husband, 2016; Vasishth et al., 2017; Wagers et al., 2009). As a result of finding these effects in comprehension, the predictions of the faulty encoding accounts were extrapolated to comprehension under the assumption that production system is being actively used to facilitate comprehension (Christiansen & Chater, 2016; Meyer et al., 2016; Pickering & Garrod, 2013). We will now review the particular mechanisms proposed in the faulty encoding accounts in more detail.

The *feature percolation* account (Franck et al., 2002a; Nicol et al., 1997; Vigliocco et al., 1995; Vigliocco & Nicol, 1998) relies on the concept of feature markedness: singular is considered to be an underspecified (unmarked) member of the binary number opposition (see, for example, Bock & Eberhard, 1993b; Harley & Ritter, 2002). According to the feature percolation account, if the subject noun is singular and therefore, unmarked for number, the plural feature that belongs to an interfering noun located within the subject noun phrase might sometimes erroneously percolate up the syntactic tree, transfer its marking to the subject noun phrase, and thus affect the subject-verb agreement computation. In that case, feature checking at the verb marked for plural returns no error signal.

The feature percolation account captures an important constraint on agreement attraction, the singular-plural asymmetry: attraction occurs when the subject noun is marked for singular, and the interfering noun for plural, but not the other way around (Pearlmutter et al., 1999a). Another strength of the account is that it utilizes an independently proposed percolation mechanism (Cole et al., 1993; Cowper, 1987) that requires only a minor change to account for the prominent class of mistakes. At the same time, the domain of feature percolation is limited to the configurations where the interfering noun belongs to the subject noun phrase, while attraction errors were also found in object relative clauses (Hartsuiker et al., 2001), questions (Vigliocco & Nicol, 1998), and direct object constructions (Schäfer et al., 2019). In all these configurations the interfering noun is located outside the subject noun phrase and its plural feature cannot percolate to the NP root node of the subject. In addition, feature percolation account cannot explain how the semantic properties of the noun phrase can influence attraction, for example, why attraction rates rise when the prepositional phrase has a distributive interpretation (Foote & Bock, 2012; Hartsuiker et al., 1999; Vigliocco et al., 1996), or when the semantic representation of a collective head noun, such as *team*, is more multitude-like (Humphreys & Bock, 2005; Smith et al., 2018), or when the subject and the attractor nouns are more closely linked in the mental model of the referred entity (e.g. *the painting of/with the flower*, Solomon and Pearlmutter, 2004).

The second faulty encoding account was introduced in order to explain how semantic properties of individual nouns or the whole noun phrase can influence agreement processing. The *marking and morphing* model (Bock et al., 2001; Eberhard et al., 2005) postulates that computation of subject number depends both on the conceptual number representation of the entity that is referred to, and on the formal number marking present in the syntactic structure. The conceptual number representation is called *notional number* — a semantic representation of the entity that is being referred to, either as a multitude or as a singular unit. Both nouns, such as *team*, and noun phrases, such as *the picture on the postcards*, can be notionally plural while being syntactically singular. According to the marking and morphing account, the subject's notional number influences the computation of number agreement over and above the number mismatch between the interfering noun and the subject. Essentially, the more multitude-like the abstract representation of the subject, the higher is the probability of using a plural verb. The assumption that notional plurality influences agreement computation received excellent support from empirical investigations.

The formal, morphosyntactic, part of the number assignment depends on the weighted sum of plural morphemes on words comprising the subject noun phrase. Consequently, a plural feature on a non-subject noun within the subject noun phrase can disrupt number computation for a singular subject. The assumption here, again, is that the singular number is the unmarked default value of the number opposition. If a plural feature on an interfering number does affect subject noun computation, and the subject receives a number value ambiguous between singular and plural, then in some proportion of cases, subject noun phrase will be encoded as plural, and, as in the case with feature percolation, feature checking at the verb marked for plural will be successful. If we relax one of the model assumptions — that only plural morphemes within the subject noun phrase affect number assignment — then the model can also cover attraction effects caused by interfering nouns located outside of the noun phrase (but this possibility is currently not instantiated, see Eberhard, Cutting, and Bock 2005, p. 544).

To summarize, the faulty encoding accounts were originally proposed to explain one particular kind of mistakes that arise both in language production and comprehension, the faulty number agreement with a non-subject noun. While different faulty encoding accounts propose distinct mechanisms underlying the observed agreement attraction effects, they still share a core property: they assume that the subject number is encoded incorrectly, either as unambiguously plural (the feature percolation account), or as somewhat plural on the plurality continuum (the marking and morphing account). In contrast to what is proposed by the similarity-based interference accounts, the faulty encoding accounts assume that the subject noun is identified correctly, only its feature markup is misspecified, and agreement computation itself proceeds correctly.

We will now turn to the particular predictions the similarity-based interference and the faulty encoding accounts make for the processing of grammatical and ungrammatical sentences with interfering nouns. We will start with ungrammatical sentences.

1.3 Processing sentences with interfering nouns

1.3.1 Ungrammatical sentences

To briefly remind the reader, we are interested in the processing of the class of ungrammatical sentences where the subject noun mismatches the verb in number while the interfering noun (or the attractor noun, i.e. some non-subject noun present in the sentence) can match or mismatch the verb in number, as in Example (3):

(3) a. The drawer with the knife apparently are ...
    b. The drawer with the knives apparently are ...

While processing of ungrammatical sentences universally leads to disruption and therefore, to slowdowns in processing times, the slowdown is greatly diminished if the interfering noun matches the number marking on the verb, as in (3-b) as compared to (3-a) (inter alia, Dillon et al., 2013; Jäger et al., 2020; Lago et al., 2015; Pearlmutter et al., 1999a; Tucker et al., 2015; Wagers et al., 2009).

The Lewis and Vasishth model straightforwardly accounts for the slowdown that arises in ungrammatical sentences without the number-matching interfering noun: in an ungrammatical sentence, the subject matches only one retrieval cue out of two — the structural `+C-COMMAND`, but not the `+PLURAL` cue. Consequently, the subject gets less spreading activation, and will be retrieved slower than if the sentence was well-formed and it received all available spreading activation. The Lewis and Vasishth model also accounts for faster processing of ungrammatical sentences with a number-matching interfering noun ((3-b) as compared to (3-a)). When an interfering noun matches the non-structural retrieval cue, both the subject noun and the interfering noun each get half of the spreading activation. The resulting activations of both nouns would also be very close. Recall that retrieval speed depends on the activation of the to-be-retrieved item. If only the subject noun receives spreading activation, retrieval (and therefore, processing) times depend solely on the activation of the subject, however high or low it might be. If both nouns receive the same amount of spreading activation and have very similar resulting activation levels, the processing times on each trial will be defined by the noun with the highest activation of the two. It means that over the course of several trials, average processing times will be faster when there are two nouns with similar levels of activation than if there is only one noun with the same average activation. The predicted speedup in the processing of ungrammatical sentences with the interfering noun matching the number retrieval cue is referred to as *facilitatory interference*. Processing speedups consistent with the predicted facilitatory interference in ungrammatical sentences have been consistently observed.

The faulty encoding accounts also predict a speedup in processing ungrammatical sentences with an interfering noun matching the number marking on the verb. According to the feature percolation account, a plural feature of the interfering noun might occasionally percolate up the syntactic tree and mark the whole subject noun phrase as plural. In that case, encountering the verb marked for plural will return no error signal. In turn, the marking and morphing account predicts that the weighted sum of plural morphemes within the subject noun phrase will lead to perceiving the subject as somewhat plural on the plurality continuum. If that happens, the subject will be in some cases encoded as plural, and encountering the verb marked for plural will again return no error signal, the sentence would seem well-formed.

To sum up, both the similarity-based interference and the faulty encoding accounts predict the same outcome for processing ungrammatical sentences with a number-mismatching interfering noun, although for different reasons, and this outcome is very consistently observed: faster reading times and more incorrect responses are reported when the morphosyntactic marking of the verb is unlicensed by the subject, but matches the marking on the interfering noun. Coinciding predictions do not allow us to differentiate between the accounts. However, the similarity-based interference accounts have broader domain and apply the same processing principles to any features used during parsing. In particular, similarity-based interference accounts predict effects similar to number attraction on the semantic level of language organization: facilitatory effects in processing the verb in ill-formed sentences where the interfering noun matches the thematic restrictions set by the verb, which the subject noun does not match, as in "The drawer with the knife apparently cuts ...". The crucial detail here is that while being semantically implausible, the sentence is grammatically well-formed and contains no agreement attraction errors. As the faulty encoding accounts explain attraction errors through faulty mechanisms of morphosyntactic number assignment and agreement, they simply cannot predict parallel effects on a non-morphosyntactic plane.

Precisely this semantic facilitatory interference effect (which we will also refer to as semantic attraction) has been demonstrated in eye movements recorded while reading (Cunnings & Sturt, 2018). However, it has not yet been compared to the facilitatory interference in number, and the similarity-based accounts predict these effects to be of the same magnitude. To replicate the semantic facilitatory interference effects in ungrammatical sentences and compare them to the well-known morphosyntactic facilitatory interference effects, we conducted three experiments described in detail in Chapter 3. Our experiments also pit the predictions of the similarity-based interference accounts and of the faulty encoding accounts against each other: while both predict agreement attraction in number in ill-formed sentences, only the similarity-based interference accounts predict semantic attraction.

Across three larger-sample single-trial online experiments, we consistently found both morphosyntactic and semantic attraction (facilitatory interference) effects, without any difference in effect sizes between morphosyntactic and semantic attraction effects. This outcome is in line with the predictions of similarity-based interference accounts and cannot be reconciled with the predictions of the faulty encoding accounts. In general, our results suggest that in processing ungrammatical sentences, people are more likely to judge the sentences as acceptable if the interpretation can be salvaged using the features of the interfering noun. In addition, we report computational simulations of both the morphosyntactic and semantic attraction effects using the interACT implementation of the Lewis and Vasishth model (Engelmann et al., 2019).

While the model successfully captures all the effects present in the acceptability judgments, we show that it cannot capture the observed reaction times due to the principled restrictions imposed by model specification.

To conclude, both the similarity-based interference accounts and the faulty encoding accounts predict morphosyntactic attraction effects in ungrammatical sentences, and the effects are consistently observed across a wide range of studies. But only the similarity-based interference accounts predict semantic attraction effects in ungrammatical sentences, the effects that were observed in the two experiments reported by Cunnings and Sturt (2018), and three experiments reported in Chapter 3. The combined evidence suggests that the faulty encoding accounts might not be required to explain comprehension of ill-formed sentences: they do not predict the observed semantic attraction effect and are not unique in predicting the morphosyntactic attraction. At the same time, the faulty encoding accounts still make unique predictions with regard to processing well-formed sentences. After all, parsing ill-formed sentences is an unconventional task that is considered by some researchers as being not worth modeling: people do not encounter ungrammatical or ill-formed sentences regularly, the main task of the human parser is to make sense of well-formed input.

1.3.2 Grammatical sentences

Again, we would like to briefly remind the reader that here, we focus on the processing of grammatical sentences that contain an interfering noun (non-subject noun present in the sentence) matching or mismatching the number marking on the subject noun and on the verb, as in Example (4):

(4) a. The admirer of the singer apparently is ...
    b. The admirer of the singers apparently is ...

Similarity-based interference accounts predict a slowdown in grammatical sentences with a number-matching interfering noun, such as (4-a), arising due to cue overload. The mechanism is as follows: the processing of verb *is* requires retrieval of the subject noun to build the subject-verb dependency. At the verb, retrieval cues `+C-COMMAND`, `+SINGULAR` can be set. When both the subject noun, *the admirer*, and the interfering noun, *the singer*, match the number cue, the spreading activation from the cue is divided equally among both nouns. The subject noun now receives less spreading activation than it would have received if it was the only chunk in memory matching the number retrieval cue (as in (4-b)). Consequently, when the interfering noun overtakes some part of spreading activation, the spreading activation and, as a result, the total activation of the subject noun will be lower, and the subject noun will be retrieved slower than in (4-b). Under some circumstances, the interfering noun could occasionally even be misretrieved instead of the subject noun. Although interference effects were widely tested in reflexive-antecedent dependencies as well as in subject-verb dependencies with the focus on thematic fit, only two studies that we know of explored number interference in grammatical sentences. The predicted inhibitory interference effect was found by Franck et al. (2015), but a large-scale study with 180 participants by Nicenboim et al. (2018) turned out inconclusive, although the direction of the observed effect was in line with the predicted slowdown.

The predictions of the faulty encoding accounts for the same set of conditions differ: a slowdown is expected in (4-b), where a number-mismatching interfering noun is present. In line with the general principles of number encoding proposed in these accounts, when a plural interfering noun is a part of the subject noun phrase, the parser would occasionally encode the number of the whole subject noun phrase as plural (due to either feature percolation or the plural morpheme on the interfering noun affecting number computation). When a singular verb is then encountered, the number marking on the verb would not correspond to the encoded plural number on the subject constituent, and the so-called *illusion of ungrammaticality* would arise. This mismatch should lead to longer average reading times on the verb in grammatical sentences with a plural interfering noun (4-b) as compared to grammatical sentences with a singular interfering noun (4-a). The illusion of ungrammaticality is rarely observed, and many of those experiments where it is observed raise internal validity concerns, in particular, that the slowdown originates not at the verb, but at the preceding plural noun (Franck et al., 2015; Lago et al., 2015; Patson & Husband, 2016; Pearlmutter et al., 1999b; Wagers et al., 2009). The lack of support for the slowdown in (4-b) predicted by the faulty encoding accounts has been perceived as evidence against applying this group of accounts to comprehension. This position has recently been challenged by Hammerly et al. (2019) who demonstrate that the predicted illusion of ungrammaticality is present in the comprehension of grammatical sentences, but concealed by a bias towards “grammatical” response in the judgment task. When the response bias is neutralized, predicted illusion in grammatical sentences is observed. While this finding is illuminating, response bias alone cannot explain the lack of predicted slowdown in reading times in setups where no grammaticality judgment response is required.

To summarize, with very few exceptions, in the processing of grammatical sentences, neither the slowdown predicted by the similarity-based interference account, nor the complementary slowdown predicted by the faulty encoding accounts were found. The majority of studies found no difference between the compared configurations (Cunnings & Sturt, 2018; Lago et al., 2015; Paspali & Marinis, 2020; Patson & Husband, 2016; Thornton & MacDonald, 2003; Tucker et al., 2015; Wagers et al., 2009). As the existence of attraction effects in the processing of grammatical sentences is far from established, many researchers that look for a parsimonious explanation of attraction effects now believe that the faulty encoding accounts do not adequately capture comprehension. They conclude that similarity-based interference is the only mechanism needed to cover the attraction effects observed in comprehension, which are in this case reduced to attraction in ungrammatical sentences (Hammerly et al., 2019; Tanner et al., 2014; Wagers et al., 2009). However, this reasoning overlooks that the predictions of the similarity-based interference accounts for processing grammatical sentences are compromised to the same degree as the predictions of the faulty encoding accounts. Both groups of accounts predict the same consistently observed facilitatory effect in processing ungrammatical sentences, and different inhibitory effects, neither of which is consistently observed, in the processing of grammatical sentences. If anything, available evidence speaks against both groups of accounts to an equal degree.

Chapter 4 tests a potential explanation for the lack of both inhibitory effects in grammatical sentences predicted by the similarity-based interference and the faulty encoding accounts. One possible reason that neither the slowdown in (4-a) predicted by the similarity-based interference accounts, nor the slowdown in (4-b) predicted by the faulty encoding accounts is observed is that both slowdowns are present simultaneously and therefore cancel each other out. We address this issue by decreasing the amount of similarity-based interference: if the effects were indeed canceling each other out, decreasing one of them should allow the other to surface. Surprisingly, throughout three larger-sample single-trial online experiments, we consistently found the slowdown predicted by the faulty encoding accounts (the illusion of ungrammaticality), and no interaction that would suggest that the illusion of ungrammaticality is normally canceled by the inhibitory interference. We then discuss and test one potential explanation for observing a replicable illusion of ungrammaticality in reading times, which contradicts the outcomes of numerous previous experiments. Based on the results of the experiments, we suggest that the faulty encoding accounts cannot be dismissed since they make a unique prediction with regard to processing grammatical sentences, a prediction that the general-purpose similarity-based interference accounts do not share.

To condense the outcomes even more, we found that to explain how ill-formed sentences are processed, the predictions of the Lewis and Vasishth model, but not those of the faulty encoding accounts, are necessary and sufficient. But for the processing of well-formed sentences, neither account is sufficient to explain all the patterns present in the data. Predictions of the similarity-based interference accounts were partly supported in experiments reported in Chapter 2, but not supported in the first three experiments reported in Chapter 4. Instead, in Chapter 4, we observed the reverse effects consistent with the broad predictions of the faulty encoding accounts. Finally, we demonstrate that the illusion of ungrammaticality arises only in the superficial processing mode; in the deep processing mode, a (delayed) slowdown consistent with inhibitory interference is observed. But our results still pose a challenge to the similarity-based interference accounts: we observed no semantic interference in reading of well-formed sentences, even when deep processing was encouraged.

Chapter 2

Retrieval and encoding interference: cross-linguistic evidence from anaphor processing

In human language processing, working memory is crucial for linking together parts of syntactic dependencies. Therefore, to understand language processing it is important to understand mechanisms and limitations of the working memory system, especially those that lead to forgetting. Although previously attributed to decay (Brown, 1958), now forgetting is often believed to stem from similarity-based interference from other entities stored in memory (Lewandowsky et al., 2008; Nairne, 2002a; Oberauer & Kliegl, 2006). Similarity-based interference may affect different working memory processes: writing (encoding) to memory, maintenance in memory, and retrieval.

2.0.1 Potential sources of similarity-based interference

Interference may arise during writing of an item to the working memory (encoding) if it shares some features with other items in memory. Such a model can be instantiated in different ways. One was proposed by Oberauer and Kliegl (2006): in their model, items in working memory are represented by sets of features that are activated together. If two items share the same feature (for example, two nouns share the same gender), they compete for it, and the competition may lead to so-called *feature overwriting* – loss of the feature in one of the sets. As a result, representation of an item that lost a feature gets less distinguishable, and the probability of the item's successful retrieval decreases. An alternative realization of encoding interference was proposed by Lewandowsky et al.: when an item is first presented, its novelty is assessed in comparison to other items already stored in memory and their feature sets. If the item is judged to be novel, it is assigned greater encoding weight than if it is judged to be similar to the items in memory. The greater the encoding weight of an item, the easier it is to retrieve. Note that although in both models interference arises during encoding of item's representation to the working memory, presence of interference affects retrieval of the item from memory.

Interference may also arise during the maintenance of an item in memory: if two or more items that share a certain feature are being stored in working memory, they may become less distinguishable from one another. The feature overwriting mechanism cited above can be thought of as maintenance interference depending on the time when the overwriting occurs. Consequently, maintenance interference is difficult to separate from encoding interference in practice, since we can only observe their effects at retrieval. Hence, in the following sections we do not distinguish between encoding and maintenance interference.

The third type of interference — retrieval interference — is assumed to arise during retrieval of an item from memory if other items share features relevant for retrieval with the target item. Among others, this type of interference is assumed in two memory retrieval models that have been applied to sentence processing: the Adaptive Control of Thought-Rational (ACT-R, see J. R. Anderson, 2014; Lewis and Vasishth, 2005; Lewis et al., 2006) and the working memory model by McElree (Martin & McElree, 2011; McElree, 2000; McElree et al., 2003). In the ACT-R model, each item is represented in memory as a bundle of features. To be retrieved, it must receive the highest activation among other items in memory. The activation of each item consists of its base-level activation (corresponding to the frequency and recency of its use), random noise and spreading activation. Spreading activation is what an item receives during retrieval: to find a specific item in memory, each retrieval cue (such as a particular gender or case) propagates activation among all items which have a feature that matches the cue. The activation that each cue spreads is divided between all items that match this cue. According to ACT-R, this mechanism is the cause of similarity-based retrieval interference. The item whose features match all the retrieval cues receives the most spreading activation, which normally results in the highest boost of activation (modulo base-level activation and noise) and therefore reaches the activation threshold first (i.e., is retrieved from memory). Importantly, the activation of an item determines the speed of its retrieval: once an item reaches a certain activation threshold, it is retrieved, i.e., the stronger the boost in activation, the faster the retrieval. If there are competitor items that match some of the retrieval cues, they receive some spreading activation, As a result, less activation reaches the target, and the target is retrieved more slowly. Therefore, the ACT-R model predicts that retrieval interference leads to a processing slowdown.

In turn, McElree and colleagues (Martin & McElree, 2011; McElree, 2000; McElree et al., 2003) suggested that while items are retrieved from memory by means of retrieval cues, the retrieval speed remains constant irrespective of the number of competitors. But constant retrieval speed does not imply constant reading times: McElree proposes that reading times represent not only the retrieval speed, but also the probability of successful retrieval — if misretrieval occurs, parser initiates a reanalysis, which takes time. Consequently, according to McElree, reading times are not diagnostic of retrieval speed, only the speed-accuracy tradeoff paradigm allows us to tease apart retrieval probability and latency. In the studies presented in this paper, we will rely on the ACT-R framework and its predictions regarding the speed of retrieval (reflected in reading times) as an indicator of interference.

The types of interference listed above are not mutually exclusive: encoding/ maintenance and retrieval interference can affect working memory independently, which is exactly what the Oberauer and Kliegl (2006) model assumes. In the psycholinguistic literature, there are very few experiments that pit the predictions of these types of interference against each other. Some exceptions — experimental results that clearly favor certain types of interference even if not rule out the others — will be reviewed below.

2.0.2 Interference effects in language processing

There are some similarity-based interference effects that can be explained only by interference arising during encoding and/or maintenance processes. The most notable example comes from the experiment of Gordon et al. (2001; replicated in Gordon et al., 2006), where participants were reading sentences such as (1):

(1) a. It was the barber/John that _ saw the lawyer/Bill in the parking lot.
b. It was the barber/John that the lawyer/Bill saw _ in the parking lot.

The authors reported that noun phrases differing in type (a common noun paired with a proper noun and vice versa) decrease reading times for object-extracted relative clauses (such as (1-b))[1] and increase question response accuracies. As retrieval occurs at the gap site, where no information about the noun type is provided, it cannot be retrieval interference that penalizes the processing of sentences with two nouns of the same type. On the contrary, encoding/maintenance interference easily accommodates these results: as the representation of similar items in working memory is degraded, retrieval of these items takes more time and is more error-prone.

In a different study, Gordon et al. (2002; see also Fedorenko et al., 2006) explored the influence of an increased memory load in a dual-task paradigm: the original sentences from Gordon et al.'s 2001 experiment with either both proper or both common nouns were preceded with triplets of proper (*Joel–Greg–Andy*) or common (*poet–voter–cartoonist*) nouns that participants had to memorize. As expected, the match between the type of nouns in memory and the ones in the sentence increased reading times and the number of errors in the answers to the comprehension questions. This effect was even stronger in the syntactically more complex object relative clauses. Again, only encoding interference can explain these results since there are no retrieval cues that could specifically trigger retrieval of only proper or common nouns and penalize the processing of sentences with similar noun types.

Retrieval interference effects, in turn, were demonstrated by Van Dyke and McElree (2006; see also Sekerina et al., 2016) in a memory-load paradigm similar to Gordon et al.'s (2002) experiment (2):

(2) a. table–sink–truck/∅
It was the boat that the guy who lived by the sea *sailed* in two sunny days.
b. table–sink–truck/∅

[1]No difference was found in subject relative clauses, such as 1a.

It was the boat that the guy who lived by the sea *fixed* in two sunny days.

While Gordon et al. (2002) manipulated the similarity between the memory load and the retrieval target, Van Dyke and McElree (2006) manipulated the match between the memory load and the retrieval cues provided by the semantics of the verb. As a result, reading times at the verb increased in condition (2-b) as compared to (2-a), but only when a memory set was present. The authors interpret these findings as evidence for interference during cue-based retrieval: semantic retrieval cues provided by the verb *sailed* can uniquely identify the to-be-retrieved item in memory (boat), while the cues provided by the verb *fixed* are compatible with all the items held in memory (table, sink, truck, and boat), which causes interference during retrieval and, therefore, a processing slowdown.

In another study, Van Dyke (2007; see also Van Dyke and Lewis, 2003) explored both syntactic and semantic interference arising within one sentence. Participants were presented with items such as (3):

(3) The worker was surprised that the resident...

a. who was living near the dangerous warehouse
b. who was living near the dangerous neighbor
c. who said that the warehouse was dangerous
d. who said that the neighbor was dangerous ...was complaining about the investigation.

The authors reasoned that to retrieve the subject while processing a verb, syntactic as well as semantic retrieval cues may be used, and indeed, a slowdown was found both in conditions with syntactic ((3-c) and (3-d)) as well as semantic ((3-b) and (3-d)) distractors. Note that these results are compatible with the encoding interference account: during encoding and maintenance both semantically and syntactically similar nouns would be predicted to lose features they share, and hence would be more difficult to retrieve. Basically, both encoding and retrieval interference accounts predict identical results in this setup. The same criticism applies to Van Dyke and McElree's 2011 study with similar experimental conditions as well as to studies by Martin and McElree (Martin & McElree, 2009, 2011).

Therefore, although many studies are conducted with the retrieval interference framework in mind, few experiments clearly demonstrate the effects of retrieval interference that cannot be explained by interference during memory encoding/ maintenance. Also, it should be noted that the only unambiguous evidence for retrieval interference comes from experiments manipulating semantic cues (Van Dyke, 2007; Van Dyke & McElree, 2006). There is, however, a common potential limitation in the studies discussed so far: they explore interference in subject-verb and filler-gap dependencies, where the second part of the dependency is predictable as soon as the first is encountered (e.g., encountering a filler posits existence of a gap later in the sentence); therefore, subjects and fillers might be maintained in focal attention (McElree, 2006), and not retrieved at encountering the verb or the gap. A more convincing demonstration of retrieval interference would come from a dependency where the first element does not posit the existence of the second, such as a retrieval of a pronoun's or a reflexive's antecedent. Indeed, many studies are investigating interference in anaphor resolution. We will discuss these studies next.

2.0.3 Interference effects in anaphor processing

In syntax, the Binding Theory (Chomsky, 1981) identifies strict syntactic constraints defining the set of grammatical antecedents for pronouns and reflexives. The question whether these constraints are considered from the early stage in online processing (Nicol & Swinney, 1989) or applied as a later filter (Badecker & Straub, 2002) has been studied extensively. Researchers tested whether distractors that are not licit antecedents of pronouns and reflexives affect anaphor resolution.

In pronouns, clear interference effects were found in some studies, as in Badecker and Straub (4):

(4) a. John thought that Bill owed him another chance to solve the problem.
    b. John thought that Beth owed him another chance to solve the problem.

In condition 4a where both the antecedent and the structurally inaccessible distractor match in gender, reading times after the pronoun *him* were elevated in comparison to condition 4b. These results are interpreted as demonstrating interference from the distractor, and the authors conclude that grammatical constraints do not rule out grammatically illicit attachment sites at an early stage of processing. This conclusion was supported by a number of other studies (Clackson et al., 2011; Kennison, 2003; Runner & Head, 2014). However note that several experiments failed to observe interference effects in pronouns (Chow et al., 2014; Cunnings et al., 2015; Patterson et al., 2014).

In reflexive binding a contradictory pattern of results is emerging: many studies found interference effects, which is inconsistent with the syntax as early filter account (Sturt, 2003), but at least as many other studies did not. For example, Badecker and Straub reported a slowdown two words downstream the reflexive when distractor matched the gender of the reflexive's antecedent (from now on, *interference* condition), as in (5):

(5) a. Jane thought that Bill owed himself another opportunity to solve the problem.
    b. John thought that Bill owed himself another opportunity to solve the problem.

Similar results were observed in several other studies (Chen et al., 2012; Clackson and Heyer, 2014; Nicol et al., 2003; Jäger, Benz, et al., 2015, Experiments 1 and 2; Jäger, Engelmann, et al., 2015, Experiment 2 in grammatical conditions, Experiment 1 in ungrammatical conditions; and Patil et al., 2016). In addition, several studies reported a speed-up in the interference condition (Sturt, 2003, Experiment 1; Cunnings and Felser, 2013, Experiment 2; Baumann and Yoshida, 2015; Cunnings and Sturt, 2014; Jäger, Benz, et al., 2015, Experiment 3). Overall, in a meta-analysis Jäger et al., 2017a no evidence was found for interference in experiments on reflexives with materials such as 5a and 5b. We will discuss the slowdown vs. speed-up interference effects in more detail in Section 2.2.5.

Interference effects were also found in a visual-world eye-tracking paradigm: Runner and Head (2014, see also Clackson and Heyer, 2014) demonstrated that distractors matching the gender of the antecedent attracted participants' attention from the onset of the reflexive more than gender-mismatching distractors, which means that participants at least sometimes attempted to bind the reflexive to the distractor. The same effects were also found in children (Clackson et al., 2011). It is not straightforward to decide whether this result patterns with a slowdown or a speed-up in reading times, but it clearly demonstrates the presence of interference effects.

However, as mentioned earlier, many experiments failed to observe any interference effects (Clifton et al., 1999; Nicol and Swinney, 1989; Badecker and Straub, 2002, Experiments 5, 6; Sturt, 2003, Experiment 2; Clackson et al., 2011; Dillon et al., 2013; King et al., 2012; Kush and Phillips, 2014; Parker and Phillips, 2016; Xiang et al., 2009). We will return to this point and discuss possible reasons for the lack of interference effects in reflexive processing later in this paper. For a more in-depth literature review of interference effects in reflexives, refer to Jäger et al. (2017a).

Most studies that targeted similarity-based interference in reflexives did not explicitly aim to test which type of interference affects reflexive processing (one exception is Jäger, Benz, et al., 2015), but rather assumed that interference arises during retrieval, when the parser is processing the reflexive and triggers the search for its antecedent. Since in most languages in which the studies were conducted, reflexives are gender- and number-marked, the reflexive's gender and number are likely to be used as retrieval cues, and all the items in memory with features that match those cues would compete for retrieval. Thus, whenever interference effects were found, they were attributed to this competition for retrieval and seen as evidence against syntax as an early filter account (Nicol & Swinney, 1989). However, Dillon et al. suggested that it might be not retrieval, but rather encoding interference that influenced the processing of reflexives. Within the encoding interference framework, if two (or more) words with the same gender and number marking are encoded to the working memory, the representation of these words would be degraded, and retrieval of those words would take more time and fail more often. If this hypothesis turns out to be true, interference effects in the literature cannot be interpreted as unambiguous evidence for retrieval interference and hence as evidence against syntax as an early filter account.

Jäger, Benz, et al. tested the encoding interference account and its predictions directly: in German, the reflexive *sich* is not gender-marked; as a result, gender cannot be used as a retrieval cue. Consequently, retrieval interference is not expected to influence the processing of sentences with gender match between the antecedent and distractor in German. In contrast, encoding interference is expected to occur any time two similar items are written to working memory, and would manifest itself in longer retrieval times and more retrieval errors. In two experiments with relatively large number of participants Jäger, Benz, et al. found no slowdown at or after the reflexive region and concluded that there is no evidence for encoding interference affecting online reflexive processing. However, some concerns were raised, mainly that the null result does not prove the absence of an effect. In Experiment 3 on Swedish possessives, a more direct evidence in favor of retrieval interference was found: fewer first-pass regressions were observed in the interference condition when possessives were gender-marked in contrast to the gender-unmarked. However, as possessives might be processed differently than reflexives, the conclusions one might draw from this result are still limited.

This brings us directly to the main point of the present paper: to find out whether it is encoding or retrieval interference that affects anaphor processing. The first of the three presented experiments contrasts reflexive and pronoun processing in German: an interference effect in pronouns and an absence of the effect in reflexives in the same sample would provide more convincing evidence against encoding interference.

2.1 Experiment 1: German reflexives and pronouns

As mentioned above, reflexives do not bear any gender marking in German; therefore, the gender feature cannot be used for retrieval, and no retrieval interference is expected if the antecedent and the distractor share the same gender. In contrast, German pronouns are gender-marked, hence gender might be used for the retrieval of the pronoun's antecedent. If we observe interference effects in pronouns but not reflexives, one can conclude that the source of interference is the retrieval process rather than processes happening during encoding or maintenance. On the other hand, if we find interference effects both in pronouns and reflexives, retrieval interference is not able to account for that pattern and we can conclude that the interference is caused by processes during memory encoding or maintenance.

2.1.1 Materials and Methods

We designed 42 sets of experimental items, manipulating interference (match or mismatch in gender between the antecedent and the distractor) and dependency type (reflexive, pronoun, or a noun phrase that does not trigger retrieval). This resulted in a 2×3 design, see Example (6). Sentences were constructed such that the reflexive/pronoun preceded the main verb in order to avoid reactivation of the antecedent before processing the anaphor. Both the antecedent and the distractor were subjects of their respective clauses and had nominative case marking in order to increase the chance to observe an effect (there is evidence suggesting that distractors in subject position induce stronger interference, see Jäger et al., 2017a). The experimental items consisted of three clauses: the main clause served as preface, while the subordinate clauses contained the actual experimental manipulation. We opted for this structure since only in a subordinate clause does German syntax allow the reflexive/pronoun to precede the main verb. The subordinate clause contained a subject (the antecedent of the reflexive) modified by a dative relative clause with the distractor in subject position, matching or mismatching the reflexive's antecedent in gender. Note that while for reflexives the antecedent is the subject of the second clause and the distractor is the subject of the dative relative clause, it is the reverse for the pronouns: the subject of the second clause is the distractor and the subject of the dative relative clause is the antecedent. We will discuss the materials with focus on the reflexive condition, but keep in mind that the order of target and distractor is reversed in the pronoun condition. The dative relative clause was followed by a direct object that triggered the retrieval in the pronoun/reflexive conditions. In the control condition this direct object was an animate noun phrase in neuter gender. Thus, no retrieval is triggered at the critical word, and therefore no difference between the interference and no interference conditions is expected. The spillover region was constant across conditions and contained a prepositional phrase and a verb. The experimental materials were additionally balanced by gender of the antecedent (21 items with a masculine and 21 with a feminine antecedent).

All materials, results and analysis files for all the experiments reported in this paper can be downloaded from Open Science framework (https://osf.io/xfthm/).

(6) a. Das Journal schreibt, dass der Bürokrat, dem der
The journal writes that the$_{masc}$ bureaucrat$_i$ the$_{Dat,\ masc}$ the$_{masc}$
Schriftsteller geraten hat umzudenken, **sich/ihn/das Mitglied** in
writer$_j$ advised to reconsider **self$_i$/him$_j$/the$_{neu}$ member** in
dem gigantischen Einkaufszentrum blamiert hat.
the giant mall embarrassed has.
The journal writes that the bureaucrat, whom the (male) writer advised to rethink, embarrassed himself/him/the member in the giant mall.

b. Das Journal schreibt, dass der Bürokrat, dem die
The journal writes that the$_{masc}$ bureaucrat$_i$ the$_{Dat,\ masc}$ the$_{fem}$
Schriftstellerin geraten hat umzudenken, **sich/sie/das Mitglied**
writer$_j$ advised to reconsider **self$_i$/her$_j$/the$_{neu}$ member**
in dem gigantischen Einkaufszentrum blamiert hat.
in the giant mall embarrassed has.
The journal writes that the bureaucrat, whom the (female) writer advised to rethink, embarrassed himself/her/the member in the giant mall.

Each sentence was followed by a yes/no comprehension question (see Example (7)). Half of the questions asked about the antecedent, and the other half about the distractor. The questions were balanced with regard to the number of yes/no answers. They were designed in such a way as to not repeat the lexical material of the corresponding sentence and required deep semantic processing of the sentence.

(7) Blieb dem Bürokraten eine Blamage erspart?
Was the$_{Dat}$ bureaucrat$_{Dat}$ an embarrassment spared?
Was the bureaucrat spared the embarrassment?

Experimental items were mixed with 83 filler sentences.

Participants completed a moving-window self-paced reading experiment programmed in Linger (Rohde, 2005). The order of presentation was pseudorandomized such that each experimental item was followed by at least one filler; each session started with five practice trials to help participants get used to the task.

2.1.2 Participants

111 participants were tested at the University of Potsdam in exchange for course credit or payment of 5 Euros. All participants were neurologically healthy native speakers of German, mostly students of the University of Potsdam. Their demographic data

were not recorded.

2.1.3 Analysis

Nicenboim et al. provide persuasive evidence that participants who do not complete syntactic dependencies and resort to guessing the answer to the comprehension questions process linguistic input qualitatively different from participants who answer questions correctly: individuals who fail to build a correct representation of the sentence read the critical retrieval region faster. Therefore, it is undesirable to conflate the data from these different categories of participants in one analysis: the slowdown in reading times of accurate participants might be concealed by a speedup in reading times of participants who do not parse the syntactic structure correctly. To avoid this, we included mean accuracy in answering the comprehension questions to experimental items as a predictor in the models of reading times. Mean participant accuracy is a reasonable approximation of the probability with which any given trial would be processed successfully by certain participant. We decided against trial accuracy because of the implicit assumption that every trial which resulted in a correct response was processed successfully. This is not necessarily true: a participant might fail in processing most of the trials but still provide correct responses for half of them due to chance. Mean subject accuracy better accounts for such cases at the expense of trial level variation.

We fit linear mixed-effects models using R (R Core Team, 2016) to the reading times from four regions: *a*) the relative clause participle (*umzudenken*); *b*) the critical region containing reflexive, pronoun, or NP (*sich/(ihn/sie)/das Mitglied*); *c*) the preposition and article after the critical region (*in dem*); and *d*) the adjective (*gigantischen*).[2]

For analysis, reading times were log-transformed. Whenever the residuals were not normally distributed, we checked whether deletion of problematic data points changed the results using the package "influence.ME" (Nieuwenhuis et al., 2012). In no case did exclusion of problematic data points change the results. For linear mixed-effects models, the "lme4" package version 1.1-8 (Bates et al., 2015) was used. Sum contrast coding was used to test the main effects and interactions. In addition,

[2]Hereafter the illustrations will always refer to the example item, in that case, Example (6).

	Noun phrase	Pronoun	Reflexive
Interference	0.63(0.018)	0.56(0.019)	0.61(0.018)
No interference	0.67(0.018)	0.70(0.018)	0.69(0.018)

Table 2.1: Mean accuracies and standard errors by conditions.

pairwise comparisons were modeled by applying sum contrasts nested within each level of dependency type factor whenever the interaction was significant.

For the analysis of response accuracies, linear mixed-effects models with a logistic link function were used. The model of question response accuracy included main effects of dependency type and interference as well as by-subject and by-item random intercepts and slopes for the main effects, but not for the interaction due to non-convergence of the full model.

The reading times models included main effects of interference, dependency type, and mean participant accuracy (centered and scaled, i.e. z-scores), the three-way interaction between them, as well as two-way interactions between dependency type and interference, and accuracy and interference. The random part of the models included random intercepts for subjects and items as well as by-item random slopes for all main effects, and by-subject random slopes for the main effects of match and dependency type. As mean accuracy is a between- rather than within-subjects predictor, it was not included into by-subject random slope structure. Interactions between main effects were also not included in the random effects structure of the model due to convergence problems.

2.1.4 Results

Accuracy

The mean accuracy rates across conditions and the corresponding standard errors are presented in the Table 2.1.

Mean accuracies by participant ranged from 0.40 to .90, with a mean of 0.64. 53 out of 111 participants had mean accuracies below chance level (defined as the highest number of mistakes a participant could make such that exact binomial test would still result in a p-value of 0.05 or lower, indicating that the number of correct responses was above chance; 14 mistakes in this experiment).

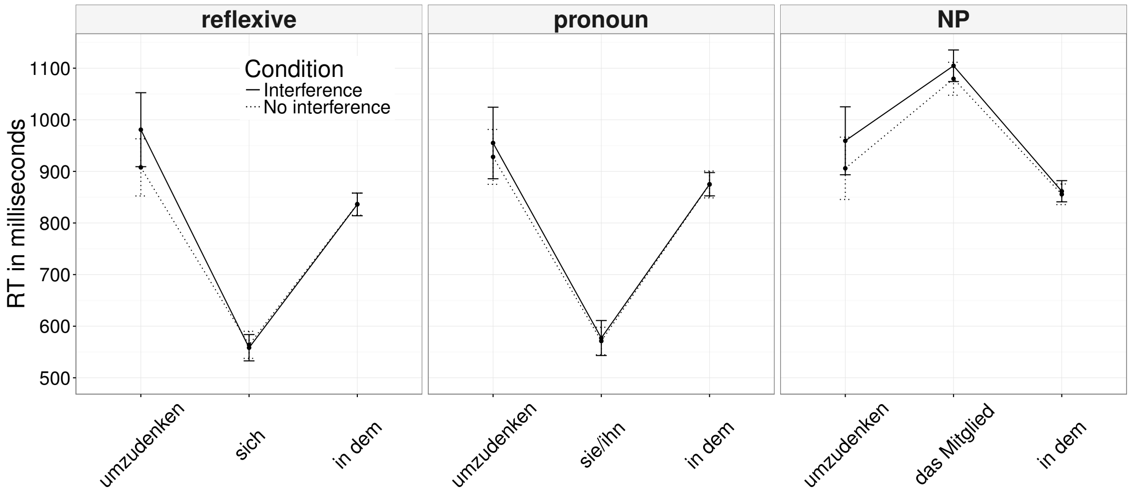

Figure 2.1: Mean reading times across conditions and their confidence intervals (Experiment 1).

Statistical analysis revealed a main effect of interference: accuracy was lower in the condition where the antecedent and the distractor shared the same gender ($\hat{\beta}$ = -0.46, SE = 0.11, z = 4.07, $p < 0.001$). There was a significant interaction between the effect of interference and the dependency type ($\hat{\beta}$ = 0.25, SE = 0.10, z = 2.44, p = 0.02). The model with pairwise comparisons revealed that in the conditions with reflexives and pronouns as compared to nouns, accuracy was lower when the antecedent and the distractor shared the same gender ($\hat{\beta}$ = -0.45, SE = 0.14, z = -3.28, $p < 0.01$ for reflexives; $\hat{\beta}$ = -0.81, SE = 0.26, z = -3.05, $p < 0.01$ for pronouns), but the effect was not present in the control condition with nouns.

Reading times

Mean reading times and their respective confidence intervals for the analyzed regions across conditions are presented in Figure 2.1.

In the pre-critical region (the verb *umzudenken* in Example (6)) a significant main effect of participants' mean accuracy was found (see Table 2.2): more accurate participants read the region more slowly. There was also a significant three-way interaction between interference, dependency type, and accuracy, but since the conditions were identical for both dependency types at that region, we discard this result as a Type I error. In the critical region, dependency type significantly affected reading times: both reflexives and pronouns were read faster than nouns. There was also a significant main effect of accuracy: the region was read more slowly by the

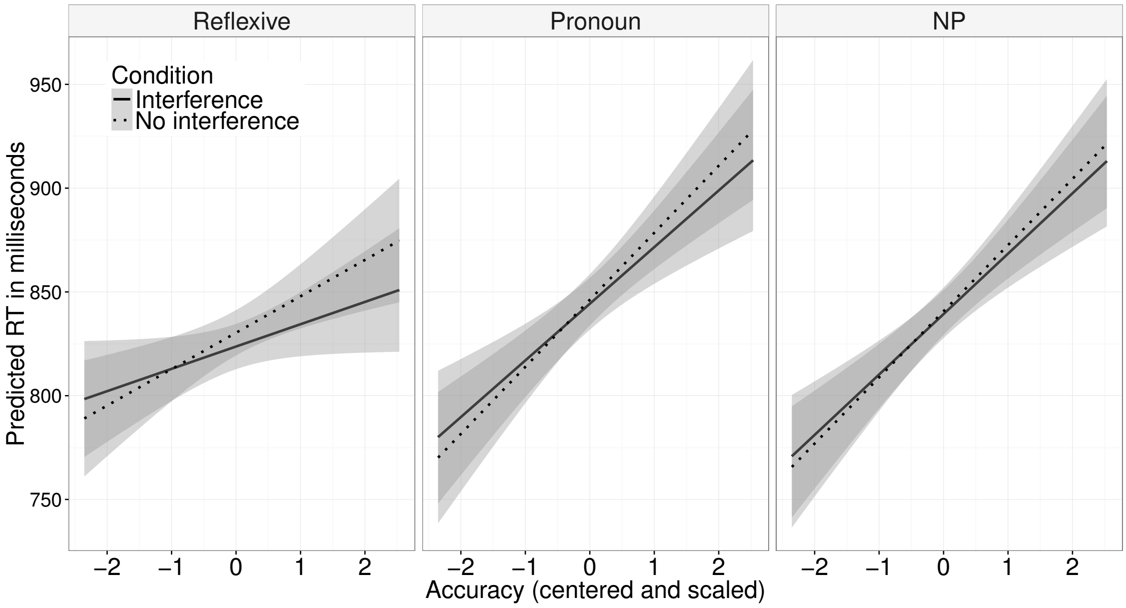

Figure 2.2: Modeled reading times (and respective standard errors) at the spillover after critical region (Experiment 1).

more accurate participants. For the analysis of reading times in the post-critical region by-item random slopes for the main effects of dependency type and accuracy were removed due to non-convergence of the model. We opted for eliminating by-item random slopes since by-item variance is usually smaller than by-subject. In this region, again, dependency type significantly affected reading times: the region was read faster in conditions where the direct object was a reflexive in comparison to a noun. There was also a three-way interaction between dependency type, interference, and accuracy (see Figure 2.2). Nested contrasts demonstrated that the interaction was driven by a two-way interaction between accuracy and dependency type: mean accuracy had less influence on the speed of reading the post-critical region after reflexives than after nouns ($\hat{\beta}$ = -0.012, SE = 0.004, t = -3). No other comparisons were significant in any region.

Table 2.2: Main effects of interference, dependency type, accuracy, and their interaction on log-transformed RTs by regions. Standard errors are given on the same scale as the estimates and represent changes to the last decimal point(s) of the estimate. For example, 0.021(8) stands for the effect of 0.021 and its SE of 0.008 (both on the log-ms scale).

	Pre-critical		Critical		Post-critical 1		Post-critical 2	
	umzudenken		*sich/ihn/sie/das Mitglied*		*in dem*		*gigantischen*	
	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t
Interference	.011(7)	1.46	.002(5)	0.32	-.002(3)	-0.7	.003(4)	0.71
Reflexive vs. NP	.004(12)	0.36	-.234(8)	-29.45	-.010(4)	-2.4	-.001(54)	-0.02
Pronoun vs. NP	.011(12)	0.92	-.223(9)	-25.16	.006(4)	1.5	-.001(5)	-0.22
Accuracy	.136(47)	2.90	.058(22)	2.53	.029(18)	1.6	.024(25)	.97
Interf.×Refl.	.008(17)	0.77	-.003(7)	-0.38	-.002(4)	-0.5	.001(5)	0.19
Interf.×Pron.	-.014(10)	-1.37	-.005(7)	-0.74	.001(3)	0.2	-.007(5)	-1.26
Interf.×Acc.	.005(7)	0.66	-.004(5)	-0.73	-.003(3)	-0.9	.005(4)	1.41
Interf.×Refl.×Acc.	-.033(16)	-2.06	.002(10)	.16	-.013(6)	-2.2	.002(8)	.20
Interf.×Pron.×Acc.	.027(15)	1.74	-.006(10)	-.61	.006(6)	.9	-.003(8)	-.39

2.1.5 Discussion

The comparison of interference effects in reflexives and pronouns revealed that question response accuracy was lower in the conditions with reflexives and pronouns when the antecedent and the distractor shared the same gender. The effect was not present in the control condition. This pattern can be explained by encoding interference, but is inconsistent with retrieval interference: when the distractor shares the gender of the antecedent, accuracy is lower independently of the anaphor type: interference is present both in gender-unmarked reflexives and in gender-marked pronouns. No difference in accuracy in the control condition with nouns is consistent with the notion that interference manipulation affects only those sentences where retrieval of the antecedent should happen. This pattern replicates the findings for German reflexives reported by Jäger, Benz, et al. (2015) in Experiments 1 and 2. However, question response results should be interpreted with caution since we were primarily testing the predictions of the interference accounts with respect to the reading times, and comprehension question accuracies might reflect processes different from those of online sentence comprehension.

It is also unclear why the overall question response accuracy was so low. It might be the case that the double nested syntactic structure was too challenging for our participants. Another factor that might have affected participants' performance was the nature of comprehension questions (see (7)): answering the question correctly required making inferences about the situation described in the experimental sentence, and not just remembering the propositions. To our knowledge, comprehension questions in most experiments are easier to answer and probe either the superficial understanding of the sentence ("Was anyone embarrassed?") or the dependency resolution ("Who was embarrassed?"). It might be possible that the combination of the double nested syntactic structure together with the demanding comprehension questions was too difficult for many participants.

An interesting point that does not directly relate to the main purpose of the study is that for the pre-critical and critical regions we found that participants' mean accuracy and reading times are correlated: participants who resolve syntactic dependencies correctly read more slowly (see also Ferreira et al., 2002). This replicates and extends the findings of Nicenboim et al. that participants who do not answer comprehension questions correctly tend to rush through the retrieval site. In our case, the effect is present not only at the retrieval site, but also at the pre-critical region. It is probable that less accurate participants might read the whole sentence more quickly. This might be explained by the limitations of working memory resources: those participants with lower WM capacity try not to lose the unresolved dependencies they have to keep track of, and speed up in order to resolve the dependencies and lift the burden as quickly as possible. However, since we did not measure participants' working memory, this must remain a speculation.

Unfortunately, we found no main effect or interactions involving the interference manipulation in reading times, and thus no evidence in favor of either encoding or retrieval interference. If anything, this suggests that there are no interference effects in the processing of anaphor dependencies, but one must be cautious interpreting the absence of the effect in favor of the null hypothesis. In addition, comparing reflexives with pronouns is potentially problematic. Interestingly, we found that the post-critical region was read faster when the critical region contained a reflexive in comparison to a noun (and even faster by more accurate participants). No such speedup was present in the post-critical region after a pronoun, although both reflexives and pronouns were read faster than nouns in the critical region. The fact that this speedup was independent of the interference manipulation suggests that it might reflect syntactic processing differences between reflexives and pronouns, whose interpretation is subject to different syntactic constraints. A better experimental design would allow us to compare gender-marked with gender-unmarked reflexives, which is not possible either in English or in German. Luegi et al. (2016) contrasted gender-marked and gender-unmarked reflexives in Portuguese, but did not find any difference in online processing. One of the possible reasons could be that in European Portuguese, the gender-marked reflexives are split constructions: first, a reader encounters an unmarked reflexive (*se*), then a verb, and only after the verb comes the gender-marked part of the reflexive (*a si mesmo/mesma*). In such configuration, retrieval is triggered at encountering the first, gender-unmarked, part of the reflexive. A better experimental design is possible in Russian, which allows us to test different interference accounts' predictions within one language.

2.2 Experiment 2A: Russian reflexives, reflexive precedes the verb

Russian has two types of reflexives with the same syntactic distribution and with binding rules generally close to those of English and German (analogous in all aspects relevant for our research question; for more detail on Russian reflexive binding, see Rappaport, 1986): gender-unmarked *sebja* (similar to German *sich*) and gender-marked *samu/samogo sebja* (similar to English *herself/himself*). This provides us with an opportunity to pit retrieval and encoding interference predictions directly against each other: the encoding interference account would predict the slowdown in the conditions where the distractor shares the gender of the antecedent, irrespective of the reflexive type. The retrieval interference account, in turn, would predict an interaction between the reflexive type and the presence/absence of interference: only in the gender-marked reflexives would gender be used as a retrieval cue; and hence we should expect an interference effect only in the gender-marked reflexives, but not in the gender-unmarked reflexives.

2.2.1 Materials and Methods

We designed 32 sets of experimental items, manipulating in a 2×2 design the interference and type of reflexive (gender-unmarked *sebja* vs. gender-marked *samogo/ samu sebja*). Experimental items consisted of a main clause and an embedded relative clause (see Example (8)). The main clause subject, the reflexive's antecedent, was followed by an object-extracted relative clause containing the distractor noun (matching or mismatching the main clause subject in gender) in subject position. The relative clause was followed by the reflexive (gender-marked or gender-unmarked), an adverb, and the main clause verb. All the verbs were in present tense in order to avoid the gender marking on the verbal past in Russian. Additionally, in the relative clause all nouns except for the distractor had neutral gender.

(8) a. Аферистка$_{i}$, которую **торговка** нанимает для ограбления,
Swindler$_{fem}$ whom **merchant$_{fem}$** hires for robbery,
себя$_{i}$/саму себя$_{i}$ серьёзно переоценивает в способности к
self$_{acc(\emptyset)}$/herself$_{acc(fem)}$ significantly overestimates in ability to

обману.
do trickery.
The swindler$_{fem}$, whom a merchant$_{fem}$ hires for a robbery, significantly overestimates her own$_{\emptyset/fem}$ trickery skills.

b. Аферистка$_i$, которую **торговец** нанимает для ограбления,
Swindler$_{fem}$ whom **merchant$_{masc}$** hires for robbery,
себя$_i$/саму себя$_i$ серьёзно переоценивает в способности к
self$_{acc(\emptyset)}$/herself$_{acc(fem)}$ significantly overestimates in ability to
обману.
do trickery.
The swindler$_{fem}$, whom a merchant$_{masc}$ hires for a robbery, significantly overestimates her own$_{\emptyset/fem}$ trickery skills.

Within an experimental item, the antecedent and both the matching and mismatching distractors had the same length (counted in number of syllables), and their lemma frequency never exceeded 100 tokens per million (Lyashevskaya & Sharov, 2009). Experimental materials were additionally balanced by gender of the antecedent (16 masculine, 16 feminine) and by noun type (16 experimental items had proper nouns, and 16 had common nouns). We employed proper nouns because distractors had to differ in gender but have the same word length within each item, and Russian has a very limited number of such common noun pairs.

Within an experimental item, the difference in frequency between matching and mismatching distractors did not exceed 50 tokens per million in common nouns and 10 tokens per million in proper nouns. The difference in frequency between the feminine and masculine antecedents across items was not significant, and neither was the difference between matching and mismatching distractors across items.[3]

The structure of 32 filler sentences superficially resembled the one of the experimental items in order to hide the experimental manipulation effectively. Each filler sentence consisted of a main and an embedded relative clause, but in contrast to the experimental items, the relative clause was subject-extracted. This discouraged participants from developing a strategy to process every sentence as containing an object relative clause, and encouraged deep structure processing. In fillers, the nouns in the main and the relative clauses had the same gender in half of the filler sentences. The fillers were additionally balanced by gender of the first noun (16 feminine, 16

[3]For the feminine and masculine antecedents: Wilcoxon rank sum test, $W = 140$, $p = 0.67$; for the matching and mismatching distractors: Wilcoxon rank sum test, $W = 496$, $p = 0.83$.

masculine) and by noun type (16 fillers had proper, and another 16 had common nouns). Instead of a reflexive, a verb with a reflexive postfix (-*sja*, which does not necessarily convey reflexive meaning in Russian) was used. An example of a filler sentence is given below in (9):

(9) Студент, который зазвал приятеля на вечеринку, основательно
Student$_{masc}$, who invited friend$_{masc}$ to party, a lot
закупается продуктами.
buys$_{sja}$ of food.
A student who invited his friend to a party buys a lot of food.

Each sentence was followed by a wh- comprehension question with two answer options to choose from (see an example comprehension question for the experimental item in (10)). In experimental items, 11 questions probed for the antecedent, 11 for the distractor, and 10 superficial questions probed for the adjuncts. To distract participants from the reflexive-antecedent dependency, in filler sentences, 20 questions probed for the adjuncts, six probed for the subject of the main clause, and the remaining six probed for the object of the relative clause. Questions were counterbalanced within each experimental list. In the questions neither lexical reflexives nor the lexical material from the experimental items were used to discourage superficial processing.

(10) Кто высоко оценивает свои способности?
Who highly thinks of own abilities?
Who thinks highly of his/her own abilities?

Each participant was assigned to one of four experimental lists arranged in a Latin square design. Each list consisted of 32 experimental items (each participant saw only one version of each item) and 32 fillers (the same across the lists). The order of experimental items and fillers was pseudo-randomized and controlled for the noun type (proper/common, maximum two of the same type in a row), question type (no more than three questions of the same type in a row) and for sentence type (experimental item/filler, no more than two of the same type in a row). In the beginning of each experimental session, the participant saw four training items.

Position of correct answers on the screen had a different randomization for each trial and participant.

2.2.2 Participants

109 volunteers completed a moving-window self-paced reading experiment programmed in Linger (Rohde, 2005). All participants were neurologically healthy native speakers of Russian, tested either at the Higher School of Economics (Moscow) or at the "Russian Reporter" Summer School. Mean age of participants was 21 (range 16-65), 17 out of 109 participants were male, 2 individuals reported to be left-handed. The study was approved by the Committee on Interuniversity Surveys and Ethical Assess of Empirical Research of the National Research University Higher School of Economics.

2.2.3 Analysis

The analysis was equivalent to the one described for the experiment on German (Section 2.1.3). The comprehension questions' responses were analyzed using a generalized linear mixed model with a logistic link function. The model included main factors of reflexive type and interference as well as interaction between them. The random effects structure included by-subject and by-item random intercepts and slopes for the main effects and their interaction.

As in Experiment 1, for reading time analyses, we computed participants' mean accuracy scores in answering the antecedent- and distractor-probing questions and used these scores as predictors. The linear models included main effects of reflexive type, interference, and accuracy, as well as the three-way interaction between these, the two-way interactions between reflexive type and interference, and accuracy and interference. The random effects structure included by-participant and by-item random intercepts and slopes for all the effects included in the model. By-participant random slopes did not include accuracy, as accuracy is a between-subjects predictor. For all linear models, correlations between random effects were not estimated.

We analyzed reading times data from the following four regions: *a*) the region preceding the reflexive (*for a robbery*); *b*) the reflexive (*sebja/samu sebja, self/herself*); *c*) the spillover after the reflexive (*significantly*); and *d*) the main clause verb (*overestimates*). Note that the reflexives *sebja* and *samogo/samu sebja* were presented and analyzed as one region. Consequently, we expected to find a trivial main effect of reflexive type in reading times: the gender-marked reflexive should take more time to be read simply because the region is longer.

2.2.4 Results

Accuracy

The mean accuracy rates across conditions and the corresponding standard errors are presented in the Table 2.3.

	Gender-marked	Gender-unmarked
Interference	0.81(0.014)	0.81(0.014)
No interference	0.88(0.012)	0.87(0.012)

Table 2.3: Mean accuracies and standard errors across conditions.

Mean participants' accuracies in answering the antecedent- and distractor-probing questions ranged from 0.45 to 1.00, with a mean of 0.79. 33 subjects out of 109 scored on average below chance (made more than six mistakes).

Statistical analysis revealed a main effect of interference: accuracy was lower in the conditions where the antecedent and the distractor shared the same gender ($\hat{\beta}$ = -.31, SE = .05, z = -5.86, $p < .001$). The effect of reflexive type and the interaction were not significant.

Reading times

Mean reading times and their respective confidence intervals for the analyzed regions across conditions are presented in Figure 2.3.

In the region preceding the reflexive, there were main effects of interference (a slowdown in the interference conditions), accuracy (more accurate participants read the region more slowly), and an interaction between these — more accurate participants slowed down even more when the antecedent and the distractor shared the same gender (see Table 2.4). In the reflexive region, we found a main effect of reflexive type with gender-unmarked reflexives being read faster than gender-marked reflexives, as expected given the respective region lengths. In the region following the

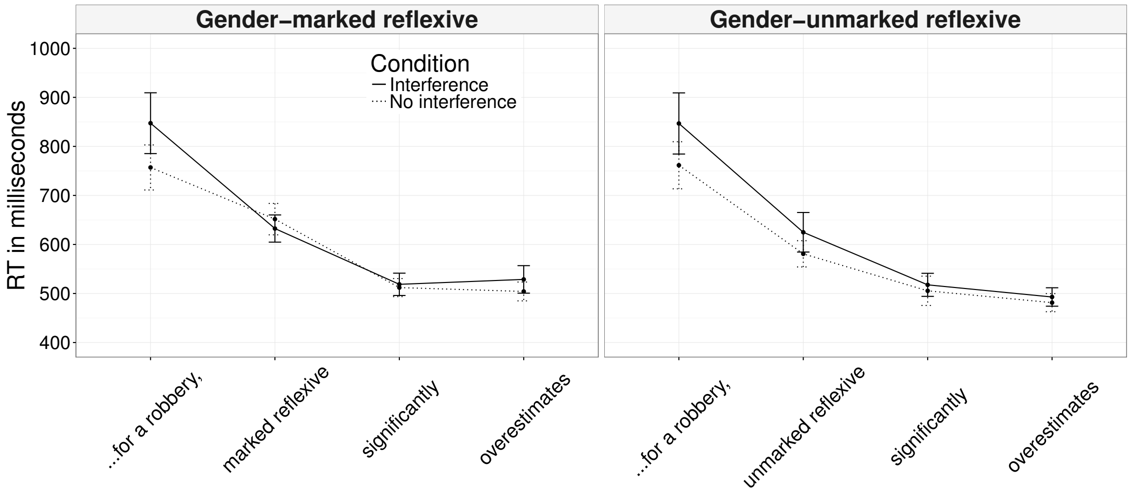

Figure 2.3: Mean reading times across conditions and their confidence intervals (Experiment 2A).

reflexive, we found an interaction between reflexive type, interference, and accuracy (see Figure 2.4). Nested contrasts testing for interference effects within each reflexive type and the interaction between these effects and accuracy did not reach significance. It seems that the interaction was driven by a difference within gender-unmarked reflexives that were read longer by more accurate participants in the interference condition ($\hat{\beta}$ = -.013, SE = .007, t = -1.66 for gender-marked reflexives; $\hat{\beta}$ = .013, SE = .007, t = 1.70 for gender-unmarked reflexives). In the following region (i.e., two words after the reflexive) we again found a main effect of reflexive type (the region was read more slowly in the conditions with gender-marked reflexives) and a main effect of interference (the region was read more slowly when the distractor matched the gender of the antecedent).

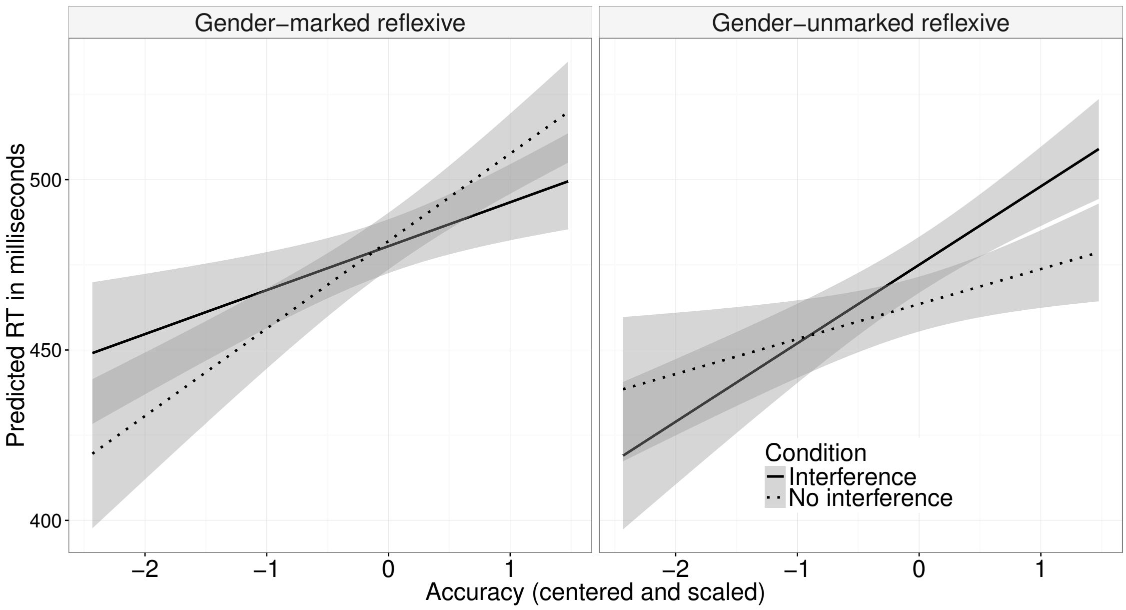

Figure 2.4: Modeled reading times (and respective standard errors) at the spillover after reflexive (Experiment 2A).

Table 2.4: Main effects of interference, reflexive type, mean accuracy, and their interactions on log-transformed RTs by regions. Standard errors are given on the same scale as the estimates and represent changes to the last decimal point(s) of the estimate. For example, 0.021(8) stands for the effect of 0.021 and its SE of 0.008 (both on the log-ms scale).

	Pre-reflexive		Reflexive		Adverb		Main verb	
	for a robbery		*sebja vs. samu/samogo sebja*		*significantly*		*overestimates*	
	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t
Reflexive type	.003(8)	.37	.034(7)	4.77	.013(7)	1.91	.024(5)	4.26
Interference	.025(9)	2.62	.006(7)	0.89	.006(6)	1.13	.012(5)	2.21
Accuracy	.109(40)	2.70	.032(25)	1.25	.031(21)	1.43	.021(22)	0.93
Int.×Acc.	.023(10)	2.26	.011(8)	1.28	.0002(70)	0.03	.003(6)	.54
Int.×Refl.	-.002(8)	-.28	-.007(7)	-.89	-.006(6)	-1.07	.0004(50)	.08
Int.×Refl.×Acc.	-.001(8)	-.17	-.014(8)	-1.71	-.013(6)	-2.01	-.013(7)	-1.89

2.2.5 Discussion

The experiment aimed at determining the type of interference that arises in reflexive processing: the encoding interference account predicts a slowdown in the interference condition independently of reflexive type, while the retrieval interference account predicts an interaction between the reflexive type and interference conditions.

In comprehension questions, similarly to Experiment 1, we observed more errors in the interference (gender match) conditions, irrespective of the reflexive type. This result is in line with with the encoding interference account and might reflect the degraded memory representation of the words that share certain features. An alternative explanation would be that this interference effect is due to some later processing that happens at the moment of answering the comprehension question, rather than due to online processes during reading.

In reading times, we found a main effect of interference at two regions: the word following the verb of the relative clause and the main verb. This is inconsistent with the predictions of the retrieval interference account: as the verbs were not marked for gender, gender could not be used as a retrieval cue, and the amount of retrieval interference should be the same regardless of gender match between the antecedent and the distractor. However, verbs were read more slowly in the conditions where the distractor matched the gender of the antecedent, which could only be explained by encoding interference: as two subjects of their respective clauses that share grammatical gender were written down to memory, their memory representations became less distinguishable, which affected retrieval speed and, consequently, slowed down reading times at the verb regions. Finding consistent evidence for encoding interference in processing subject-verb dependencies is an important result of the present experiment, but it does not necessarily translate to anaphoric dependencies.

The critical interaction that should allow us to disentangle the encoding and retrieval interference accounts in the processing of anaphoric dependencies was found in the region following the reflexive. However, the interaction went into an unexpected direction: we found that gender-unmarked reflexives were read more slowly in the interference condition by accurate participants, while there was no difference in the gender-marked reflexives across conditions. The slowdown in the gender-unmarked reflexives can only be explained by the encoding interference account (and is consistent with the evidence for encoding interference in subject-verb dependencies), but that account predicts a slowdown in the gender-marked reflexives that we do not observe. The retrieval interference account also standardly predicts a slowdown for gender-marked reflexives, although several studies reported a speedup (Sturt, 2003, Experiment 1; Cunnings and Felser, 2013, Experiment 2; Baumann and Yoshida, 2015; Cunnings and Sturt, 2014; Jäger, Benz, et al., 2015, Experiment 3). Similarly, in a study on anaphoric noun phrases, Autry and Levine (2014) found that increase in number of potential referents (from two to five) decreased rather then increased reading times at the noun phrase.

Although our results for the gender-marked reflexives are seemingly in conflict with the predictions of both interference accounts, we propose a post-hoc explanation that is consistent with the literature and with the ACT-R model: we suggest that both retrieval and encoding interference affect processing of gender-marked reflexives, and counteract each other. In that case, processing of both the gender-unmarked and the gender-marked reflexives is slowed down in the interference condition, but for the gender-marked reflexives, there is also a speedup in processing due to retrieval interference. Engelmann et al. have shown that a speedup in the interference condition is actually in line with the retrieval interference as implemented in ACT-R model of sentence processing under certain conditions. Engelmann et al. demonstrated that if a distractor is particularly activated and matches most of the retrieval cues, it would be misretrieved instead of the antecedent in a large proportion of trials. Due to a race-like scenario, the mean retrieval latencies will be faster in such a configuration (the more items are gaining activation, the sooner on average one of them crosses the activation threshold), which, in turn, would lead to a speedup in mean reading times in the respective condition. Since we constructed the experimental items such that the distractor is particularly prominent in order to maximize potential retrieval interference effects, it is reasonable to assume that the distractor was highly activated. Two factors contribute to the distractor's prominence: it occupies subject position and stays linearly closer to the reflexive than the antecedent. Earlier we mentioned that being a subject might be one of the retrieval cues (Van Dyke, 2007), and if this is the case, the distractor in our setup matches all but one retrieval cue (being an NP, gender, number, "subjecthood", but not c-command). Additionally, the meta-analysis (Jäger et al., 2017a) shows that distractors that are subjects of their clauses increase the amount of interference. As to recency, it contributes to the base-level activation of an item because ACT-R assumes decay: base-level activation decreases as time since the last retrieval of this item passes. To summarize, there are reasons to believe that in our design, distractors were particularly highly activated, which lead to a speedup due to retrieval interference, and that speedup counteracted the slowdown due to encoding interference in the gender-marked reflexives.

One of the reasons retrieval interference in reflexive dependencies is still a controversial subject is that many studies failed to find interference effects. Among possible reasons could be the insufficient number of participants and resulting low statistical power, or the joint analysis of the data from participants who are accurate in answering comprehension questions and participants who are at chance (see discussion is Section 2.1.3 and Nicenboim et al., 2015). As could be seen from the results of Experiment 2A, the interference effect is only found in the data from participants who generally answer the comprehension questions above chance. Thus our results can be seen as an additional evidence for the pattern proposed by Nicenboim et al.: participants who lack the resources to fully parse dependencies and are thus generally poor at answering comprehension questions often rush through the retrieval site and mask the effect that shows up in the data from the more accurate participants.

Another promising account explaining why retrieval interference effects are often not found in English was suggested by Parker and Phillips, who found that illusory negative polarity licensing is modulated by the position of the dependent element with regard to the verb (i.e., *ever* in the *no ... ever* dependency). The authors proposed that at the point of processing the verb, the part of sentence that precedes it is consolidated and becomes opaque for retrieval interference. For this reason, they argue, illusory licensing is possible only when both elements precede the verb, and does not occur when the dependent element follows the verb. Parker and Phillips suggest that the same might be true for reflexive processing. From this point of view, the distractor gets enclosed in the opaque representation that is not able to cause retrieval interference as soon as the main verb in encountered. If the reflexive follows the main verb, it is unable to retrieve the distractor from this representation, and hence no retrieval interference effects are observed at or following the reflexive.

Within the ACT-R framework, the position of the reflexive with regard to the main verb is also crucial, albeit for a different reason: the main verb triggers the retrieval of the subject, which is also the reflexive's antecedent. If the reflexive follows the verb and triggers the retrieval of its antecedent, the antecedent is relatively easy to retrieve since it has just received a boost of activation. Consequently, interference from the distractor is less likely to have any measurable effects. This might account for the lack of interference effects in many studies conducted in English, since in English, configurations where the reflexive precedes the main verb are structurally prohibited. There was at least one experiment that aimed at finding interference in a setup where reflexive preceded the verb (in Hindi), but no interference effects were found (Kush & Phillips, 2014). However, in this study the distractor did not bear ergative marking, which might have been one of the retrieval cues for Hindi.

It is possible that in Experiments 1 and 2A the antecedent of the reflexive might have been maintained in focal attention at the point of processing the reflexive, because the antecedent of the reflexive is also a subject that had not yet formed a dependency with the verb. In that case no retrieval would take place and no retrieval interference is expected. Whether an item in focal attention is predicted to be susceptible to encoding interference, must depend on the model of encoding interference one assumes. No model explicitly posits existence of the focal attention slot, but the model of Oberauer and Kliegl can be reconciled with it. Since both the reflexive's antecedent and the distractor are subjects of their respective clauses (and must both be in focal attention at some point during sentence processing), encoding interference might be possible. That account readily accommodates the slowdown in the interference condition for gender-unmarked reflexives, but fails to explain the absence of a slowdown in gender-marked reflexives: if there is no speedup due to retrieval interference, it is unclear why no slowdown due to encoding interference is found in reading times for the gender-marked reflexives. In any case, the focal attention explanation would be ruled out in a setup where the verb precedes the reflexive.

Our third experiment aims at testing Parker and Phillips' 2016 hypothesis that retrieval interference will be blocked if the main verb precedes the reflexive by replicating the second experiment with one important modification — the main verb and the manner adverb that followed the reflexive will now precede it.

2.3 Experiment 2B: Russian reflexives, reflexive follows the verb

Experiment 2B seeks to test the hypothesis that the relative order of the reflexive and the main verb might affect the presence of retrieval interference effects. In addition, we expect to replicate the encoding interference effects found in Experiment 2A on the main and relative clause verbs because word order should not affect encoding interference. For example, within the Oberauer and Kliegl (2006) model, both target and distractor have equal chances of losing a feature due to the proposed feature-overwriting mechanism and thus becoming less accessible. Therefore, retrieval of the target item given a feature-sharing distractor should have a longer latency and be more error-prone.

2.3.1 Materials and Methods

The experimental materials consisted of the same 32 sets of items as in Experiment 2A. In each sentence, the manner adverb and the main verb were placed between the relative clause and the reflexive. No other changes to the experimental materials were made. An example item is given in (11):

(11) a. Аферистка$_i$, которую **торговка** нанимает для ограбления,
Swindler$_{fem}$ whom **merchant$_{fem}$** hires for robbery,
серьёзно переоценивает **себя$_i$/саму себя$_i$** в способности к
significantly overestimates **self$_{acc(\emptyset)}$/herself$_{acc(fem)}$** in ability to
обману.
do trickery.
The swindler$_{fem}$, whom a merchant$_{fem}$ hires for a robbery, significantly overestimates her own$_{\emptyset/fem}$ trickery skills.

b. Аферистка$_i$, которую **торговец** нанимает для ограбления,
Swindler$_{fem}$ whom **merchant$_{masc}$** hires for robbery,
серьёзно переоценивает **себя$_i$/саму себя$_i$** в способности к
significantly overestimates **self$_{acc(\emptyset)}$/herself$_{acc(fem)}$** in ability to
обману.
do trickery.

The swindler$_{fem}$, whom a merchant$_{masc}$ hires for a robbery, significantly overestimates her own$_{\emptyset/fem}$ trickery skills.

The same procedure as in Experiment 2A was used, see Section 2.2.2.

2.3.2 Participants

112 volunteers who had not participated in the previous experiment took part in the study. All participants were neurologically healthy native Russian speakers and were tested at the Higher School of Economics, Moscow. Their mean age was 26 (range 16-70), 77 participants were female; 15 individuals reported to be left-handed or ambidextrous. The study was approved by the Committee on Interuniversity Surveys and Ethical Assess of Empirical Research of the National Research University Higher School of Economics.

2.3.3 Analysis

The data analysis was analogous to the one of Experiment 2A, see Section 2.2.3.

2.3.4 Results

Accuracy

The mean accuracy rates by condition and the corresponding standard errors are presented in the Table 2.5.

	Gender-marked	Gender-unnmarked
Interference	0.81(0.014)	0.76(0.015)
No interference	0.86(0.012)	0.85(0.012)

Table 2.5: Mean accuracies and standard errors by condition.

Participants' mean accuracies in answering antecedent- and distractor-probing questions ranged from 0.27 to 1.00 with a mean of 0.76. 34 out of 112 participants had mean accuracies below chance level (made more than 6 mistakes).

Statistical analysis revealed a main effect of interference: accuracy was lower in the conditions where the antecedent and the distractor shared the same gender ($\hat{\beta}$ = -.27, SE = .07, z = -4.13, p < .001). The main effect of reflexive type was also

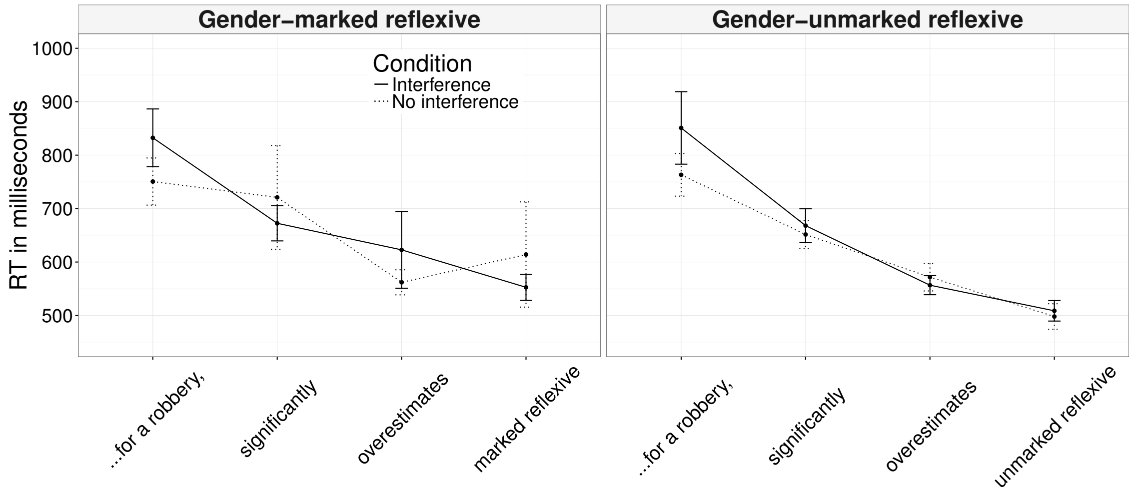

Figure 2.5: Mean reading times across conditions and their confidence intervals (Experiment 2B).

significant: accuracy was lower in conditions with gender-unmarked reflexives ($\hat{\beta}$ = .11, *SE* = .05, z = 2.29, p = .022). The interaction was not significant.

Reading times

Mean reading times and their respective confidence intervals for the analyzed regions for each experimental condition are presented in Figure 2.5.

Main effects of interference and accuracy were found in the region following the verb of the relative clause: the region was read more slowly by the more accurate participants and in the interference condition (see Table 2.6). In the two following regions (*significantly overestimates*) a main effect of accuracy was found: accurate participants read these two regions more slowly. In the reflexive region, we found a significant main effect of accuracy (accurate participants read the region more slowly) and an interaction between interference and reflexive type. Nested contrasts testing for interference effects within each reflexive type did not reach significance. It seems that the interaction was driven by the difference between interference and no interference conditions within gender-unmarked reflexives since there was no difference in the gender-marked reflexives ($\hat{\beta}$ = -.006, *SE* = .007, t = -.86 for gender-marked reflexives; $\hat{\beta}$ = .15, *SE* = .07, t = 1.94 for gender-unmarked reflexives).

Table 2.6: Main effects of interference, reflexive type, mean accuracy, and their interactions on log-transformed RTs by regions. Standard errors are given on the same scale as the estimates and represent changes to the last decimal point(s) of the estimate. For example, 0.021(8) stands for the effect of 0.021 and its SE of 0.008 (both on the log-ms scale).

	RC ending		Adverb		Main verb		Reflexive	
	for a robbery		*significantly*		*overestimates*		*sebja vs. samu/samogo sebja*	
	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t	$\hat{\beta}$(SE)	t
Reflexive type	-.006(7)	-0.83	.007(7)	.95	.004(6)	0.70	.039(6)	5.62
Interference	.021(8)	2.74	.002(7)	.30	.007(5)	1.37	.004(5)	0.72
Accuracy	.138(41)	3.31	.108(32)	3.35	.078(29)	2.65	.064(27)	2.36
Int.×Acc.	.013(7)	1.77	-.002(7)	-0.34	.009(5)	1.69	.003(5)	.54
Int.×Refl.	.0007(70)	.10	-.004(6)	-.72	.008(5)	1.57	-.011(5)	-2.08
Int.×Refl.×Acc.	-.004(7)	-.60	.006(6)	0.95	.005(5)	1.00	.001(5)	0.37

2.3.5 Discussion

Contrary to what is predicted by both the ACT-R cue-based retrieval model of sentence processing (Lewis & Vasishth, 2005) and Parker and Phillips' 2016 hypothesis (the presence of the verb blocks pre-verbal elements from retrieval interference), in the syntactic configuration where the main verb preceded the reflexive we replicated the main results of Experiment 2A. This means that word order alone cannot explain the absence of interference effects in many studies conducted in English: interference effects are still present in case the verb precedes the reflexive (Badecker and Straub, 2002; Clifton et al., 1999; Nicol and Swinney, 1989, Experiments 5, 6; Sturt, 2003, Experiment 2; Clackson et al., 2011; Dillon et al., 2013; King et al., 2012; Parker and Phillips, 2016; Xiang et al., 2009).

We once again found a correlation between participants' mean accuracy and reading times: in all the analyzed regions, more accurate participants read significantly more slowly. Interestingly, in Experiment 2A, we found this effect only in the pre-critical region and in the spillover after the reflexive. It is unclear why it was not present in other regions, since the accuracies in Experiments 2A and 2B are comparable.

As encoding interference does not depend on the word order, we expected to replicate the encoding interference effects (slower reading times) found in Experiment 2A on the main and relative clause verbs. We found a main effect of interference at the region following the relative clause verb, but not at the main verb. As the region following the main verb was the reflexive, it is impossible to disentangle spillover effects from processing of the reflexive itself. At any rate, the evidence for encoding interference is present in two regions (as compared to three in Experiment 2A): the region following the relative clause verb and the reflexive region.

At the reflexive region, the pattern of reading times is similar to the one observed in Experiment 2A: we again found a slowdown in the interference condition in gender-unmarked, but not gender-marked reflexives, but this time the interaction did not depend on participants' accuracy. The fact that the reading times pattern found in Experiment 2A was again replicated in Experiment 2B is an argument in favor of its systematic nature. However, the post-hoc explanation we provided for the effect in Experiment 2A does not fit Experiment 2B equally well: we reasoned that in gender-marked reflexives, the slowdown due to encoding interference is present, but concealed by a speedup caused by retrieval interference. However, the speedup in processing gender-marked reflexives is only predicted by cue-based retrieval as implemented in ACT-R (Lewis & Vasishth, 2005) if the distractor is particularly active. In Experiment 2B at the point of processing the reflexive the distractor must be less active than the antecedent because of the recent reactivation of the antecedent at the main verb. In such a case retrieval interference account predicts a slowdown at the reflexive region, not a speedup. Therefore, we should observe a slowdown in reading times at gender-marked reflexives when the gender of the distractor matches the gender of the antecedent. Our results contradict this prediction and therefore cannot be reconciled with the retrieval interference account.

If retrieval interference cannot account for the absence of interference effects in gender-marked reflexives, what can? One straightforward option is that gender-marked reflexives differ in some important way from the gender-unmarked reflexives. There is indeed a semantic difference: gender-marked reflexives put emphatic focus on the antecedent. As Lyutikova (1997) puts it, gender-marked reflexives (as opposed to gender-unmarked reflexives that take a purely syntactic function) signal that despite the expectations of a listener, the same person plays two different central roles in the situation (cf. “You did it to yourself”). It means that in our experimental conditions, gender-marked reflexives not only established coreference between the reflexive and the antecedent, but also provided higher-level discourse and/or semantic information, putting the emphatic focus on the antecedent.

Two additional facts may be seen as a post-hoc indirect support for the claim that gender-marked reflexives were processed differently. First, in Experiment 2A, there was a main effect of reflexive type two words downstream the reflexive: the word was read longer in conditions with gender-marked reflexives. Second, in Experiment 2B (but not 2A), question response accuracies were higher in conditions with gender-marked reflexives. These results might indicate that processing the emphatic focus on the antecedent took longer than establishing purely syntactic relationship, but the resulting interpretations were more stable, as demonstrated by question response accuracies. However, any post-hoc interpretation must remain a speculation until further tests.

Even though gender-marked reflexives might require some additional extra-syntactic processing, at present it is unclear why we did not find encoding interference effects in gender-marked reflexives. Every encoding interference account predicts the same effects regardless of gender marking, and if the slowdown in processing gender-unmarked reflexives is caused by encoding interference, there should be a similar slowdown in processing gender-marked reflexives. We suggest that in sentences with gender-marked reflexives, establishing emphatic focus at the point of retrieving the antecedent is assosiated with greater variance in processing times that conceals the main effect of interference.

An alternative explanation would be that the processing slowdown in marked reflexives might be concealed by a slowdown in the control condition: if on some proportion of trials participants erroneously predicted that the upcoming words should bear the gender marking of the distractor, encountering the gender marking consistent with the target should cause processing delays. No delays of such nature are expected either in the interference condition (since the prediction would always be confirmed), or in the gender-unmarked reflexives (since the prediction could never be disconfirmed). Only in marked reflexives slowdowns might arise in each condition and undermine the comparison between those.

To summarize, in Experiment 2B, we replicated the main results of Experiment 2A: the correlation between reading times and mean accuracies (more accurate participants read more slowly), the encoding interference effects in reading times at the relative clause verb and reflexive, and the unexpected pattern of reading times at reflexive (a slowdown in the interference condition in gender-unmarked, but not gender-marked reflexives). In Experiment 2B, the reading times at reflexive cannot be explained by the retrieval interference account, and the retrieval interference explanation of interference effects in processing reflexives in Russian is therefore ruled out.

2.4 General discussion and conclusions

The main goal of the present paper was to ascertain whether it is retrieval or encoding interference that accounts for the similarity-based interference effects in reflexive processing. The answer to this question would allow us, from the one side, to accept or reject the syntax as an early filter account of sentence processing, and from the other side, to obtain a more general insight into the functioning of working memory in online sentence processing.

In order to disentangle encoding and retrieval interference accounts' predictions, we conducted three experiments: one in German, contrasting reflexive and pronoun processing, and two in Russian, contrasting the processing of gender-marked and gender-unmarked reflexives. In the first experiment, we failed to find any interference effects, presumably due to the difficulty of the experimental materials. In the second experiment, we pitted the predictions of the encoding and the retrieval interference accounts against each other within reflexives: the encoding interference account predicts that both in gender-marked and gender-unmarked reflexives, the interference condition would be processed more slowly. On the contrary, the retrieval interference account predicts that only in the gender-marked reflexives would the difference between the interference and no interference conditions appear, since only in the processing of gender-marked reflexives gender can be used as a retrieval cue. In Experiment 2A, we encountered an unexpected pattern of reading times at the region following the reflexive – a slowdown in the interference condition in the gender-unmarked, but not in the gender-marked reflexives. This reading times pattern was replicated in Experiment 2B, where the order of reflexive and the main verb was reversed (as in English, the reflexive followed the verb). While the results of Experiment 2A might be reconciled with the retrieval interference account under certain conditions, the results of Experiment 2B contradict the predictions of the ACT-R model: when the reflexive is preceded by the verb whose subject is the reflexive's antecedent, retrieval interference effects are expected to lead to a slowdown, not a speedup in mean reading times. Since retrieval interference cannot account for the results of Experiment 2B, and the same pattern of reading times was found in Experiments 2A and 2B, we expect that the underlying cause was the same in both experiments, and therefore the retrieval interference explanation must be rejected. To summarize, we found no retrieval interference effects in the three experiments reported in this paper.

On the contrary, in the two experiments carried out in Russian, we found evidence in favor of encoding, but not retrieval, interference, both in reflexive-antecedent and in subject-verb dependencies. This stands in marked contrast to German, where no encoding interference in the processing of reflexives was found in two higher powered studies (Jäger, Benz, et al., 2015) and in the Experiment 1 reported in this paper. It does not seem likely that the existence of encoding interference depends on the language, rather our ability to detect interference effects might depend on the syntactic structure in question and the skill of the readers. As we already noted, the syntactic structure of the sentences used in Experiment 1 was more complicated than that of Experiments 2A and 2B (double vs. single embedding), which might have caused the observed difference across experiments.

At the same time, our results are not fully consistent with the predictions of the encoding interference account: while the slowdown at the reflexive is predicted for all sentences where the distractor matches the gender of the antecedent, we only found it in gender-unmarked, but not in gender-marked reflexives. We suggest that this might have two explanations. The first is that gender-marked reflexives require additional semantic processing: Lyutikova (1997) suggests that in Russian, gender-marked reflexives not only establish referential relationship between the reflexive and its antecedent, but also put emphatic focus on the antecedent. It is possible that additional semantic processing associated with establishing emphatic focus might conceal the encoding interference effect. The second explanation concerns a possible fault in the control condition: if in some proportion of trials participants erroneously expect the gender marking of the distractor on the upcoming words, their predictions can be disconfirmed only in the no interference condition in sentences with gender-marked reflexives. That would lead to delays in reading times, which could in turn undermine the comparison with the interference condition.

Interestingly, consistent evidence for encoding interference was found in the question response accuracies in all three experiments, including the experiment in German. The same pattern of results was also reported for German in Jäger, Benz, et al. (2015). Although this is not explicitly discussed, we assume that both retrieval interference accounts considered in this paper (Lewis & Vasishth, 2005; McElree et al., 2003) predict that in answering comprehension questions, the resulting representation that was built during sentence comprehension is used. Even if later reanalysis was postulated, it would engage the retrieval mechanisms specified in the models. In comprehension questions, the retrieval of verb arguments (required to provide a correct answer) would be initiated at the verb. In all the reported experiments, verbs were gender-unmarked, so gender could not be used as a retrieval cue, and retrieval interference account predicts equal accuracies across all conditions. This contradicts the pattern of observed accuracies. We must either suggest that mechanisms involved in sentence processing and building a faithful representation differ from those that provide access to the resulting representation (as in answering comprehension questions), or interpret comprehension question accuracies as evidence for encoding and against retrieval interference.

Finally, across the three experiments presented in this paper, the correlation between participants' accuracies and reading times seems to be robust: more accurate participants read more slowly. In Experiment 2A accuracy was crucial for uncovering the critical interaction between interference and reflexive type: the interaction was present only in the more accurate participants' reading times. However, no such relationship was found in Experiment 2B – the critical interaction was not modulated by participants' accuracy. Therefore, we replicated the relationship between reading speed at the retrieval site and comprehension accuracy reported by Nicenboim et al. (2015) only in one of the two experiments. Nevertheless, researchers who investigate long-distance dependencies might benefit from being aware of this relationship and in particular of the fact that reading times from the participants who do not build syntactic dependencies correctly might conceal the effect present in the reading times of the more accurate participants.

To conclude, in two out of three experiments reported in this paper we found a reading times pattern that is inconsistent with the retrieval interference account, but can be explained by encoding interference. Feature-matching distractors influence how coreference between the antecedent and the reflexive is established, and that goes against the strong version of the syntax as an early filer account (Nicol & Swinney, 1989). However, the main claim of the account – that reactivation of the antecedent is restricted by grammatical constraints – still holds true: encoding interference attributes the slowdown in processing the reflexive to feature overwriting and degraded memory representation of the antecedent, not to competition for retrieval between all the nouns.

Chapter 3

Agreement attraction and semantic attraction in ill-formed sentences

One way to understand how the human language processing system operates is to study the errors people make and the circumstances that affect these errors. One particularly well-studied type of errors is called agreement attraction (Bock & Miller, 1991; Kimball & Aissen, 1971). Agreement attraction refers to an erroneous agreement typically between the verb and a non-subject noun that seizes morphosyntactic control of the verb from the subject, as in:

(1) *The difference between the studies stem from ...

Here, the verb agrees with 'studies' – both are plural – instead of with the subject 'difference' which is singular. Even though the resulting sentence is clearly ungrammatical such sentences are regularly produced (Haskell & MacDonald, 2005) and often go unnoticed in comprehension (Clifton et al., 1999; Tanner & Bulkes, 2015).

Agreement attraction has been studied intensively in language production and more recently also in language comprehension. This research has identified various constraints on agreement attraction. For instance, agreement attraction has been found more reliably when the subject is singular, as in (1), than when it is plural (referred to as *singular-plural* asymmetry, see, for example, Bock and Cutting, 1992; Bock and Eberhard, 1993b; Bock and Miller, 1991; Deutsch and Dank, 2011;

Eberhard, 1997, but see Franck et al., 2002b, for a counter-example). While the position of the attractor seems to have some impact on the strength of agreement attraction (e.g., Franck et al., 2006; Franck et al., 2002b), there is currently little evidence suggesting that syntactic constraints can completely prevent a noun from interfering with the subject-verb dependency (but see Franck et al., 2010, who report some evidence for immunity to agreement attraction in complement clauses). Agreement attraction has also been demonstrated in a variety of languages other than English and there is some evidence that languages with richer morphosyntax, e.g., Russian and Spanish, may be more robust to agreement attraction (Foote and Bock, 2012; Lorimor et al., 2008, but see Lago et al., 2015). Finally, it has been found that patterns of agreement attraction errors in production largely mirror the effects in comprehension, which has raised the question whether the underlying mechanisms are the same (Pearlmutter et al., 1999a).

All models aiming to explain agreement attraction errors in production share the assumption that attraction is manifested only on the morphosyntactic level of language organization, that is, attraction is caused by mechanisms that can derail the formation of morphosyntactic relationships in a sentence (i.e. agreement) but not other aspects of a sentence (Bock et al., 2001; Eberhard et al., 2005; Franck et al., 2002b). According to these models, agreement attraction is a phenomenon with a rather narrow scope.

Meanwhile, more general language processing models have been used to explain agreement attraction errors in comprehension: the Lewis and Vasishth model (Lewis and Vasishth model, Engelmann et al., 2019; Lewis & Vasishth, 2005; Nicenboim & Vasishth, 2018) and the self-organized sentence processing model (self-organized sentence processing model, Smith et al., 2018; Tabor & Hutchins, 2004). According to these accounts, attraction errors arise from the particular way in which linguistic structure is stored in content-addressable memory. While these accounts have so far only been used to explain (morpho-)syntactic attraction effects, the principles they are based on are thought to be more domain-general. However, if we assume, as the Lewis and Vasishth model and self-organized sentence processing model do, that agreement attraction arises from domain-general mechanisms, there is no reason why attraction should be limited to the morphosyntactic level. Instead, we would expect that attraction effects should also arise in other linguistic domains, for instance, on the level of meaning.

To test this prediction, we ran three experiments in which participants where presented with a verb and a sentence fragment and after which they had to decide whether the verb was a viable continuation of the fragment or not. Specifically, we tested whether language users would accept mismatching verbs as sentence completions when there was a another noun (the attractor) that satisfied the verb's demand for a semantically matching subject. More concretely, we tested (among other things) whether the singular verb form 'cuts' would be accepted more often as a completion to fragments like (2-a) than to fragments like (2-b) even though 'cuts' thematically fits the subject equally badly in both sentences.

(2) a. The drawer with the knife ... (cuts?)
b. The drawer with the handle ... (cuts?)

If attraction is limited to morphosyntax, we expect no difference in completions for (2-a) and (2-b). However, if there were more errors in (2-a) than (2-b), this would constitute evidence that attraction is a more general phenomenon than has typically been assumed in the literature and this finding would therefore favor more unifying theories of sentence processing. Hence, this research asks not just questions about attraction phenomena in particular, but also promises new insights into the mental representation of linguistic structure and the modularity of linguistic processes.

In the following, we will briefly review the most influential accounts of agreement attraction in production and comprehension and then outline their predictions with regard to semantic attraction errors. Then, we will report three experiments and follow up with a computational simulation examining one model's predictions at a more fine-grained level.

3.0.1 Production accounts of agreement attraction

The *feature percolation* account (Franck et al., 2002b; Nicol et al., 1997; Vigliocco & Nicol, 1998) was formulated to explain attraction effects in the number domain and heavily relies on the notion of markedness. Singular is considered an unmarked member of the number opposition and just plural is assumed to be marked (e.g., Bock & Eberhard, 1993b; Eberhard, 1997; Harley & Ritter, 2002). The key idea is that in sentences like (2), where the attractor noun ('studies') is part of a complex subject noun phrase ('The difference between the studies'), the plural feature of the attractor can erroneously "percolate" up the syntactic tree and override the correct number marking of the noun phrase. As a result, the sentence processor expects a plural verb (e.g., 'stem') even though the subject ('The difference') requires singular. Thus, feature percolation posits that the culprit is faulty encoding of the subject noun phrase, not the agreement computation itself.

The beauty of this account is its parsimony and the fact that it makes rich predictions about the circumstances under which agreement attraction can arise. For instance, feature percolation correctly predicts more attraction errors when the subject noun is singular than when it is plural which was confirmed many times (Bock & Cutting, 1992; Bock & Eberhard, 1993b; Bock & Miller, 1991; Deutsch & Dank, 2011; Eberhard, 1997). Another strong prediction is that agreement attraction only arises in configurations where the attractor is embedded within the subject noun phrase, such as in (2) and (3). However, studies have also shown agreement attraction effects in constructions where the attractor is located outside the subject noun phrase ('The cabinets that the key … *open', Staub, 2009, 2010), questions ('*Are the helicopter for the flights safe?', Vigliocco and Nicol, 1998), and direct object constructions (Dutch subject-object-verb constructions: Hartsuiker et al., 2001; French object-subject-verb cleft constructions: Franck et al., 2006; German subject-object-verb constructions: von der Malsburg et al., 2020). All these findings pose problems for feature percolation.

(3) The soldier that the officers accused … *were

Further, feature percolation cannot explain why attraction errors increase when the subject is syntactically singular but denotes a set of items as in 'The label on the bottles …' (Foote & Bock, 2012; Hartsuiker et al., 1999; Vigliocco et al., 1995; Vigliocco et al., 1996) or 'The team with the red shirts … were' (Humphreys & Bock, 2005; Smith et al., 2018; Solomon & Pearlmutter, 2004).

An alternative account that can explain the latter class of cases is the *marking and* *morphing* account (Bock et al., 2001; Eberhard et al., 2005). Like feature percolation, it assumes faulty encoding of the subject, but unlike feature percolation, it relies on the concept of *notional number* — a semantic representation of the entity that is referred to, either as a multitude or as a single unit. Both nouns, such as 'team', and noun phrases, such as 'the picture on the postcards', can be notionally plural while being syntactically singular. The marking and morphing account builds upon feature percolation and postulates that the subject's notional number influences the computation of number agreement over and above the morphosyntactic number match between the attractor and the verb. Essentially, the more multitude-like the abstract representation of the subject, the higher the probability of using a plural verb. Just as feature percolation, the account is well-suited for explaining agreement attraction effects in the number domain. And like feature percolation, it only covers the configurations when attractor is located within the subject noun phrase (although, unlike feature percolation, it could potentially be extended to cover object attraction; Eberhard et al., 2005).

A shortcoming of feature percolation is that it fails to account for instances of agreement attraction involving case or gender features (Antón-Méndez et al., 2002; Badecker & Kuminiak, 2007; Bader & Meng, 1999; Slioussar & Malko, 2016; Slioussar et al., 2015). The specification of marking and morphing allows the model to account for gender attraction in systems with two genders, however, it is unclear how it can be extended to systems with more than two features, such as many gender or case systems.

While both feature percolation and marking and morphing were designed to explain attraction errors in production, they were also invoked to explain analogous effects in sentence comprehension (Pearlmutter et al., 1999a; Wagers et al., 2009):

(4) a. *The key to the cells were ...
    b. *The key to the cell were ...

In sentences with attraction errors, such as (4-a), reading times at the verb were shown to be faster than in control sentences, (4-b), where the number marking of the attractor noun does not match the verb (see also Avetisyan et al., 2020; Lago et al., 2015; Tucker et al., 2015; Villata et al., 2018, for similar results in Spanish, Eastern

Armenian, Arabic, and Italian). In addition, sentences with attraction errors are more often judged as grammatical or acceptable than analogous sentences without an attractor noun matching the verb (Hammerly et al., 2019; Patson & Husband, 2016; Vasishth et al., 2017; Wagers et al., 2009). Feature percolation and marking and morphing both can explain these effects by assuming that the plural feature of the attractor in (4-a) sometimes compromises the subject's number marking, in which case the unlicensed plural verb is actually expected and consequently doesn't cause as much processing difficulty as in the control condition (4-b).

Another prediction of both accounts is that the occasional misspecifications of the subject's number should cause processing difficult when the verb in fact agrees with the subject. In this scenario, the misspecification leads the parser to predict a different number marking on the verb and, upon encountering the (correct) verb stumbles, which should be reflected in a slowdown at the verb. The result is an *illusion of ungrammaticality*. Although some studies found evidence for such a slowdown (Franck et al., 2015; Lago et al., 2015; Nicol et al., 1997; Patson & Husband, 2016; Pearlmutter et al., 1999b; Wagers et al., 2009), most of them had design shortcomings, and the majority of studies did not find evidence for an illusion of ungrammaticality (inter alia, Cunnings & Sturt, 2018; Lago et al., 2015; Nicenboim et al., 2018; Patson & Husband, 2016; Thornton & MacDonald, 2003; Tucker et al., 2015; Wagers et al., 2009). The lack of support for the illusion of ungrammaticality lead researches to believe that production-based models might not adequately explain comprehension (but note that this position has recently been challenged by Hammerly et al., 2019).

3.0.2 Comprehension theories of agreement attraction

We will now briefly review two general models of language comprehension that can potentially explain attraction effects in comprehension even though they were not explicitly designed for this purpose. The *Lewis and Vasishth 2005 model* (henceforth Lewis and Vasishth model05, Lewis & Vasishth, 2005) is based on the content-addressable memory architecture ACT-R (J. R. Anderson, 1996). The model assumes that syntactic chunks are activated in working memory when they are encountered and later retrieved in order to build syntactic dependencies. Syntactic chunks (including words) are represented as bundles of features and are retrieved by querying a subset of these features relevant at the moment of retrieval. The model was first applied to the comprehension of sentences with agreement attraction errors by Wagers et al. (2009). To understand how it can explain agreement attraction, consider the grammatical sentence (5) from their study:

(5) The cabinets that the key opens ...

When encountering the verb ‘opens’, the parser triggers a retrieval of the previously processed subject to complete the subject-verb dependency. The verb is marked for number, and the parser will therefore spread activation to every word that has the features `+SUBJECT`[1] and `+SINGULAR`. The word with the highest activation (that also exceeds a so-called *retrieval activation threshold*) will be retrieved and used to complete the dependency. In (5), the only word that fully matches the retrieval cues is the subject ‘key’ and it will therefore be retrieved in virtually all cases.

Now consider the sentence (4-a) which contains an agreement attraction error. The parser will spread activation to every word that has features `+SUBJECT` and `+PLURAL`. Now, both the subject ‘key’ and the attractor ‘cells’ fail to fully match the retrieval cues, each has only one matching feature, either `+SUBJECT` or `+PLURAL`. The attractor and the subject therefore receive an equal amount of activation and noise in the system will determine which word will be retrieved. As a result, the attractor will be retrieved in half of the cases.

An important difference to the production accounts discussed above is that the Lewis and Vasishth model predicts attraction effects in ungrammatical sentences irrespective of their syntactic structure — any noun, irrespective of position can be misretrieved instead of the subject as long as it matches sufficiently many features required by the verb. The Lewis and Vasishth model can also explain the increase in attraction error rates when the attractor superficially resembles the sentential subject (Engelmann et al., 2019), for instance, when the attractor’s case marking is ambiguous between nominative and the actual case (Badecker & Kuminiak, 2007; Hartsuiker et al., 2003; Slioussar & Malko, 2016). A further difference to the production accounts is that in grammatical sentences, the Lewis and Vasishth model predicts the opposite

[1] `+SUBJECT` feature is a commonly used simplification, adopted in Jäger et al., 2017b, and elsewhere.

of the illusion of ungrammaticality. Nicenboim et al. (2018) found some inconclusive evidence in favor of this effect, but the majority of studies found no effect.

Unlike production accounts, the Lewis and Vasishth model cannot explain is the singular-plural asymmetry present in many studies. The reason is that, unlike feature percolation, the Lewis and Vasishth model assumes that singular is marked just as plural.[2] Similar asymmetries that the Lewis and Vasishth model cannot explain out of the box have been found in gender (more errors in sentences with a masculine subject noun and feminine attractor that the other way around in Slovak and Russian, see Badecker and Kuminiak, 2007; Slioussar and Malko, 2016).

Another general model of sentence comprehension than can potentially explain agreement attraction is the *self-organized sentence processing model* (henceforth SOSP, Smith et al., 2018; Tabor & Hutchins, 2004; Vosse & Kempen, 2000). This account assumes that every word tries to form a connection with every other encountered word, and that such connections – treelets – combine further in a bottom-up fashion to form larger meaning-bearing structures. The strength of connections between these treelets depends on the goodness of fit which is assessed based on all features, morphosyntactic, semantic, and otherwise. Strong connections grow stronger over time and weak connections taper off. While the underlying dynamics in this model look rather different from those assumed in Lewis and Vasishth model, many predictions are similar. Crucially, if two connections have approximately equal strength, as the verb-subject and the verb-attractor connections in (4-a), the winning attachment depends largely on noise in the system. Hence, self-organized sentence processing model, just like the Lewis and Vasishth model, predicts attraction effects in a wide range of syntactic configurations. Moreover, self-organized sentence processing model also covers notional plurality effects (Smith et al., 2018), but it is less clear how it could explain attraction asymmetries in number and gender.

Both the Lewis and Vasishth model and self-organized sentence processing model have also been invoked to account for attraction effects also in production (Badecker & Kuminiak, 2007; Konieczny et al., 2004; Smith et al., 2018). The idea being that, to build syntactic structure for production, we need to keep in memory what has already been said and what we are planning to say, and that the memory substrate

[2]Note, though, that some comprehension studies did not find evidence for the singular-plural asymmetry, e.g., Häussler (2009) and Acuña-Fariña et al. (2014).

used in this process is likely the same as for comprehension.

3.0.3 Differences between production and comprehension accounts

Production and comprehension accounts differ not only in the mode of language use, they also postulate different mechanisms underlying attraction effects. According to the Lewis and Vasishth model and self-organized sentence processing model, attraction occurs during dependency formation – an incorrect syntactic chunk can be retrieved to form the dependency, whereas according to both feature percolation and marking and morphing attraction is caused during the encoding of the subject's number. As a consequence, the Lewis and Vasishth model and self-organized sentence processing model predict that if attraction occurs, the attractor noun will be perceived to be the subject, while both production accounts predict that the subject noun will be identified correctly, only the number marking of the whole noun phrase will be incorrect. Available evidence (Schlueter et al., 2019) favors the production accounts by demonstrating that attractor noun is perceived to be the subject only in a minority of attraction cases.

Further, the two classes of accounts differ in the role they assign to semantic information. While marking and morphing allows some semantic properties either of the subject noun, such as conceptual plurality, or of the whole noun phrase, such as distributivity, to affect feature computation and assignment, other semantic properties are not assumed to have an impact on attraction. For example, the model cannot account for the increase in attraction rates due to the goodness of thematic fit between the attractor and the verb (Thornton & MacDonald, 2003), or due to higher semantic integration between the subject and the attractor within the noun phrase (Solomon & Pearlmutter, 2004), or due to attractor being an animate noun (Bock & Miller, 1991, Experiment 3). Many of these semantic influences on agreement computation are easily explained by the comprehension accounts since in both the Lewis and Vasishth model and self-organized sentence processing model semantic features receive the same treatment as all other types of features, including morphosyntactic.

Crucially, since semantic features are being treated in the same way as morphosyntactic features their role is not limited to influencing agreement computations (and therefore modulating agreement attraction); both the Lewis and Vasishth model and self-organized sentence processing model make the surprising prediction that attraction effects should occur in other linguistic domains as well, independently of agreement. As Lewis and Vasishth state (2005, p. 411):

> In this model, we have realized only syntactic cues, which are used primarily to reactivate predicted structure to unify with. However, the model can accommodate a richer set of cues—for example, there may also be semantic cues derived from specific lexical constraints (e.g., the semantic constraints that a verb places on its subject).

Similarly, Smith et al. state (2018, p. 24):

> In SOSP, linguistic tree-representations form via continuous feedback interactions among treelets that are guided by vectors of syntactic and semantic features.

This means that both the Lewis and Vasishth model and self-organized sentence processing model predict not only agreement attraction errors, but also analogous *semantic attraction errors*. These could be reflected in the acceptance of a verb that thematically fits the attractor but not the subject noun. For example, in the sentence ‘The drawer with the knife cuts . . . ’ the semantic attractor ‘knife’ satisfies the semantic restrictions set by the verb better than the subject noun ‘drawer’. In comprehension, the verb ‘cuts’ should therefore be easier to process in the presence of the attractor that can perform the cutting action than in the presence of a noun that cannot, such as ‘handle’ in ‘The drawer with the handle cuts . . . ’. These effects would precisely mirror agreement attraction effects, but crucially, they could arise independently of morphosyntactic processing — note that the example of semantic attraction os morphosyntactically well-formed.

Note that this proposal differs from the one made in Thornton and MacDonald (2003) where semantic features were shown to influence agreement computations. While highly relevant in the present context, the Thornton and MacDonald proposal is more narrow in scope than the idea of purely semantic attraction effects predicted by the Lewis and Vasishth model and self-organized sentence processing model. Thus finding evidence for purely semantic attraction effects would considerably widen the scope of the attraction phenomenon, increase its relevance, and improve our understanding of it.

There is one study that provides evidence for semantic attraction. In two eye-tracking experiments, Cunnings and Sturt (2018) tested sentences like (6):

(6) a. Sue remembered the letter that the butler with the *cup* accidentally shattered.
b. Sue remembered the letter that the butler with the *tie* accidentally shattered.

Both sentences are implausible, but Cunnings and Sturt found that the verb 'shattered' was processed faster in condition (6-a), where the local non-subject noun 'cup' was semantically a good fit for the verb 'shattered', than in (6-b), where the local noun was 'tie', i.e. an object that cannot be shattered. These results mirror agreement attraction effects in comprehension, but it is not clear how they compare to attraction effects. Both Lewis and Vasishth model and self-organized sentence processing model predict the same proportion of misinterpretations and the same processing times profiles for semantic and morphosyntactic attraction from a non-subject noun. However, Cunnings and Sturt (2018) only tested semantic, but not morphosyntactic attraction in their study, so that effect sizes could not be compared.

The purpose of the present study is, first, to conceptually replicate the semantic attraction effect demonstrated by Cunnings and Sturt (2018), and second, to build on their work by examining more closely whether the effects they observed constitute genuine semantic attraction effects arising from the same mechanisms underlying agreement attraction. To do so, we compared configurations with semantic attraction and morphosyntactic attraction side by side using a slightly modified version of the forced-choice paradigm that has been used extensively to study agreement attraction. The goal of Experiment 1 was to establish whether semantic attraction errors occur in this paradigm and, if yes, whether their rate is comparable to that of morphosyntactic (agreement) attraction errors as is predicted by Lewis and Vasishth model and self-organized sentence processing model. Experiment 2 replicated the findings of Experiment 1 and included two additional conditions that give us further insight into how semantic and morphosyntactic attraction interact. Experiment 3 mitigated a possible confound in the item design, and replicated the results using the same experimental conditions, but a new set of experimental items. Finally, we report simulations with a modified version of the Lewis and Vasishth model to see how the model could potentially account for the results of the three experiments.

3.0.4 Disclosures

All reported studies had been carried out in accordance with the Declaration of Helsinki. All participants provided informed consent. The full list of materials used in both reported experiments, the collected data and analysis code are available from the project page at the Open Science Framework, doi: doi:10.17605/OSF.IO/P9HS7. The full list of materials is also provided in Appendix 6.1.

3.1 Experiment 1

To demonstrate attraction effects in the semantic domain and to compare them to classical agreement attraction, we used a forced choice task. The classical version of the task presents participants with a sentence preamble and prompts them to choose one out of two verbs as a plausible continuation. Instead of two verbs, we showed only one and asked participants to judge whether or not it was a plausible continuation of the preamble. If the rate of mistakes is increased when the attractor matches the verb thematically, that would constitute an attraction effects in the semantic domain and thus suggest that attraction is not limited to agreement processing.

A secondary goal was to compare semantic attraction effects to morphosyntactic attraction effects: Are they equally sized? And how do semantic and morphosyntactic attraction interact? Are the effects of morphosyntactic and semantic attraction additive, or under- or super-additive? If there was evidence for an interaction, that would favor a single underlying mechanism (Roberts & Sternberg, 1993; Sternberg, 1998), while a lack of interaction would be consistent both with a single common and independent underlying mechanisms.

3.1.1 Methods

Participants

Participants (N=1,100) were recruited on Prolific, a crowd-sourcing platform for academic studies. Participants were prescreened for being self-reported native speakers of English who were born in the US/UK, citizens of the US/UK and residents of the US/UK at the time of participation. Participation took approximately 1 minute and was compensated with 10p (0.1 GBP). After finishing the experimental task, participants had to indicate (again) whether they are native speakers of English, US/UK citizens, and that they spent the first five years of their life in the US/UK. After excluding data from participants who answered negatively to at least one of these questions, 1,072 individuals were left in the analysis.

Materials

We tested twenty-five item sets (see Table 3.1) in which the verb never fully matched the subject. It mismatched either the subject's number (morphosyntactic violation), or meaning (semantic violation), or both (double violation). At the same time, the verb could mismatch or match the attractor in number (morphosyntactic attraction), meaning (semantic attraction), or both (double attraction). This set of conditions allowed us to test morphosyntactic attraction (conditions b vs. a), semantic attraction (d vs. c), as well as double attraction (f vs. e).

The items had the following structure: The subject noun was followed by a prepositional phrase containing the attractor. The verb had clear thematic restrictions that allowed for only a subset of nouns to plausibly serve as subject. Subject- and attractor-verb combinations were created with the aim to avoid metonymic and metaphorical sense transfers (e.g., a person glowing with joy).

Procedure

The study was conducted as a single-trial online experiment where each participant saw only one item and only one of the experimental conditions. This way we avoided adaptation of processing strategies to the stimuli, in particular, to the ungrammatical or otherwise non-well-formed sentences. Arguably, this also allowed us to detect the

Condition			Violation	Attraction
a.	The drawer with the handle	OPEN	morphosyntactic	none
b.	The drawer with the handles	OPEN	morphosyntactic	morphosyntactic
c.	The drawer with the handle	CUTS	semantic	none
d.	The drawer with the knife	CUTS	semantic	semantic
e.	The drawer with the handle	CUT	double	none
f.	The drawer with the knives	CUT	double	double
g.	The drawer with the knife	CUT	double	semantic
h.	The drawer with the handles	CUT	double	morphosyntactic

Table 3.1: Example experimental item. Conditions (a-f) were tested in Experiment 1, conditions (g) and (h) were added in Experiment 2. 'Double' stands for simultaneous morphosyntactic and semantic attraction and/or violation.

biggest possible attraction effect as compared to the same number of probes tested with a smaller number of participants in a repeated-measures design, as participants likely become aware of the nature of the mismatches and hence more efficient at detecting them (Baayen et al., 2017; Demberg & Sayeed, 2016; Fine et al., 2013).

The experiment consisted of instructions, the experimental probe, and the debriefing questions mentioned above (native language, citizenship, country of residence during first five years of life). Within the experimental task, participants were presented with a verb in capitals (see Table 3.1). After memorizing the verb, they had to press the spacebar key to see a sentence fragment and the two response buttons below it. They had to read the fragment and indicate via a mouse click whether the memorized word was an acceptable continuation of that fragment (we did not explicitly state that the word was a verb). Thornton and MacDonald (2003) showed that presenting the verb before the preamble produced the same results as the more common version of the oral production task where the verb is presented following the preamble.

To indicate whether the verb was a possible continuation of the sentence, participants had either to click on one of the symbols (green check mark or red X mark) or to press 1 or 2 on the keyboard, where 1 corresponded to 'good fit' and 2 to 'bad fit'. Note that the verb never perfectly matched the subject and the correct response was therefore always to reject the verb. However, since each participant performed only one trial, the correct response could not be guessed based on knowledge from prior trials.

The experiment was programmed using the Ibex[3] software and run on the IbexFarm cloud service.

Data analysis

All analyses were conducted with the R system for statistical computing (R Development Core Team, 2009). Data were analyzed using generalized linear mixed models fit in the Bayesian framework (Vasishth et al., 2018) using the 'brms' package (Bürkner et al., 2017). Plots were produced with the 'ggplot2' and 'tidybayes' packages (Kay, 2019; Wickham, 2016). Inferences were based on the posterior distributions of the parameters, which are reported in terms of the posterior mode and 95% percentile intervals (CrI). If nearly all of the posterior mass for an estimate fell on one side of zero, we considered that as evidence that the effect was reliable. However, note that we do not adopt a strict threshold here, we instead evaluate the strength of evidence in a graded fashion.

Accuracy in the judgment tasks was modeled using hierarchical logistic regression. Treatment contrasts were used to code the two factors: the type of violation (morphosyntactic, semantic, or both) with morphosyntactic violations serving as the reference level, and attraction (none, morphosyntactic, semantic) with no attraction being the reference level (Schad et al., 2020). We estimated both simple effects as well as the interaction between them. For modeling accuracy, we used regularizing priors for the main effects and interactions (Normal(0, 1)). The model also included full by-item random effects (Barr et al., 2013). Random effects for participants were not needed since each participant contributed only one measurement (single-trial design).

We also analyzed reaction times, but since these were not of primary interest, results are reported in Appendix 6.1.4.

3.1.2 Results

The estimated proportions of correct responses in each condition are shown in Figure 4.1A and the posterior distributions of the parameters in Figure 4.1B.

Accuracy in condition (a), the baseline for morphosyntactic attraction, was 77%

[3]http://spellout.net/ibexfarm

($\hat{\beta} = 1.22$, 95%-CrI: $[0.86, 1.60]$). Accuracy in the baseline for semantic attraction (c) did not differ from the baseline for morphosyntactic attraction (a) (77% vs. 73%, $\hat{\beta} = 0.25$, 95%-CrI: $[-1.13, 0.64]$, $P(\beta < 0) = 0.73$). However, accuracy in condition (e), the baseline for double attraction, was higher than that in (a) (77% vs. 89%, $\hat{\beta} = 0.85$, 95%-CrI: $[0.21, 1.57]$, $P(\beta < 0) = 0.004$), which suggests that double subject-verb fit violations were easier to spot than isolated morphosyntactic or semantic violations.

We found the classic agreement attraction effect, i.e. accuracy was considerably lower in condition (b) with morphosyntactic attraction compared to baseline (a) without attraction (77% vs. 56%, $\hat{\beta} = -1.00$, 95%-CrI: $[-1.50, -0.49]$, $P(\beta < 0) = 0.999$). Neither semantic nor double attraction effects differed from the morphosyntactic attraction effect (semantic attraction: 49% vs. 45%, $\hat{\beta} = -0.17$, 95%-CrI: $[-1.07, 0.68]$, $P(\beta < 0) = 0.65$; double attraction: 75% vs. 78%, $\hat{\beta} = 0.17$, 95%-CrI: $[-0.75, 1.12]$, $P(\beta < 0) = 0.35$).[4]

To assess more directly whether semantic attraction also decreased response accuracy, we combined the posterior of the morphosyntactic attraction effect with the posterior of the difference between the morphosyntactic and semantic attraction effects (McElreath, 2016). The resulting posterior for the size of the semantic attraction effect (comparison between conditions c and d) suggested a great decrease in response accuracy in the presence of semantic attraction (73% vs. 45%, $\hat{\beta} = -1.17$, 95%-CrI: $[-1.96, -0.47]$, $P(\beta < 0) = 0.999$).

These effect sizes are slightly bigger but largely in line with those reported in earlier research using similar tasks: 17% in (Schlueter et al., 2019), 18% in (Staub, 2009), 13% and 19% in the sentence repetition paradigm used by (Thornton & MacDonald, 2003).

3.1.3 Discussion

In line with the predictions of the Lewis and Vasishth model and self-organized sentence processing model, we found a semantic attraction effect similar in manifestation and size to the classic morphosyntactic attraction effect. We will review the broader

[4]The percentage values for the last two effects indicate the expected accuracy if semantic and double attraction had the same magnitude as morphosyntactic attraction vs. the observed accuracy in the respective attraction conditions (d) and (f).

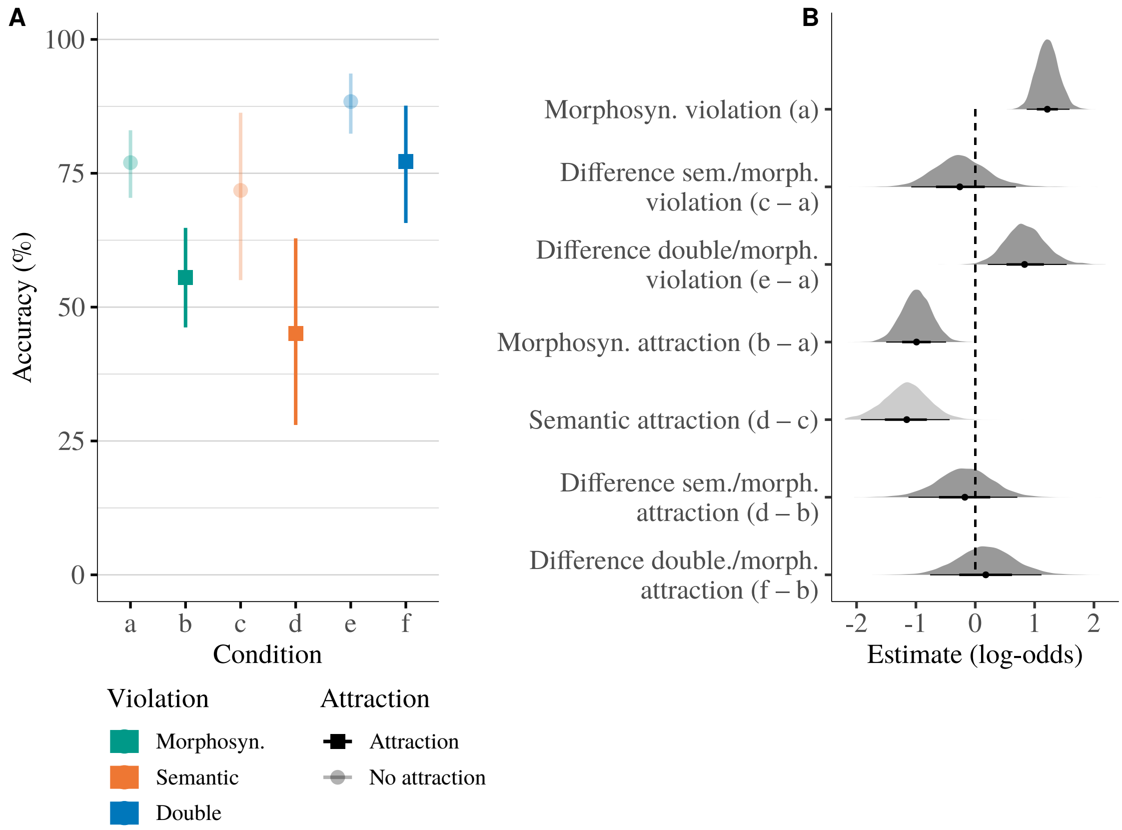

Figure 3.1: Results of Experiment 1. Panel A: Estimated condition means with 95% credible intervals. Panel B: Posterior distributions for the model parameters (log-odds scale). The posterior for the semantic attraction effect (light gray) was obtained by combining the posteriors for morphosyntactic attraction and the posterior for the difference between the morphosyntactic and semantic attraction. Error bars around the estimates represent 66% (thick) and 95% (thin) credible intervals.

implications of this finding in the general discussion. Our second goal was to assess whether these two types of attraction effects interact: under- or over-additive effects would favor a common underlying mechanism. While our analysis was suggestive of an interaction — the effect of double attraction was not larger than morphosyntactic and semantic attraction (in log odds) — the relevant comparison of conditions may have been flawed: Isolated morphosyntactic and semantic attraction effects were tested with subject-verb combinations that violated either morphosyntactic agreement or semantic plausibility. In contrast, double attraction was tested with subject-verb combinations that mismatched along both dimensions, morphosyntax and semantic plausibility. The results show that this double violation was easier to spot than single violations (higher accuracy in condition e than in a and c). So, while double attraction might be stronger than single attraction, that effect may have been partly counteracted and canceled out by the easier detection of the subject-verb mismatch in (e).

To address this shortcoming of the design, we conducted Experiment 2 with two additional conditions. Both contained double subject-verb fit violations combined with attraction along a single dimension, either morphosyntactic or semantic. Therefore, Experiment 2 allows us to cleanly compare morphosyntactic, semantic, and double attraction in the presence of the same double violation. A secondary goal of Experiment 2 was to replicate the semantic attraction effect found in Experiment 1.

3.2 Experiment 2

We retained all conditions from Experiment 1 and included conditions (g) and (h) that introduce morphosyntactic and semantic attraction manipulation in the presence of a double violation of subject-verb fit (see Table 3.1).

3.2.1 Methods

Participants

Participant recruitment procedure and exclusion criteria were the same as for Experiment 1. Individuals who participated in Experiment 1 were blocked from participating in Experiment 2. We tested more participants in order to maintain the same number of observations per condition as in Experiment 1 and thus the same statistical power: 1,450 individuals took part in the experiment; after applying exclusion criteria, data from 1,426 individuals were left in the analysis.

Materials

The same materials as in Experiment 1 were used, with the addition of two conditions, (g) and (h), see Table 3.1.

Procedure

Experimental procedure was identical to that of Experiment 1.

Data analysis

To establish the reliability of the semantic attraction effect, we replicated the analysis from Experiment 1 but excluded conditions (e) and (f), since comparisons with these conditions are flawed as explained above. This left us with a 2×2 design with factors *type of violation* (morphosyntactic or semantic) and *attraction* (present or not). As in Experiment 1, these factors were coded as treatment contrasts with morphosyntactic violation as the reference level for factor *type of violation* and no attraction as the reference level for the factor *attraction*.

To assess the interaction of morphosyntactic and semantic attraction in conditions (e)–(h), we fit a separate model with factors *morphosyntactic attraction*, *semantic attraction*, and their interaction. Morphosyntactic and semantic attraction were coded with sum contrasts such that the parameter estimates captured the main effects of morphosyntactic and semantic attraction (i.e. the effect averaged across the levels of the respective other factor). As before, the models included full by-item random effects.

3.2.2 Results

The estimated proportions of correct responses in each condition can be seen in Figure 4.2A and posterior distributions of the parameters in Figure 4.2B and Figure 4.2C.

Analysis replicating results of Experiment 1 (conditions a–d). Accuracy in the baseline condition for morphosyntactic attraction (a) was 76% ($\hat{\beta} = 1.18$, 95%-CrI: $[0.69, 1.72]$). Accuracy in the baseline condition for semantic attraction (c) did not differ from the baseline for morphosyntactic attraction (76% vs. 67%, $\hat{\beta} = -0.45$, 95%-CrI: $[-1.37, 0.48]$, $P(\beta < 0) = 0.83$). The morphosyntactic attraction effect (a vs. b) was in the expected direction but not reliable this time (76% vs. 70%, $\hat{\beta} = -0.36$, 95%-CrI: $[-0.88, 0.17]$, $P(\beta < 0) = 0.90$). The effect of semantic attraction was numerically bigger but did not differ from the effect of morphosyntactic attraction (59% vs. 43%, $\hat{\beta} = -0.65$, 95%-CrI: $[-1.63, 0.27]$,

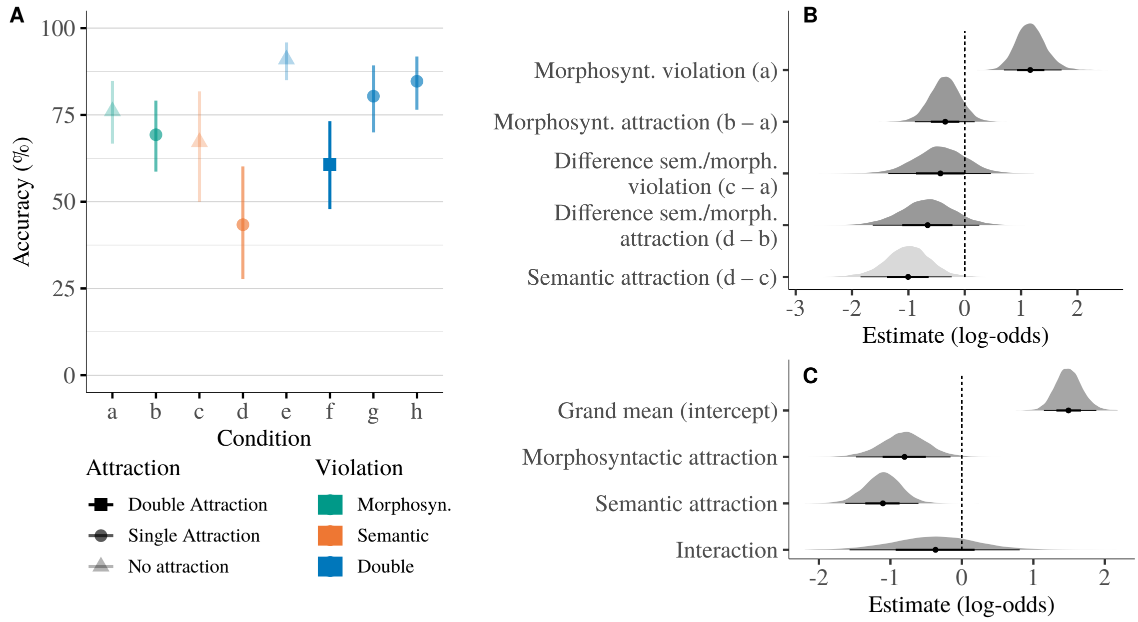

Figure 3.2: Results of Experiment 2. Panel A: Estimated condition means with 95% credible intervals. Panels B: Posterior distributions for the model of conditions (a)-(d). The posterior for semantic attraction (light gray) was obtained by combining the posteriors for morphosyntactic attraction and the difference between the semantic and morphosyntactic attraction. Panel C: Posterior distributions for the model of conditions (e)-(h). All parameters are on the log-odds scale. Error bars around estimates represent 66% (thick) and 95% (thin) credible intervals.

$P(\beta > 0) = 0.93$).[5] As in Experiment 1, we combined posteriors to get a direct estimate of the semantic attraction effect (d vs. c). The result shows that semantic attraction greatly decreased response accuracy (67% vs. 43%, $\hat{\beta} = -1.01$, 95%-CrI: $[-1.83, -0.24]$, $P(\beta < 0) = 0.993$) thus replicating the semantic attraction effect found in Experiment 1.

Analysis testing the interaction of morphosyntactic and semantic attraction (conditions e–h). The average accuracy across conditions was 82% ($\hat{\beta} = 1.5$, 95%-CrI: $[1.1, 1.9]$). Morphosyntactic attraction decreased response accuracy (87% vs. 75%, $\hat{\beta} = -0.8$, 95%-CrI: $[-1.5, -0.17]$, $P(\beta < 0) = 0.99$). Likewise semantic attraction decreased accuracy (89% vs. 72%, $\hat{\beta} = -1.1$, 95%-CrI: $[-1.6, -0.63]$, $P(\beta < 0) = 0.999$). There was no interaction of morphosyntactic and semantic attraction, i.e. their effects were approximately additive (83% vs. 80%, $\hat{\beta} = -0.37$, 95%-CrI: $[-1.6, 0.82]$, $P(\beta < 0) = 0.74$).

[5]The percentage values for the last effect indicate the expected accuracy if semantic attraction had the same magnitude as morphosyntactic attraction vs. the observed accuracy in the semantic attraction condition (d).

3.2.3 Discussion

The two goals of Experiment 2 were to test whether the semantic and morphosyntactic attraction effects are additive given appropriate control conditions, and to confirm the reliability of the semantic attraction effect. With regard to the first goal, the outcomes of Experiment 2 and of the analysis of the pooled data set suggest effect additivity, which is consistent with a single common substrate but also with separate substrates for morphosyntactic and semantic attraction. With regard to the second goal, we successfully replicated the semantic attraction effect, both in the context of single and double subject-verb fit violations.

While these results are in line with the key predictions of the Lewis and Vasishth model and self-organized sentence processing model, a reviewer pointed out a potentially critical confound in the design of our experimental items that could account for the semantic attraction effect: in semantic attraction conditions with single subject-verb fit violations (d), the attractor and the verb could sometimes form locally coherent noun-noun compounds, such as 'tree blossoms', 'knife cuts', 'fountain bubbles', and so on. Thus, it is possible that participants accepted the continuation in (d) not due to semantic attraction but because they adopted a noun-noun compound interpretation. This is possible in particular because we did not instruct participants to interpret the continuation word as a verb. To assess how much this design confound influenced our estimate of semantic attraction effect, and to get an unbiased estimate of semantic attraction, we replicated Experiment 2 using a new set of items that does not allow the noun-noun compound interpretation.

3.3 Experiment 3

3.3.1 Methods

Participants

Participant recruitment procedure was the same as for Experiments 1 and 2. Individuals who participated in Experiments 1 and 2, as well as individuals who participated in the pretest of experimental materials, were blocked from participating in Experiment 3. We tested 2,600 participants; after applying exclusion criteria, data from 2,454 participants were left in the analysis (compare to the pooled N=2,498 in Experiments 1 and 2).

Materials

We created a new set of experimental items. To exclude the possibility of forming a noun-noun compound interpretation, the attractor noun was followed by an adverb unambiguously signaling that the memorized word must be a verb, see example item in Table 3.2.

Table 3.2: Example experimental item from Experiment 3. 'Double' stands for simultaneous morphosyntactic and semantic attraction and/or violation.

Condition			**Violation**	**Attraction**
a.	The newsstand near the bench usually	SELL	morphosyntactic	none
b.	The newsstand near the benches usually	SELL	morphosyntactic	morphosyntactic
c.	The newsstand near the bench usually	SMELLS	semantic	none
d.	The newsstand near the coffee shop usually	SMELLS	semantic	semantic
e.	The newsstand near the bench usually	SMELL	double	none
f.	The newsstand near the coffee shops usually	SMELL	double	double
g.	The newsstand near the coffee shop usually	SMELL	double	semantic
h.	The newsstand near the benches usually	SMELL	double	morphosyntactic

To ensure that the semantic match/mismatch was actually perceived as such by native English speakers, we conducted a plausibility norming pretest. For each of the 32 items sets, we constructed five sentence preambles using the three NPs (the subject and the two attractors) and the two verbs as follows: 'The newsstand/the bench usually sells ...', 'The newsstand/the bench/the coffee shop usually smells ...'. Participants then rated these preambles on a 1-7 Likert scale. Every participant (N=50, recruited online on Prolific) saw all 32 experimental items, each item in one out of five conditions. Lists were created following Latin square design. We analyzed the results using Bayesian ordinal regression (for details, see Appendix 6.1.2). As expected, the preambles constructed to be plausible received systematically higher ratings than the ones constructed to be implausible. Based on model estimates, we excluded four items for which the estimated difference between the plausible and implausible conditions was the smallest and not reliably different from zero. We additionally excluded one item for which a distributive interpretation of the number attraction condition was available. This left us with 27 experimental items for Experiment 3. See Appendix 6.1 for the list of items.

Procedure

The experimental procedure was similar to that of Experiments 1 and 2 with a small modification: we introduced two training sentences so that participants could familiarize themselves with the experimental procedure. As for the experimental sentence, participants had to memorize a word and judge whether the word fit the sentence preamble. One of the training sentences was ill-formed ('The house by the new FURIOUSLY ...'), and we excluded data from participants who failed to notice the ill-formedness. This lead to exclusion of 5% of data points, but the results remain the same if data from these participants is retained.

Data analysis

We replicated both analyses of Experiment 2.

3.3.2 Results

The estimated proportions of correct responses in each condition can be seen in Figure 4.3A and posterior distributions of the parameters in Figure 4.3B and Figure 4.3C.

Conditions a–d. Accuracy in the baseline condition for morphosyntactic attraction (a) was 67%, $\hat{\beta} = 0.72$, 95%-CrI: $[0.28, 1.2]$). Accuracy in the baseline condition for semantic attraction (c) did not differ from the baseline for morphosyntactic attraction (a) (67% vs. 71%, $\hat{\beta} = 0.18$, 95%-CrI: $[-0.81, 1.2]$, $P(\beta < 0) = 0.37$). Morphosyntactic attraction (a vs. b) decreased response accuracy (67% vs. 38%, $\hat{\beta} = -1.2$, 95%-CrI: $[-1.6, -0.81]$, $P(\beta < 0) = 0.999$). The effect of semantic attraction was numerically bigger but did not differ from the effect of morphosyntactic attraction (42% vs. 38%, $\hat{\beta} = -0.17$, 95%-CrI: $[-0.9, 0.55]$, $P(\beta < 0) = 0.67$).[6] As in Experiment 1, we combined posteriors to get a direct estimate of the semantic attraction effect (d vs. c). The result shows that semantic attraction decreased response accuracy ($\hat{\beta} = -1.36$, 95%-CrI: $[-1.99, -0.78]$, $P(\beta < 0) = 0.999$) thus replicating the semantic attraction effect found in Experiments 1 and 2. Note that both attraction effects, morphosyntactic and semantic, were larger in Experiment 3 than in Experiments 1 and 2.

Conditions e–h. The average accuracy across conditions was 80%, ($\hat{\beta} = 1.4$, 95%-CrI: $[1, 1.8]$). Morphosyntactic attraction decreased response accuracy (85% vs. 74%, $\hat{\beta} = -0.67$, 95%-CrI: $[-1.1, -0.3]$, $P(\beta < 0) = 0.999$). Likewise semantic attraction decreased accuracy (88% vs. 68%, $\hat{\beta} = -1.2$, 95%-CrI: $[-1.6, -0.87]$, $P(\beta < 0) = 0.999$). There was again no interaction of morphosyntactic and semantic attraction (82% vs. 78%, $\hat{\beta} = -0.46$, 95%-CrI: $[-1.1, 0.21]$, $P(\beta < 0) = 0.91$). Again attractions effects were numerically larger in Experiment 3 than in Experiment 2.

To obtain an even more precise estimate of the interaction, we also combined data from all three experiments and repeated the last analysis (the confound in the stimulus design in Experiments 1 and 2 did not affect the relevant conditions). The analysis of the combined dataset (N=2,338) still showed no interaction (82% vs. 80%, $\hat{\beta} = -0.27$, 95%-CrI: $[-0.78, 0.24]$, $P(\beta < 0) = 0.85$).

Bayes factor analysis. We additionally conducted a Bayes factor analysis to quantify the evidence in favor of both attraction effects and against their interaction. In particular, Bayes factor estimates how much more likely is one model to have

[6] The percentage values for the last effect indicate the expected accuracy if semantic attraction had the same magnitude as morphosyntactic attraction vs. the observed accuracy in the semantic attraction condition (d).

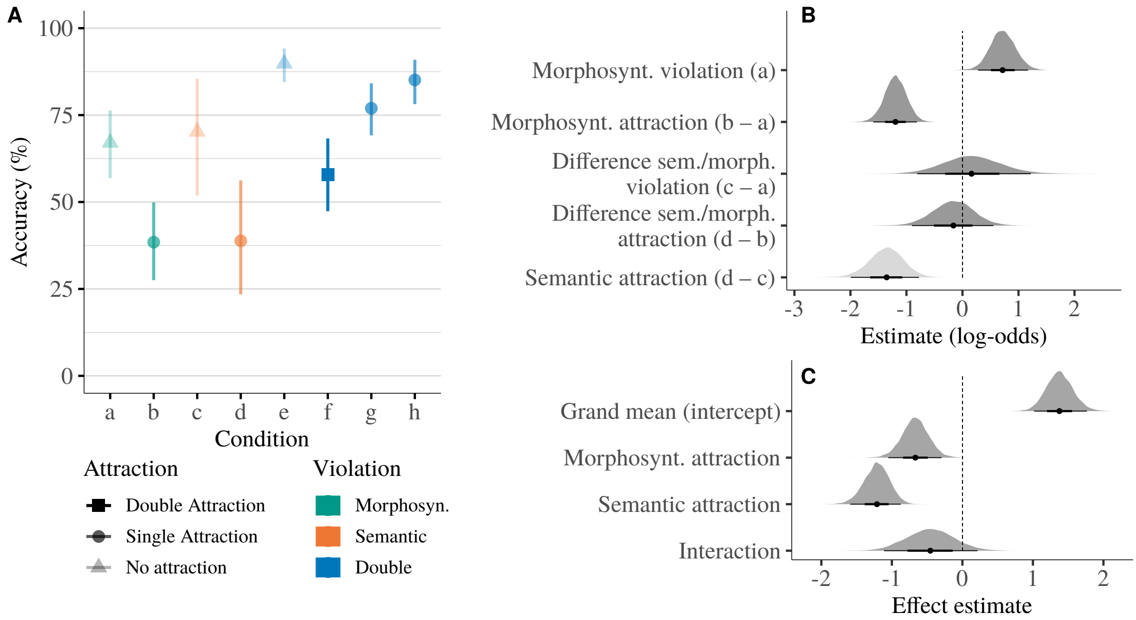

Figure 3.3: Results of Experiment 3. Panel A: Estimated condition means with 95% credible intervals. Panels B: Posterior distributions for the model of conditions (a)-(d). The posterior for semantic attraction (light gray) was obtained by combining the posteriors for morphosyntactic attraction and the difference between the semantic and morphosyntactic attraction. Panel C: Posterior distributions for the model of conditions (e)-(h). All parameters are on the log-odds scale. Error bars around estimates represent 66% (thick) and 95% (thin) credible intervals.

generated the data as compared to some other model, in our case, the model that does not include the predictor in question. As Bayes factor is sensitive to priors, we computed Bayes factors for a small range of plausible priors: the priors used in the analysis ($Normal(0, 1)$), a more informative prior ($Normal(0, 0.1)$), and an even less informative regularizing prior ($Normal(0, 1.5)$). For each model, we ran four chains with 20000 iterations each, the first 2000 samples were discarded as warm-up samples. The resulting Bayes factor values can be seen in Table 3.3.

Table 3.3: Experiment 3. Bayes factor values quantifying evidence in favor of the effects for a range of priors.

	Conditions (a–d)			Conditions (e–h)		
Prior	Morphosyntactic attraction (b-a)	Difference sem./morph. violation (c-a)	Difference sem./morph. attraction (d-b)	Morph. attraction	Semantic attraction	Interaction
Normal(0, 0.1)	6.09	1.01	1.28	1.50	3.20	1.07
Normal(0, 1)	54366847	0.34	0.32	198	5820245	0.82
Normal(0, 1.5)	53617260	0.21	0.21	123	4947099	0.43

Note that in our case, informative priors turn out to be too restrictive: with an intercept close to 65% and mean attraction effect size of more than 20% (effect estimate on the log-odds scale is ~ -1.20), this prior strongly biases attraction effects towards the smaller range of effect sizes. But even with such prior, we still have moderate evidence for attraction effects (except for morphosyntactic attraction in double violation condition). With less restrictive priors, we have strong evidence in favor of morphosyntactic attraction in single subject-verb violation conditions and for both morphosyntactic and semantic attraction effects in double subject-verb fit violation conditions. Semantic attraction in single violation conditions is also supported: there is anecdotal evidence against difference between morphosyntactic and semantic attraction effects in the single violation conditions. Nothing can be stated with regard to the interaction between morphosyntactic and semantic attraction effects, the evidence is inconclusive.

3.3.3 Discussion

The main goal of Experiment 3 was to replicate the results of Experiment 2 with an improved set of stimuli that are free from the design confound that would allow forming a noun-noun compound interpretation in one of the two semantic attraction conditions (d). All effects reported in previous experiments were successfully replicated.

To summarize, we found semantic attraction effects both in single and double subject-verb fit violation configurations, and the effect size of semantic attraction was similar to that of morphosyntactic attraction. These results are qualitatively consistent with the predictions of Lewis and Vasishth model and self-organized sentence processing model as outlined in the introduction. In the following section we investigate whether the Lewis and Vasishth model also provides a good quantitative fit to the data.

3.4 Computational simulation with Lewis and Vasishth model

In the following we explain the predictions of the Lewis and Vasishth model in more detail and investigate whether these predictions can be improved by modifying the parameters of the Lewis and Vasishth model. For this purpose we use an implementation of the Lewis and Vasishth model in R, the so-called interACT model (Engelmann et al., 2019).

We first define a linking hypothesis that allows us to link model dynamics to the response variable produced by our task. Out of the box, the model predicts the resulting parse and the time it will take to build this parse. The Lewis and Vasishth model does not internally track sentence grammaticality or well-formedness: a syntactic structure is either built or not if retrieval from memory failed. Additional assumptions are therefore necessary to let the model predict grammaticality judgments needed for the task. We adopt a simple mapping from failure to build a structure onto rejecting the sentence as ill-formed (correct response in our task), and from retrieving of a noun from memory and subsequent formation of a subject-verb dependency (correct or not) onto accepting the verb (incorrect response). The first scenario corresponds to correctly failing to build a parse (after all there is no correct parse), whereas the second scenario corresponds to the illusion of a correct parse when there is none.

Recall that retrieval failure happens when the activations of all chunks in memory lie below the retrieval threshold — the lower the activation of each chunk, the higher the probability of retrieval failure, and therefore, of a correct response. In all attraction conditions the attractor matches more retrieval cues than in the respective control conditions, which increases the activation of the attractor noun and the probability that it will be retrieved and attached. Therefore, in attraction conditions, the probability of a correct response is always predicted to be lower.

Furthermore, retrieval failure (leading to correct responses) should happen more often in conditions with double violation of subject-verb fit than in conditions with single violation of subject-verb fit and the Lewis and Vasishth model therefore predicts higher accuracy in those conditions (the exception is condition (f) where the attractor matches two features of the verb).

However, while it appears that the Lewis and Vasishth model could in principle explain the qualitative pattern of results in our data (see Figure 3.4A), the default version of the Lewis and Vasishth model, as reported by Lewis and Vasishth (2005), has a retrieval activation threshold that is so low that some item will always be retrieved from memory even if it mismatches two out of three retrieval cues. Consequently, the Lewis and Vasishth model with default parameter settings predicts no failed retrievals, and hence 0% accuracy in all condition we tested which is clearly implausible. To address this shortcoming of the model, we next explore through simulations whether changes in two relevant parameter values allow the model to fit the observed pattern qualitatively and quantitatively.

Condition			Retrieval cues	Subject matches	Attractor matches
a.	The drawer with the handle	OPEN	+SUBJ +PL +OPENABLE	+SUBJ +OPENABLE	
b.	The drawer with the handles	OPEN	+SUBJ +PL +OPENABLE	+SUBJ +OPENABLE	+PL
c.	The drawer with the handle	CUTS	+SUBJ +SG +CAN_CUT	+SUBJ +SG	+SG
d.	The drawer with the knife	CUTS	+SUBJ +SG +CAN_CUT	+SUBJ +SG	+SG +CAN_CUT
e.	The drawer with the handle	CUT	+SUBJ +PL +CAN_CUT	+SUBJ	
f.	The drawer with the knives	CUT	+SUBJ +PL +CAN_CUT	+SUBJ	+PL +CAN_CUT
g.	The drawer with the knife	CUT	+SUBJ +PL +CAN_CUT	+SUBJ	+CAN_CUT
h.	The drawer with the handles	CUT	+SUBJ +PL +CAN_CUT	+SUBJ	+PL

Table 3.4: Summary of cue-feature matches for the subject and the attractor nouns across experimental conditions and mean rates of retrieval failures predicted by Lewis and Vasishth model

3.4.1 Simulations

We used grid search to systematically vary two parameters that affect the probability of a retrieval failure: the retrieval activation threshold and the noise parameter. We then identified the set of parameters that most closely reproduced the observed effects in Experiment 3. Prediction error was quantified in terms of the average mean-squared error across the eight experimental conditions. The simulation was run for 5000 iterations for each combination of parameters.

The interACT implementation of the Lewis and Vasishth model05 (Engelmann et al., 2019)[7] only supports two types of cues and was therefore modified to support all three cues needed for the present purposes: structural (indicating whether a noun is in subject position, `+SUBJ`), morphosyntactic (`+SG`, `+PL`), and semantic (e.g., `+CAN_CUT`). Table 3.4 shows cue-feature match patterns for all conditions of one example item.

In Lewis and Vasishth model, the probability of retrieving a word from memory depends on three parameters:

$$\text{Probability of retrieval} = \frac{1}{1 + e^{\frac{\tau - A}{s}}}$$

We varied two of those parameters: τ and s. Parameter τ is the *retrieval activation threshold*: the higher the threshold, the lower the probability that some item will be retrieved from memory. If none of the candidates reaches the activation threshold, parsing fails. In ACT-R (but not the Lewis and Vasishth model) the default value of this parameter is 0. We varied it around 0 within the boundaries of -1.5 to 1.5 in 13 steps of size 0.25.

Parameter s represents the amount of noise in the system, e.g. random fluctuations in activation. It can increase or decrease item activation, which affects the probability of its retrieval. The more noise there is in the system, the less likely it is that the correct item will be retrieved. If noise is close to 0, the transition from low to high probability of retrieval is abrupt, and when noise is greater, the transition will follow a sigmoidal function. We varied noise between 0.05 and 0.5 in 10 steps of 0.05 (the default value is 0.2, and in general ACT-R modeling it is typically varied below 0.5).

[7]The code of the model is publicly available at https://github.com/felixengelmann/inter-act/, also available as a Shiny App: https://engelmann.shinyapps.io/inter-act/.

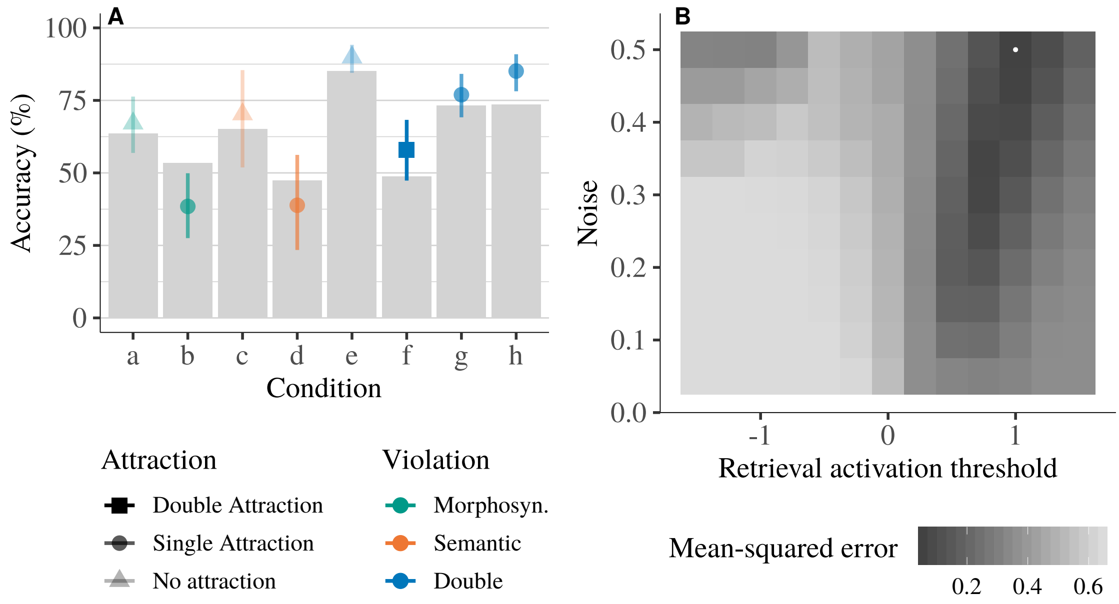

Figure 3.4: Modeling results. Panel **A**: model predictions compared to the observed data. Gray bars represent predictions of Lewis and Vasishth model with the best-fitting set of parameters. Colored lines represent the 95% credible intervals of the observed condition means from the pooled dataset. Panel **B**: assessment of the fit between the modeled and observed effects as a function of two parameters. The lower the mean-squared error, the better the fit. The white dot marks the best parameter set.

3.4.2 Results

The model predictions generated by the best-fitting set of parameters (retrieval activation threshold: 1, noise: 0.50) are shown in Figure 3.4A (grey bars). The model qualitatively predicts all effects we observed: the morphosyntactic and semantic attraction effects both in single and in double subject-verb fit violation conditions, as well as the double attraction effect (the standard error of the model's predictions are below 1%, which means that a difference of several percent between conditions is likely robust). Quantitatively, the model's predictions lie within the 95% credible intervals for six out of eight conditions.

Figure 3.4B shows how parameter values affected model fit. The retrieval activation threshold affects the model fit to a greater degree than noise, but higher noise values also contribute to a better fit because noise can reduce the activation of the most active item and thus lead to retrieval failures.

To assess whether the predictions of the Lewis and Vasishth model are sufficiently constrained and the model does not predict reverse attraction effects under other

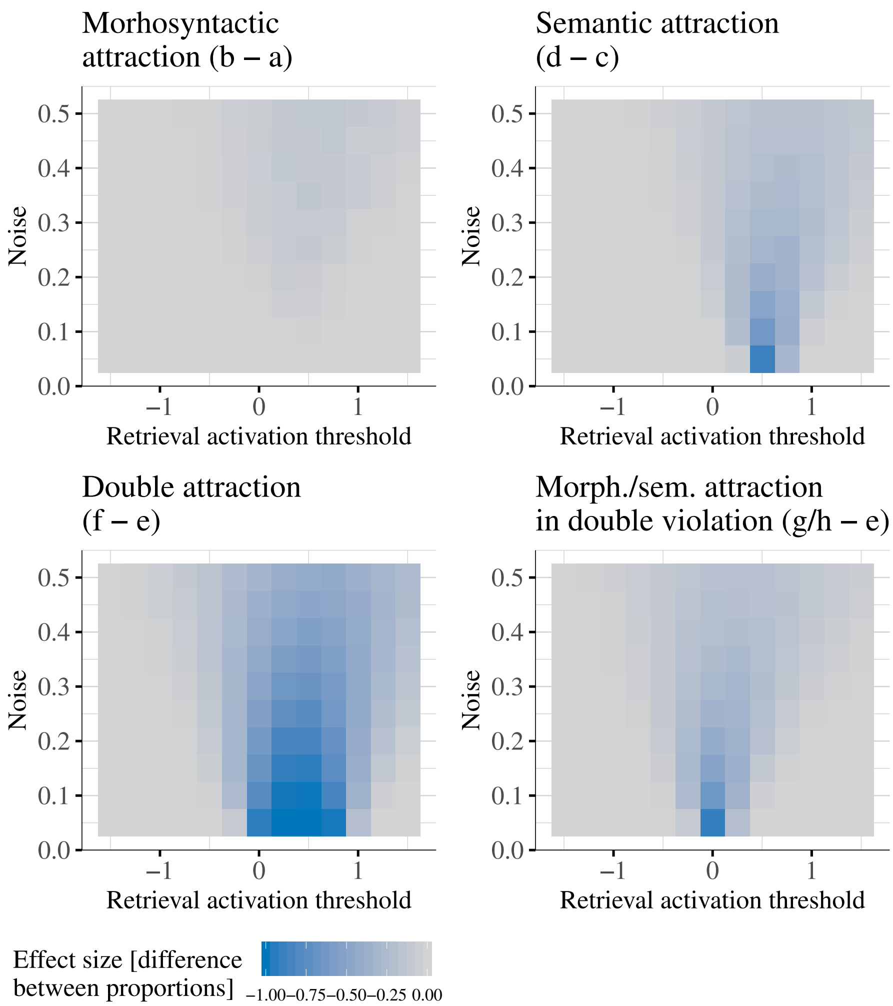

Figure 3.5: Attraction effect predicted by Lewis and Vasishth model for acceptability judgments as a function of parameter value. Note that every predicted attraction effect goes in the right direction (lower accuracy in attraction conditions). Figure contains only four panels for five attraction effects since from the point of view of the model, semantic and morphosyntactic attraction effects in double verb violation setup are the same, predictions for conditions (g) and (h) do not differ.

parameter configurations, we computed the whole range of the Lewis and Vasishth model predictions for the five attraction effects generated by all possible parameter values, see Figure 3.5. The crucial insight is that Lewis and Vasishth model always predicts correct effect direction (decrease in accuracy due to attraction) or no effect, but never an incorrect effect direction (increase in accuracy due to attraction).

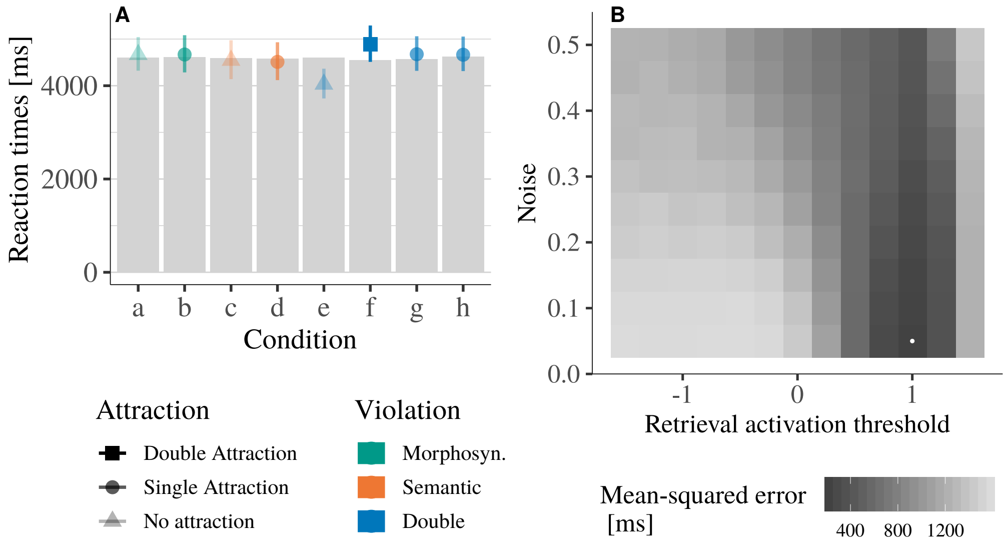

Figure 3.6: Modeling RTs. Panel **A**: model predictions compared to the data observed in Experiment 3. Gray bars represent predictions of Lewis and Vasishth model with the best-fitting set of parameters. Colored lines represent the 95% credible intervals of the observed condition means from the pooled dataset. Panel **B**: assessment of the fit between the modeled and observed effects as a function of two parameters. The lower the mean-squared error, the better the fit. The white dot marks the best parameter set.

3.4.3 Modeling reaction times

As we already mentioned, Lewis and Vasishth model predicts both the parsing outcome and the time it takes to build the parse. All of the previous modeling work focused largely on evaluating Lewis and Vasishth model predictions with regard to RTs rather than parsing outcomes (accuracies). Therefore, to more fully assess how the model's predictions fit our data, we also modeled reaction times. Note that this assessment is of a more limited nature, since in our task, reaction times incorporate both the time it takes to read the whole sentence preamble (and preambles vary in lengths across conditions) and the time it takes to make the decision. Under these circumstances, no full match between model predictions and the observed data can be expected, but a fundamental mismatch might still be informative.

We varied the retrieval activation threshold and noise parameter values within the same boundaries as before. In addition, we varied the latency factor (between 7.5 and 15 in 16 steps of 0.5), the most freely valued parameter in the ACT-R framework that scales model's predictions into a numerical range comparable with the data. The rest of the modeling settings remained the same.

The fit provided by the best-fitting set of parameters (retrieval activation threshold = 1, noise = 0.00) is presented on Figure 3.6A. We immediately see that the best-fitting values of the noise parameter are radically different in the modeling of RTs and acceptability judgments, and that lower noise values provide better fit in case of RTs (Figure 3.6B). Still, higher noise values that provided the best fit for acceptability judgments allow a reasonable fit for reaction times as well. The most important outcome of the evaluation of Lewis and Vasishth model predictions for RTs is that despite small numerical mean-squared error, the model fails to capture any of the slowdowns due to attraction effects that are present in the data (for double attraction effect and for the morphosyntactic and semantic attraction effects in double subject-verb fit violation[8]). Instead, the best-fitting set of parameters predicts no difference in RTs due to attraction. When we assess the range of model predictions generated by all possible parameter values (see Figure 3.7), we see that the model predicts either no difference, or a speedup due to attraction, while we observe a slowdown.

The speedup that Lewis and Vasishth model predicts for attraction conditions follows directly from the specification of the model and the mapping from modeling outcomes to acceptability judgments. Recall that retrieval failures are mapped onto correct responses, and compare the time it takes to register a retrieval failure to the time needed for a successful retrieval:

$$\text{Retrieval failure} = latency factor \times e^{-\tau}$$
$$\text{Successful retrieval} = latency factor \times e^{-A}$$

Here, A is the activation of the chunk that is retrieved, and τ is the retrieval activation threshold. For any chunk to be retrieved from memory, its activation A must be greater than τ, therefore, any retrieval will by definition be faster than retrieval failure. It follows that control conditions without attraction with higher proportion of retrieval failures are predicted to be processed longer than conditions with attraction. One exception is the configuration where A and τ are so close that the latency of successful retrieval is almost the same as the latency of retrieval failure. In such a case, there will be little difference in processing times between conditions —

[8]The slowdown is present for morphosyntactic and semantic attraction effects in single subject-verb fit violation as well, but only in trials with correct responses, see Appendix 6.1.4.

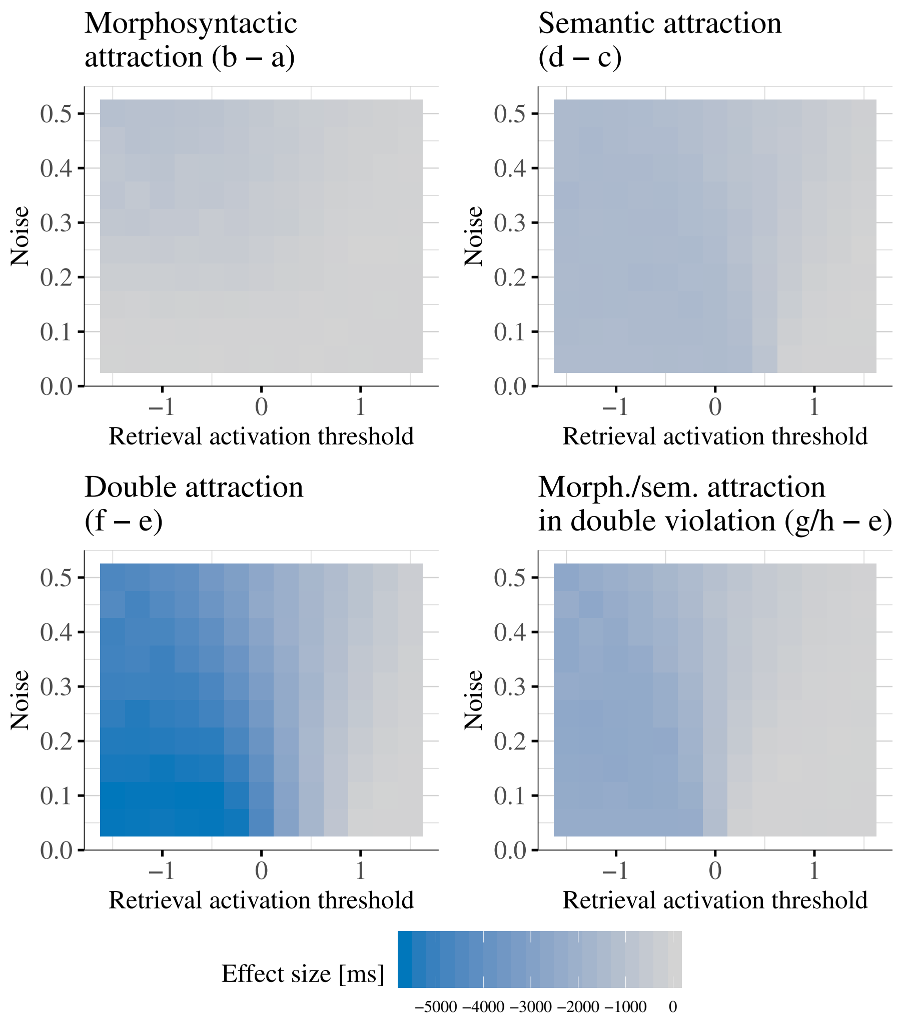

Figure 3.7: Attraction effect predicted by Lewis and Vasishth model for response times as a function of parameter value. Note that every predicted attraction effect goes in the wrong direction (predicted lower RTs in attraction conditions).

this is exactly the best-fitting parameter combination for the RTs. Importantly, there are no parameter configurations that predict a positive difference (corresponding to the observed slowdown) between the conditions with attraction and their respective control conditions without attraction.

A natural objection would be that the slowdown we observed might stem from reading sentence preambles of varying lengths rather than from processing attraction. But modeling of reaction times suggests that this is not the case: when the length of the preamble is taken into account, there is still a clear slowdown in the attraction conditions (see Appendix 6.1.4). In addition, slowdowns in judgment times are consistently observed in attraction conditions in studies without confounds in measuring RTs (e.g., Avetisyan et al., 2020; Lago & Felser, 2018; Reifegerste et al., 2020; Schlueter et al., 2019; Staub, 2009). Therefore, even if we put aside reaction times from the presented experiments, there will still be a fundamental discrepancy between the speedup predicted by the Lewis and Vasishth model and the repeatedly observed slowdowns in judgment times, providing a systematic challenge for Lewis and Vasishth model. To accommodate the data, Lewis and Vasishth model is likely to require an additional processing component that operates on top of structure building and specifically models the processes deployed in the grammaticality judgment task.

3.4.4 Discussion of simulation results

We demonstrated that Lewis and Vasishth model in general predicts the correct direction of all attraction effects in acceptability judgments, and that varying the values of two parameters allowed Lewis and Vasishth model to approximate condition means with a good quantitative fit. However, the best-fitting value of noise seems implausible: the estimated value (0.50) was higher than the estimate obtained for participants with aphasia (Mätzig et al., 2018). As our participants did not have speech- or language disorders, the high value seems unlikely. Moreover, there was a fundamental discrepancy between model predictions and the observed condition means for the reaction times: while we observed longer RTs for attraction conditions, Lewis and Vasishth model predicts either no difference in RTs or faster RTs for the attraction conditions, which perfectly fits attraction effects in reading ill-formed sentences, but not in judging them.

3.5 General discussion

The main goal of this study was to establish whether the well-known agreement attraction effect has an analogue in the semantic domain, as Cunnings and Sturt (2018) claimed. In doing so, we also aimed to disambiguate between morphosyntactic theories of agreement attraction, which do not predict semantic attraction effects, and more general sentence processing theories, which do predict semantic attraction. In two out of three experiments, we replicated the classic morphosyntactic agreement attraction effect (recall that the agreement attraction effect was relatively small in Experiment 2) and in all three we also found robust evidence for semantic attraction effects that were similar in size. Specifically, participants were more likely to accept an unlicensed plural verb as a continuation of sentence fragments containing a singular subject when another plural noun was present (agreement attraction: 'The drawer with the handles open'). Likewise, participants were also more likely to accept a verb that mismatched the subject semantically as a continuation of the sentence when another noun matched the verb's semantic requirements (semantic attraction: 'The drawer with the knives cuts'). The fact that morphosyntactic and semantic attraction effects were similarly sized suggests that both types of errors may be subserved by a common processing mechanism. The lack of interaction between the morphosyntactic and semantic attraction effects is consistent with both a common and with two distinct processing mechanisms.

Both Lewis and Vasishth model and self-organized sentence processing model predict the observed effects qualitatively. We conducted computational simulations with Lewis and Vasishth model (Engelmann et al., 2019) in order to test whether the model also provides a good qualitative fit. In the following, we briefly discuss the implications of our findings for the individual theoretical accounts.

3.5.1 Feature percolation and marking and morphing

Semantic attraction is not covered by purely morphosyntactic models of attraction, such as feature percolation (Nicol et al., 1997; Vigliocco & Nicol, 1998) and marking and morphing (Bock et al., 2001), because the phenomenon manifests entirely on the semantic level with no involvement of morphosyntax. To incorporate semantic attraction effects, these accounts would need to be either significantly expanded by changing some of their core assumptions, or their principles be incorporated into a more general model of attraction mechanism. The latter option seems preferable, as it acknowledges that these accounts elegantly capture some unique properties of agreement attraction on a particular level of language organization, such as the singular-plural asymmetry and the influence of notional number on morphosyntactic attraction effects.

The integration of ideas from feature percolation and marking and morphing and more general models such as Lewis and Vasishth model and self-organized sentence processing model can take many different shapes, and a detailed discussion and evaluation is far beyond the scope of the present research. For instance, it is not clear how precisely the percolation of features could be implemented Lewis and Vasishth model and self-organized sentence processing model and whether that would even make sense within these models. However, implementing ideas about markedness of number features might be relatively straightforward. For instance, to account for the singular-plural asymmetry in agreement attraction, it might be sufficient to have only `+PLURAL` but no corresponding `+SINGULAR` features in Lewis and Vasishth model, as originally proposed by Wagers et al. (2009). As to the notional plurality effects, the part of the marking and morphing model that accounts for these is already covered by self-organized sentence processing model (Smith et al., 2018): the effects were successfully modeled by decomposing them into several smaller-scale semantic features.

3.5.2 Lewis and Vasishth 2005 model

The Lewis and Vasishth model predicts semantic, morphosyntactic, and double attraction effects, and by allowing the values of some parameters to take non-default values it can closely, though not perfectly, reproduce the observed condition means and effect sizes. This suggests that the Lewis and Vasishth model might claim the place of the universal account explaining attraction effects in all possible configurations. But some evidence speaks against that: first, the value of the noise parameter that provided the best fit to the data is problematic from the point of view of cognitive processing. High noise value has no external justification as our participants had no known language disorders. Second, recall that the Lewis and Vasishth model does not cover the full range of findings about agreement attraction effects. The singular-plural asymmetry as well as notional plurality effects currently lie beyond the scope of the model. While the singular-plural asymmetry could in principle be captured, it is unclear how or whether at all notional plurality effects could be captured by Lewis and Vasishth model. Finally, the Lewis and Vasishth model fails to capture the pattern of reaction times in acceptability judgment task both in our experiments and in other reported studies (Avetisyan et al., 2020; Lago & Felser, 2018; Reifegerste et al., 2020; Schlueter et al., 2019; Staub, 2009).

3.5.3 Self-organized sentence processing model

As in Lewis and Vasishth model, semantic features in self-organized sentence processing model are on par with other types of features. While the abstract description of the two models' mechanics differ, the predictions of self-organized sentence processing model seem to mirror those of Lewis and Vasishth model. self-organized sentence processing model predicts both semantic and morphosyntactic attraction effects in ungrammatical sentences if we assume the same mapping from parsing outcome to the acceptability judgments as for Lewis and Vasishth model.

In their current forms both self-organized sentence processing model and Lewis and Vasishth model have almost the same strengths and weaknesses. Both predict semantic attraction effects, and can be extended to account for the singular-plural asymmetry. Unlike Lewis and Vasishth model, however, self-organized sentence processing model also covers the notional plurality effects (Smith et al., 2018). This makes self-organized sentence processing model so far the most comprehensive model potentially able to explain all of the observed attraction effects. Whether this is indeed the case, can only be confirmed via simulations.

3.5.4 Limitations

The scope of our study was limited to the processing of ungrammatical sentences, therefore we cannot fully evaluate the performance of the theoretical accounts and the models we considered. Further evaluation on grammatical and semantically well-formed sentences would provide important insight, as the two broad groups of accounts make contradicting predictions with respect to processing such sentences. The lack of comparison with well-formed sentences necessarily limits the conclusions we can draw: despite the good model fit for Lewis and Vasishth model in acceptability judgments, it is entirely possible that the attraction effects we observe reflect not the miscasting of the attractor noun as the subject of the sentence, as Lewis and Vasishth model and self-organized sentence processing model predict, but rather participants' efforts to reanalyze the input they have correctly identified as ill-formed (as suggested by Lago et al., 2015) and to make sense of it. Our study cannot reliably distinguish which noun was considered to be the subject of the sentence, and what representations participants built as a result. Schlueter et al. (2019) claim that attractor noun is misrepresented to be the subject of the sentence only in some instances of agreement attraction, not always when attraction errors are made. However, it is in general difficult to establish which noun was retrieved during parsing, as question responses might reflect not the structure built during online processing, but rather some salvageable post-hoc interpretation (Bader & Meng, 2018), and these general reservations apply also to the Schlueter et al. (2019) findings.

If our conclusions are limited to the processing of ungrammatical structures, a question might arise about why even evaluate the performance of the models on ill-formed linguistic material. We believe that such evaluation clearly defines the scope of application of a processing model: it is important for the models of sentence comprehension to distinguish between ill- and well-formed structures, as humans do. After identifying the structure as ill-formed, the model can still try to make sense of it, as humans also do, whether according to the principles of rational inference (Levy, 2008b), or in some other way. Finally, models can be seen as cognitively plausible if they make the same kinds of mistakes as humans make, and do not make the mistakes that humans do not make (for example, see the evaluation of neural networks processing subject-verb agreement by Arehalli and Linzen, 2020; Linzen and Leonard, 2018).

A related concern is that semantic attraction errors are hardly ever encountered outside of the experimental setup, while morphosyntactic attraction is more common (approximately 0.1% to 0.5% rate in written corpora, Stemberger, 1984). Again, this could be seen as evidence that the effects we observed might reflect reanalysis rather than failing to notice sentence ill-formedness. Even in that case, our results are still highly informative: they show that despite only morphosyntactic, but not semantic attraction occurring naturally in production, both have similar profiles in comprehension. This lack of difference suggests that morphosyntactic attraction effects in comprehension are not mainly driven by the processes postulated in feature percolation or marking and morphing models. Of course, the evidence provided by our study is indirect, and further evidence is needed to disentangle these options.

With regard to modeling, one further limitation is that neither Lewis and Vasishth model nor interACT currently take into account human tendency to consider the sentences to be well-formed by default, demonstrated by Hammerly et al. (2019). We currently map failed parsing onto rejecting the sentence as ill-formed, but it could also be the case that failed parsing would be still mapped onto accepting the sentence as well-formed by default in some proportion of cases. As this model modification would affect each condition to a different degree, it is difficult to predict how it could influence the modeling outcomes.

3.6 Conclusion

In this study, we provided evidence for a semantic attraction effect mirroring the well-known agreement attraction effect in sentence comprehension. The semantic attraction effect is predicted by two general language-processing models (Lewis and Vasishth model and self-organized sentence processing model), and reading time results by Cunnings and Sturt (2018) recently provided initial evidence for its existence.

In three experiments, we thoroughly investigated semantic attraction using an experimental paradigm designed for attraction phenomena and compared it directly to agreement attraction. We found that the semantic attraction effect is similar in size (and reaction times profile) to the classic morphosyntactic agreement attraction effect. This finding suggests that both effects may be subserved by the same underlying mechanism and/or processing principles. If true, it follows that the focus of models specifically designed to explain morphosyntactic attraction may be too narrow, and that agreement attraction is just one instance of a potentially much broader phenomenon. Beyond semantics and morphosyntax, attraction could also manifest on the phonological level (although disentangling attraction from coarticulation might prove difficult).

Regarding the mechanism that might explain both semantic and morphosyntactic attraction, our findings are most compatible with theoretical accounts assuming that all possible linguistic features — morphosyntactic and semantic alike — are evaluated concurrently, such as Lewis and Vasishth model (Lewis & Vasishth, 2005) and self-organized sentence processing model (Smith et al., 2018; Tabor & Hutchins, 2004; Vosse & Kempen, 2000).

At the same time, our data pose a challenge at least for Lewis and Vasishth model that fails to capture a broad and robust pattern in acceptability judgment reaction times: it predicts faster RTs in attraction conditions, while slower RTs are consistently observed. Addressing this shortcoming might be a fruitful topic for future research.

Chapter 4

Agreement attraction and inhibitory interference in well-formed sentences

In psycholinguistic theory development, studying the particular mechanism postulated by a certain theory in isolation helps to determine the limits of the theory's explanatory power. However, never going further and ignoring that other theories postulate other, possibly counteracting mechanisms that operate in the same circumstances, could potentially hinder the progress of every theory involved, and of the field in general. One example of two counteracting mechanisms predicted to operate in the same circumstances is the case of similarity-based interference and agreement attraction, predicted by two broad classes of theoretical accounts, which we will refer to as similarity-based interference accounts (Lewis & Vasishth, 2005; McElree, 2000) and the faulty encoding accounts (Bock & Eberhard, 1993a; Eberhard et al., 2005). Both make predictions in regards to the processing of grammatical sentences that contain a subject noun, a verb, and some interfering noun(s) (called *attractor(s)* in the faulty encoding accounts, and *distractor(s)* in the interference accounts) matching or mismatching the morphosyntactic marking of the subject noun, see example (1) adapted from Bock and Miller (1991).

(1) a. The computer installed in the Russian antiballistic missile is outdated.
b. The computer installed in the Russian antiballistic missiles is outdated.

The interference accounts predict a processing slowdown at the verb in (1-a), the faulty encoding accounts predict a complementary slowdown in (1-b), and neither effect is consistently observed. We propose that the effects might be not absent, but rather present at the same time, canceling each other out and thus seemingly undermining the predictions of both classes of accounts.

If this is the case, and the predictions of both classes of accounts are correct, the theories aiming to account for language processing in general, such as the similarity-based interference accounts (Lewis & Vasishth, 2005; McElree, 2000) and the expectation-based accounts (Hale, 2001; Levy, 2008a) would need to incorporate the slowdown in grammatical sentences such as (1-b) predicted by the faulty encoding accounts. Currently this effect is largely believed to be non-existent and is not predicted by any of the general language-processing accounts.

One of the main reasons that the issue of conflicting predictions has not been addressed before is that the interference and the faulty encoding accounts have historically (until Wagers et al., 2009) been investigated by non-overlapping researcher communities. A further complication is that the empirically observed effects are contradictory: while several studies reported a slowdown predicted by the faulty encoding accounts, most of these studies had design shortcomings that compromise the interpretation of the results. At least one large-scale study found some inconclusive evidence for the opposite slowdown predicted by the interference accounts (Nicenboim et al., 2018), but the overwhelming majority of studies found no difference at all.

The inconsistent outcomes and the lack of difference are equally problematic for both groups of accounts: for the interference accounts, they cast doubt on the existence of morphosyntactic interference and limit interference to the semantic domain; for the faulty encoding accounts, they limit the scope of the agreement attraction phenomenon in comprehension to the ungrammatical sentences, which undermines its usefulness for explaining normal sentence processing.

We will next review the mechanisms proposed by the two groups of accounts to drive the predicted effects. Based on these mechanisms, we'll propose an experimental design that should differentiate the slowdowns across conditions, which could bring one of the predicted effects to the surface, and thus demonstrate that in a typical experimental setup both effects were at play simultaneously.

4.0.1 Faulty encoding accounts

The faulty encoding accounts were originally developed to explain the *agreement attraction* phenomenon is sentence production. Agreement attraction in number refers to erroneous selection of verb number controller in production, see (2).

(2) We speculate that the difference between the two studies in the pairwise effects stem from...

Parallel effects were reliably observed in comprehension, so the mechanisms postulated by the faulty encoding accounts were consequently extended to affect comprehension. The extension is based on the assumption that comprehension heavily relies on language production system (Christiansen & Chater, 2016; Meyer et al., 2016; Pickering & Garrod, 2013). These parallel effects in comprehension include overlooking attraction errors: sentences containing such errors are more often judged as grammatical than sentences without an interfering noun matching the number of the verb (Hammerly et al., 2019; Patson & Husband, 2016; Wagers et al., 2009). This is referred to as an *illusion of grammaticality.* Another attraction effect in comprehension is reflected in reading times: the verb in ungrammatical constructions such as (3-b) is read faster than in (3-a). This facilitation is observed very consistently (inter alia, Dillon et al., 2013; Jäger et al., 2019; Lago et al., 2015; Pearlmutter et al., 1999b; Tucker et al., 2015; Wagers et al., 2009), and the accumulated evidence is very persuasive: a recent meta-analysis estimated a facilitatory effect of -22 ms, with a 95% credible interval (CrI) lying between $[-36, -9]$ ms (Jäger et al., 2017b).

(3) a. *The computer installed in the Russian antiballistic missile are outdated.
    b. *The computer installed in the Russian antiballistic missiles are outdated.

The faulty encoding accounts propose two distinct mechanisms underlying the illusion of grammaticality that nevertheless share a core property: they assume that the number of the subject is erroneously encoded – either as unambiguously plural (the feature percolation account), or as somewhat plural on the plurality continuum (the marking and morphing account).

The feature percolation account (Bock & Eberhard, 1993a; Franck et al., 2002a;

Vigliocco et al., 1995) posits that a plural feature of the attractor might occasionally erroneously percolate up the syntactic tree and contaminate subject number marking. In that case, the subject is encoded as plural, and feature checking at the verb marked for plural returns no error signal.

The marking and morphing account (Bock et al., 2001; Eberhard et al., 2005) postulates that computation of subject number depends, among other factors, on the weighted sum of plural morphemes on words comprising the subject noun phrase. This means that a plural feature on an interfering noun in the subject noun phrase can disrupt number computation for a singular subject. If that happened and the subject received a number value ambiguous between singular and plural, the subject will be with some probability encoded as plural, and feature checking at the verb marked for plural will be successful. While the proposed mechanisms of the feature percolation and the marking and morphing accounts differ, the predictions are essentially the same, except that the marking and morphing account can potentially cover attraction arising from nouns outside of the noun phrase (but this possibility is not instantiated in the current version of the model, see Eberhard, Cutting, and Bock 2005, p. 544).

Feature percolation and marking and morphing also share their predictions about the processing of grammatical sentences, such as (1-a) and (1-b). Just as in the previous scenario, the parser would occasionally encode the number of the subject as plural, and when a singular verb is encountered, an *illusion of ungrammaticality* would arise. This should lead to longer average reading times on the verb in (1-b) as compared to (1-a). The actual findings are contradictory: some studies reported the predicted slowdown in processing grammatical sentences (Franck et al., 2015; Lago et al., 2015; Nicol et al., 1997; Patson & Husband, 2016; Pearlmutter et al., 1999b; Wagers et al., 2009), but many of these studies had design confounds. The main concern is that the slowdown at reading the verb might actually be a spillover from the processing of the previous region, the plural noun. Plural nouns could be difficult to process for several reasons: they are longer than singular nouns (and might thus take longer to read), they are less frequent, more morphologically complex, and might be more difficult to integrate into a context that contains singular nouns and into discourse representation. Although many studies are dismissed due to the spillover concern, there is some evidence for the predicted slowdown: it was found using the maze paradigm (Experiments 1, 2, and 4 by Nicol et al., 1997), which is claimed to be free from spillover effects (Boyce et al., 2019).

Those reading studies free from the spillover confound report no difference between conditions (Avetisyan et al., 2020; Cunnings & Sturt, 2018; Lago et al., 2015; Paspali & Marinis, 2020; Patson & Husband, 2016; Thornton & MacDonald, 2003; Tucker et al., 2015; Wagers et al., 2009). The meta-analysis conducted by Jäger et al. turned out inconclusive, although there is some indication of the predicted facilitation: Est. = -7 ms, 95% CrI = [-16, 4] ms. But the outcome of the meta-analysis depends on the data it is based on: the meta-analysis itself cannot resolve the problem with spillover effects contaminating the measure of interest.

To summarize, both faulty encoding accounts predict an illusion of ungrammaticality in processing grammatical sentences, but this effect is rarely observed in experiments that do not raise any internal validity concerns. The lack of support for the prediction is seen as evidence against applying the production-based faulty encoding models to comprehension. This position has recently been challenged by Hammerly et al. (2019) who argue that the ungrammaticality illusion is present in comprehension of grammatical sentences, but concealed by a bias towards "grammatical" response in the grammaticality judgment task. They show that when response bias is neutralized, attraction effects in grammatical sentences surface. While this finding is important for the tasks that require explicit reasoning about the experimental materials, it is not immediately apparent how this might be applicable to reading. No grammaticality judgment is usually required in the studies measuring reading speed, and therefore response bias alone cannot explain the lack of the illusion of ungrammaticality in reading times.

As existence of attraction effects in the processing of grammatical sentences is under doubt, the recent decade brought a general shift of the paradigm: many researchers now believe that since attraction is not consistently observed during comprehension of grammatical sentences, faulty encoding accounts do not adequately capture comprehension, and similarity-based interference is the only mechanism needed to cover all the attraction effects observed in comprehension, which are in this case reduced to attraction in ungrammatical sentences (Hammerly et al., 2019; Tanner et al., 2014; Wagers et al., 2009).

4.0.2 Similarity-based interference accounts

In contrast to faulty encoding accounts, the group of similarity-based interference accounts (Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, 2000; Van Dyke & McElree, 2006) was developed to model language comprehension. They provide a broad theoretical framework making predictions about language processing in general, not limited to sentences of a certain structure. The similarity-based interference accounts (also referred to as cue-based retrieval accounts) assume that sentence processing relies on a series of fast retrievals of previously processed constituents from content-addressable memory in order to build a syntactic structure in real-time. The speed and/or accuracy of these retrievals depends on how unique the features of the to-be-retrieved element are. If the to-be-retrieved element shares features with other elements present in memory, retrieval could take longer or another element could be erroneously retrieved instead. Both the slowdown in processing times and the erroneous retrieval are referred to as (the consequences of) interference.

Two models, in particular, were applied to language processing — the Lewis and Vasishth model (Lewis & Vasishth, 2005) built on the cognitive architecture ACT-R (J. R. Anderson, 1996) and the direct access model (Martin & McElree, 2011; Van Dyke & McElree, 2006). The direct access model makes quantifiable predictions only with regard to retrieval accuracy, but not processing times (although see an instantiation of the model that also predicts processing times in Nicenboim and Vasishth, 2018), while the Lewis and Vasishth model makes predictions regarding both the accuracy and processing times. Since we are mostly interested in processing times, we will focus on the predictions of the Lewis and Vasishth model. To do that, we will briefly review the model's mechanics.

Consider the ungrammatical sentences from (3). When the verb *are* is being processed, it provides retrieval cues such as `+C-COMMAND, +PLURAL` in order to retrieve the subject and complete the subject-verb dependency. These retrieval cues send a fixed amount of spreading activation (divided equally among all cues by default) to all items in memory that have matching features. The spreading activation these items receive adds to the base-level activation they already have. The item in memory with maximal activation (that is also greater than the retrieval activation threshold) will be retrieved to form the dependency. The higher the activation of the item, the greater is the retrieval speed.

The Lewis and Vasishth model straightforwardly accounts for the slowdown in processing ungrammatical sentences such as (3-a) as compared to their grammatical counterparts. In an ungrammatical sentence, the subject matches only the `+C-COMMAND`, but not the `+PLURAL` cue. It gets spreading activation from only one cue out of two, and will be retrieved slower than if the sentence was grammatical and it received spreading activation from both cues. The Lewis and Vasishth model also accounts for faster processing of (3-b) as compared to (3-a), but the mechanism of the speedup is a bit more complicated. When an ungrammatical sentence has an interfering noun that matches the other retrieval cue, as *missiles* matches `+PLURAL` in (3-b), the subject noun and the interfering noun each get half of the spreading activation from the verb. Since their baseline activation levels are also comparable, the resulting activations of the subject noun and the interfering noun would be very similar. Recall that the retrieval speed depends on the activation of the to-be-retrieved item. On each trial, the noun with a (slightly) higher activation will be retrieved. That means, on every trial, retrieval will be a bit faster when there are two nouns as in (3-b) as compared to one noun in (3-a), where the processing times depend exclusively on the activation of the subject, however high or low it might be. The predicted speedup in the processing of ungrammatical sentences with the interfering noun matching some of the retrieval cues is called *facilitatory interference*. Facilitatory interference has been extensively tested and universally found — it is exactly the facilitation that the faulty encoding accounts also predict in ungrammatical sentences, albeit for different reasons.

Now consider the predictions of the Lewis and Vasishth model for the processing of grammatical sentences such as (1-a) and (1-b): The verb *is* sets retrieval cues `+C-COMMAND, +SINGULAR`[1]. Recall that if more than one word matches a certain retrieval cue, such as both the subject and the interfering noun *the missile* matching the number cue in (1-a), the spreading activation from the `+SINGULAR` cue will be divided equally between the words. The subject will now get less spreading activation than in (1-b), its total activation will be lower, and it will take longer to retrieve than

[1] Hammerly et al., 2019; Wagers et al., 2009 argue against introducing a separate `+SINGULAR` feature, assuming that singular number being represented by the absence of the `+PLURAL` feature. We will return to this in more detail in the General discussion section.

in (1-b). This is referred to as *inhibitory interference*. Due to inhibitory interference arising in (1-a), the Lewis and Vasishth model predicts longer processing times in (1-a) than in (1-b).

This particular prediction was only rarely tested using grammatical number: the predicted inhibitory interference effect was found by (Franck et al., 2015), but a large-scale study with 180 participants by (Nicenboim et al., 2018) turned out inconclusive. The 95% credible interval for the number interference effect included 0 ($Est. = 9\,\text{ms}$, $95\%CrI = [0, 18]\,\text{ms}$). The lack of persuasive number interference effect is problematic for the interference accounts, as the number feature does not have any special status and is predicted to create interference just as any other feature.

Interestingly, while the lack of attraction effects in comprehension of grammatical sentences drove researchers to discard the faulty encoding accounts as being unable to explain language comprehension, no comparable revision arose in the interference literature. This is surprising since the evidence against number interference in grammatical sentences is literally the same as the evidence against attraction effects: the studies that failed to find the illusion of ungrammaticality predicted by the faulty encoding accounts also failed to find the inhibitory interference predicted by the interference accounts. The lack of concern is somewhat puzzling, but we believe that the explanation is simple: agreement attraction literature has been disconnected from the interference research, used different terminology, and null results in the attraction studies never came to the attention of the researchers interested in interference until Jäger et al. (2017b) systematically reviewed the existing literature from these two subfields.

4.0.3 The rationale for the proposed experiments

We demonstrated that grammatical sentences such as (1-a) and (1-b) are exactly where the predictions of the two groups of accounts diverge: the faulty encoding accounts predict an illusion of ungrammaticality, a slowdown in (1-b) as compared to (1-a), while the Lewis and Vasishth model predicts an inhibitory interference effect, a slowdown in (1-a) as compared to (1-b). The contradictory predictions do not necessarily imply a win-or-lose situation: the competing accounts assume different underlying mechanisms, both of which could be at play simultaneously. If both mechanisms are deployed, the effects might cancel each other out giving the impression that, on one hand, agreement attraction effects do not arise in comprehension of grammatical sentences (Lago et al., 2015; Wagers et al., 2009), and on the other hand, that interference effects do not arise in number (Nicenboim et al., 2018).

Indirect support for both mechanisms being deployed simultaneously comes from event-related potentials: Martin et al. (2014) found effects compatible with both the interference and agreement attraction accounts. During the processing of grammatical Spanish sentences with ellipsis, gender-matching interfering nouns lead to early anterior negativity reflecting difficulties in processing due to interference, but gender-mismatching interfering nouns lead to an increased P600 (also reported in Martin et al., 2012) indicating difficulties in syntactic processing predicted by the attraction accounts.

To test whether both interference and faulty encoding mechanisms affect reading times, we modify typical experimental materials to decrease the inhibitory interference and to allow the illusion of ungrammaticality to surface. The inhibitory interference predicted to arise in (1-a) consists of two components, number and semantic interference: the interfering noun *missile* shares the number marking of the verb and is a plausible theme of the verb. Semantic interference had been demonstrated in a series of studies (Van Dyke, 2007; Van Dyke & McElree, 2006, 2011): in grammatical sentences, non-subject nouns semantically matching the verb create inhibitory interference. In the meta-analysis, semantic interference was one of the most reliable effects consistent with the predictions of the Lewis and Vasishth model; it was estimated to lie within a 95% CrI between 1.7 and 28.1 ms, with a mean expected effect size of 13 ms (Jäger et al., 2017b).

We plan to capitalize on the well-established semantic interference component contributing to the interference effect: to decrease the overall interference in (1-a), we will eliminate the semantic interference component.[2] To do that, it should suffice to make the interfering noun inanimate and therefore semantically incompatible with the verb that requires an animate subject. When the interference, and hence processing

[2]Note that it's impossible to eliminate interference altogether as long as a number-matching interfering noun is present in the sentence.

slowdown in (1-a) is decreased, we should be able to observe the complementary slowdown in (1-b) predicted by the faulty encoding accounts.

The set of experimental conditions is presented in (4): (4-a) and (4-b) mirror the traditionally tested conditions where the interfering noun matches the verb in number and thematic requirements. If both interference and agreement attraction effects influence parsing, we expect equal reading times in these conditions.

(4) a. The admirer of the singer apparently thinks
    b. The admirer of the singers apparently thinks
    c. The admirer of the play apparently thinks
    d. The admirer of the plays apparently thinks
    ... the show was a big success.

In contrast, in conditions (4-c) and (4-d) the interfering noun is inanimate and does not meet the thematic requirements of the verb.[3] In that case, the Lewis and Vasishth model model predicts faster retrieval of the subject and faster reading times in (4-c) than in (4-a). The predictions of the faulty encoding accounts are not affected by this manipulation — equal amounts of attraction are expected in (4-b) and (4-d). Therefore, we expect a particular interaction: no difference between (4-a) and (4-b), and a slowdown due to the illusion of ungrammaticality in (4-d) in comparison to the control condition (4-c). Such interaction would demonstrate agreement attraction in grammatical sentences, and indicate that it was indeed masked by similarity-based interference.

If we observe the predicted interaction in the average reading times, our further goal is to test whether our explanation of an absence of both attraction and interference effects is supported by more fine-grained properties of the data. The Lewis and Vasishth model predicts a small slowdown in every reading time measurement in condition (4-a). In contrast, the faulty encoding accounts predict very high reading times (due to the illusion of ungrammaticality) in only a subset of trials in condition (4-b). Bayesian mixture modeling could help us determine whether the proportion of

[3]The reader might notice that the design of our experimental conditions is identical to the design of grammatical conditions reported in (Thornton & MacDonald, 2003). However, we cannot evaluate the results reported by Thornton and MacDonald as they do not report the interaction that is critical for our argument. Moreover, with eight experimental conditions and 24 participants, their experiment is likely to be underpowered.

extremely high reading times is greater in (4-b) than in (4-a).

4.1 Experiment 1

The hypotheses, number of participants, and analyses planned for Experiment 1 were pre-registered on OSF, doi:10.17605/OSF.IO/PD8KY.

4.1.1 Methods

Participants

Participants were recruited through the academic crowdsourcing platform Prolific and compensated for their time based on the recommended hourly rate of 6£ per hour. A compensation of 10 pence was offered for the task of reading and rating four sentences, which took approximately a minute. Inclusion criteria for participants were: (i) being a native speaker of English and (ii) being a resident of the US, UK, Ireland, New Zealand, or Australia.

Based on the power calculations (provided in the pre-registration) we estimated that 4,160 participants (65 independent observations per item per condition) would ensure a reasonable statistical power between 61% and 88%, depending on the effect size (ranging from 0.017 to 0.025 log milliseconds). In order to ensure that we acquire data from at least 4,160 participants who do not fall under the exclusion criteria, we collected data from 4,300 participants.

We excluded data from those participants who:

(i) admitted in a questionnaire following the experiment that English is not their native language or that they do not currently live in an English-speaking country;

(ii) gave exactly the same rating to the three practice sentences (two well-formed sentences and one sentence with an apparent agreement error);

(iii) had reading times for any word in the experimental sentence that fell below 180 ms or above 3,000 ms.

After applying the exclusion criteria, 4,296 participants entered the analysis.

Materials

We created 16 items similar to Example (4) in a 2×2 design manipulating the semantic and the number match/mismatch between the interfering noun and the verb (the subject noun always matched the verb). The subject was always singular and animate, while the properties of the interfering noun varied across conditions: it was singular in the number match conditions (a) and (c), plural in the number mismatch conditions (b) and (d); animate in the semantic match conditions (a) and (b), and inanimate in the semantic mismatch conditions (c) and (d). In the semantic match conditions both nouns were chosen such that they could potentially perform the action denoted by the verb. The interfering noun never referred to a multitude (such as *team*, *collective*, etc.). Within the sentence, the noun phrase was followed by an adverb and a verb with correct number marking, the same across all conditions. The verb was followed by a region that was the same across conditions and did not indicate the number of the head noun (no personal pronouns etc.).

Each item was followed by a comprehension question with five answer options, as in Example (5). The question rephrased the sentence and contained a verb marked for past simple tense. This way, the verb provided no information about the number of the head noun. The answer options were: the head noun in singular and plural forms, the interfering noun in singular and plural forms, and *I'm not sure*, presented in random order.

(5) Who considered the show a success? — Admirer/Admirers/(Singer/Singers or Play/Plays)/I'm not sure

The full set of experimental items and comprehension questions is presented in Appendix 6.2.

4.1.2 Item norming

To ensure that the semantic match/mismatch was actually perceived as such by native English speakers, we conducted a plausibility norming pretest. Based on each item, we created three sentences (see Example (6)) whose subjects were the head noun, the animate interfering noun, or the inanimate interfering noun of the original item. If the head noun was typically used with complements (as *admirer*, *opponent*, etc.) the whole noun phrase served as subject. All the nouns were singular.

We conducted two online questionnaires, both prompted participants to rate sentences on the Likert scale from 1 (bad, unnatural) to 7 (good, perfectly natural). In the first questionnaire, full sentences were presented; in the second questionnaire, sentences were truncated after the main verb (the truncated part is denoted by square brackets in (6)). We tested truncated sentences to ensure that the mismatch between the attractor noun and the verb was apparent right at the verb and not later in the sentence so that we could detect the effect at the verb.

(6) a. The admirer of the play supposedly thinks [the show was a big success].
  b. The singer supposedly thinks [the show was a big success].
  c. The play supposedly thinks [the show was a big success].

277 individuals took part in the pretests, 179 saw full sentences, and 98 other individuals saw truncated sentences. Each participant saw every item in one out of three conditions. The results of both norming studies confirmed that sentences with animate subjects ((6-a) and (6-b)) consistently received similarly high ratings (that is, we found no difference in ratings), while the sentences with inanimate subjects received lower ratings. We conclude that in the semantic match condition both the subject and the interfering noun are likely to perform the action denoted by the verb, and the interfering noun in the semantic mismatch condition is not. Further details on the pretests can be found in Appendix 6.2.4.

4.1.3 Procedure

The experiment was programmed using the Ibex[4] software and run on the IbexFarm cloud service. Each participant first saw the instructions, then three training sentences to get used to the non-cumulative centered self-paced reading procedure, and then one experimental sentence in one of the four conditions. Each participant saw only one experimental sentence — this way, participants could not get used to the manipulation and could not develop experiment-specific processing strategies. For each sentence including training ones, acceptability ratings on the scale from 1 (bad)

[4]http://spellout.net/ibexfarm.

to 7 (good) were collected to ensure that participants paid attention to the task, and to get an offline measure of the attraction effect. For the experimental sentence only, the acceptability rating task was followed by a comprehension question probing the final interpretation of the sentence.

4.1.4 Planned analyses

All analyses were conducted with the R system for statistical computing (R Development Core Team, 2009). Data were analyzed using generalized linear mixed models fit in the Bayesian framework (Vasishth et al., 2018) using the 'brms' package (Bürkner et al., 2017), which, in turn, relies on 'Stan' (Carpenter et al., 2017), a statistical system for full Bayesian inference. Plots were produced with the 'ggplot2' and 'tidybayes' packages (Kay, 2019; Wickham, 2016). Inferences were based on the posterior distributions of the parameters, which are reported in terms of the posterior mode and 95% percentile intervals (CrI). We used principled priors for the main effects and interactions ($Normal(0, 0.2)$). If nearly all of the posterior mass for an estimate fell on one side of zero, we considered that the effect was reliable. Every model included the main effects of number and semantic match/mismatch and their interaction, as well as random intercepts for items (but not for subjects, as only one observation comes from each subject) and by-item random slopes for the main effects and their interaction.

For reading times analysis, we assumed underlying lognormal distribution, and planned to analyze two regions: the critical verb and the region following the verb. For every experiment, we ran an exploratory analysis probing whether successful comprehension modulates the effects of interest: the models included, in addition to the previously specified structure, the main effect of trial accuracy and by-item random slopes for trial accuracy. For all reported experiments, the results of these additional analyses replicated those of the main analyses; we do not report the additional analyses in this chapter.

To analyze acceptability ratings, we use ordinal ordered logistic mixed-effects regression models. We opted to not model acceptability ratings on the linear scale as this could increase Type I and Type II errors, as well as lead to the inversion of the effects (Liddell & Kruschke, 2018).

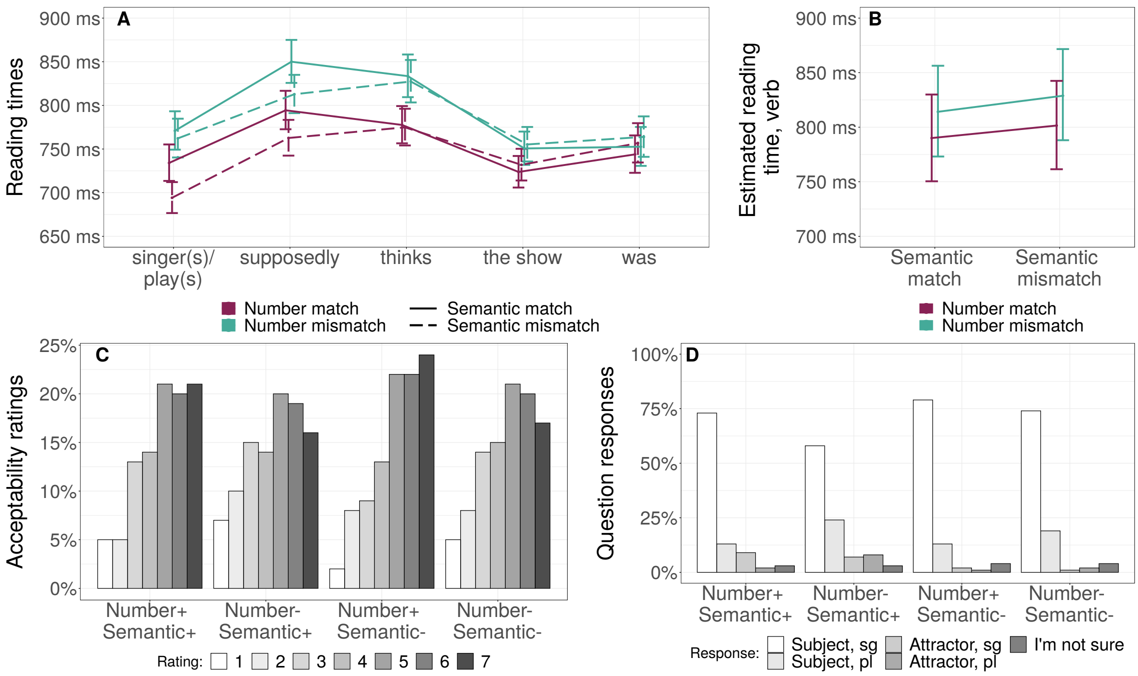

Figure 4.1: Results of Experiment 1. Panel A: geometric mean reading times across sentence regions. Panel B: Estimated reading times at the verb with 95% credible intervals (spillover from the previous region is accounted for in the modeling). Panel C: proportions of acceptability ratings across conditions. Panel D: proportions of question responses across conditions. In panels C and D, *Number+* stands for number match, *number-* for number mismatch; similarly, *semantic+* stands for semantic match, and *semantic-* for semantic mismatch.

Comprehension question responses had five answer options. In the descriptive statistics, we present proportions of responses of every category in each condition, but for statistical analysis, we simplify the data and code responses just as correct/incorrect. These binary coded responses were analyzed using mixed-effects linear models with a binomial link function. In the modeling of both acceptability ratings and comprehension question responses, we used principled priors for the main effects and interactions ($Normal(0, 0.3)$).

4.1.5 Results

Summaries of reading time, acceptability ratings, and question response accuracies are presented in Figure 4.1.

Planned analyses

Reading times. As can be seen from Figure 4.1A, we encountered an unexpectedly long-lasting plural complexity effect (inhibitory effect in the number mismatch conditions) that spanned for three more words following the plural interfering word: the adverb, the verb, and the region following the verb. This renders the planned comparison of reading times at the verb and at the region following the verb uninformative: the difference could be attributed to the plural complexity effect spilling over from the interfering noun, and not to the processing of the verb itself. We opted to use statistical control and describe the resulting analysis in the following section.

Acceptability ratings. Acceptability ratings were lower in the number mismatch condition (see Table 4.1). There was a tendency towards lower ratings for the semantic match conditions. The effects did not interact.

Table 4.1: Experiment 1. Statistical modeling of acceptability ratings.

Predictor	Estimate	95%-CrI	$P(\beta < 0)$
Intercept[1]	-3.10	-3.38 – -2.84	>0.999
Intercept[2]	-2.04	-2.30 – -1.79	>0.999
Intercept[3]	-1.14	-1.39 – -0.90	>0.999
Intercept[4]	-0.45	-0.69 – -0.21	>0.999
Intercept[5]	0.46	0.21 – 0.70	0.0005
Intercept[6]	1.50	1.24 – 1.74	<0.001
Number mismatch	-0.20	-0.31 – -0.09	>0.999
Semantic match	-0.11	-0.22 – 0.01	0.967
Number mismatch × Semantic match	0.02	-0.05 –0.10	0.244

Exploratory analyses

Reading times. As planned analyses of reading times were rendered void by the plural complexity effect, we corrected for the spillover effects by including reading times from the previous word as a predictor for reading times at the current word (Vasishth, 2006). This allows us to find out whether processing the current word introduces any additional difficulties over and above those inherited from the previous word. After applying this procedure, we found a slowdown in the number match conditions on the attractor noun itself, but not on the following adverb. At the verb, we found a main inhibitory effect of number mismatch: the verb was read slower in conditions with plural attractors (see Table 4.2 and Figure 4.1B; the slowdown comprised 26 ms, CrI:[0.30, 50] ms). This is precisely the slowdown predicted by the faulty encoding accounts.

Table 4.2: Experiment 1. Statistical modeling of reading times controlling for the reading times on the previous region.

Predictor	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$
Intercept	6.69	6.65–6.74	>0.999
Number mismatch	0.02	0.00–0.03	0.976
Semantic match	-0.01	-0.02–0.01	0.126
Previous region RT	0.26	0.23–0.29	>0.999
Number mismatch × Semantic match	0.00	-0.01–0.01	0.448

As we did not find evidence that agreement attraction in grammatical sentences is masked by interference, we did not run the pre-registered mixture-modeling analysis.

Question response accuracies. We found that both number mismatch and semantic match decreased the probability of giving a correct response (see Table 4.3). There was an interaction between the effects: nested comparisons demonstrated that the decrease in accuracy due to number mismatch was greater within the semantic match than within semantic mismatch conditions, (p($\beta > 0$) = 97.9%).

Table 4.3: Experiment 1. Statistical modeling of question response accuracies.

Predictor	Estimate (log-odds)	95%-CrI	$P(\beta < 0)$
Intercept	0.97	0.68 – 1.24	<0.001
Number mismatch	-0.24	-0.34 – -0.15	>0.999
Semantic match	-0.29	-0.42 – -0.16	>0.999
Number mismatch × Semantic match	-0.09	-0.17 – -0.01	0.982

4.1.6 Discussion

An unexpectedly prolonged plural complexity effect spanning three regions rendered the planned analyses of reading times uninformative. The scope of the effect is surprising as we used a typical design that takes the standard one-word spillover effects into account. Similar design was implemented, among others, in Wagers et al. (2009) and Lago et al. (2015), and prolonged plural complexity effects have never been reported. We suggest that the prolonged effect might stem from the single trial procedure: all effect sizes are likely to be bigger when participants do not adapt to the stimuli. We will return to this point in the General discussion.

The exploratory analysis of reading times mitigating the spillover effect supports the faulty encoding accounts: we found a slowdown at the verb in number mismatch conditions, which is precisely what the marking and morphing and feature percolation accounts predict. Acceptability judgments mirror reading times in demonstrating a clear decrease in ratings for the number mismatch conditions. We found no semantic interference effects and no interaction between number and semantic match conditions. This goes against our hypothesis that attraction would be detectable only in the semantic mismatch conditions, and concealed by number interference in the semantic match conditions. We will address the lack of semantic interference in the General discussion.

In the question response accuracies analysis, the main effects of number match and semantic mismatch can be dismissed as trivial: in each case, the number of potentially viable response options is lower than in the conditions they are contrasted with. That is, in number match conditions, two responses marked for plural are not viable as there were no words in the sentence marked for plural. Similarly, in the semantic mismatch conditions, the inanimate attractor cannot perform the action denoted by the verb, as established in the norming test. Therefore, participants simply have fewer options to choose from, which is sufficient to account for higher accuracy. However, the interaction cannot be dismissed on these grounds. In the number mismatch conditions, accuracy was lower in the semantic match than in the semantic mismatch conditions. This is compatible with semantic, but, importantly, not number interference. This outcome is in line with the results of several studies (Jäger, Benz, et al., 2015; Laurinavichyute et al., 2017; Mertzen et al., 2020) reporting interference effects in grammatical sentences only in question responses, but not in reading times measures. The common caveat with interpreting question response accuracies, however, is that they could reflect postinterpretative processing and the outcomes of reanalysis rather than the structure formed during processing of the verb (Bader & Meng, 2018).

To summarize, a slowdown (over and above the one spilling over from the previous regions) on the verb in the number mismatch conditions is compatible only with the faulty encoding accounts, but this conclusion could be compromised by the statistical correction for spillover effects. Similar outcome in an experiment without the spillover confound would be more convincing. To address the issue of the long-lasting plural complexity effect, we conducted Experiments 2 and 3. In Experiment 2, we retain the materials from Experiment 1 and introduce a long parenthetical phrase between the interfering noun and the verb. In Experiment 3, we employ sentences with object relative clauses, where the interfering noun is located further away from the verb and its subject both linearly and structurally. Faulty encoding accounts predict an illusion of ungrammaticality in prepositional phrases (Experiment 2), but not in the object relative clause setup (Experiment 3). The Lewis and Vasishth model model predicts inhibitory interference effects in both experiments irrespective of the syntactic structure. This means, in Experiment 2 we expect either to find the interaction we were originally testing for or to replicate the illusion of ungrammaticality that we found in Experiment 1. In Experiment 3 we expect to observe only the inhibitory interference predicted by the Lewis and Vasishth model.

4.2 Experiment 2

The hypotheses, number of participants, and analyses planned for Experiment 2 were pre-registered on OSF (doi:10.17605/OSF.IO/VM5BW).

4.2.1 Methods

Procedure and analysis are the same as in Experiment 1, except for the differences in the number of participants and experimental materials that are described below.

4.2.2 Participants

Participant recruitment and exclusion procedure followed that of Experiment 1. We recruited only individuals who did not take part in Experiment 1. 4,100 participants took part in the study. After applying exclusion criteria, 3,920 participants entered the analysis.[5]

4.2.3 Materials

Materials from Experiment 1 were modified such that within the sentence, the interfering noun and the verb were separated by a parenthetical phrase three to five words long (see Example (7)). The parenthetical contained either personal pronouns (I, you) or proper nouns (Daily Mail), but not common nouns in order to minimize additional interference. The parenthetical phrase was followed by an adverbial used in Experiment 1. In total, the buffer region between the interfering noun and the verb comprised four to six regions (4.5 on average), see (7):

(7) a. The admirer of the singer, according to the Daily Mail, apparently thinks
    b. The admirer of the singers, according to the Daily Mail, apparently thinks

[5]We first pre-registered N=1,956 based on the limit on available funding. We found no effects and decided to collect more data based on the power analysis (which suggested 3,900 samples) to be able to at least demonstrate the predicted interference effect.

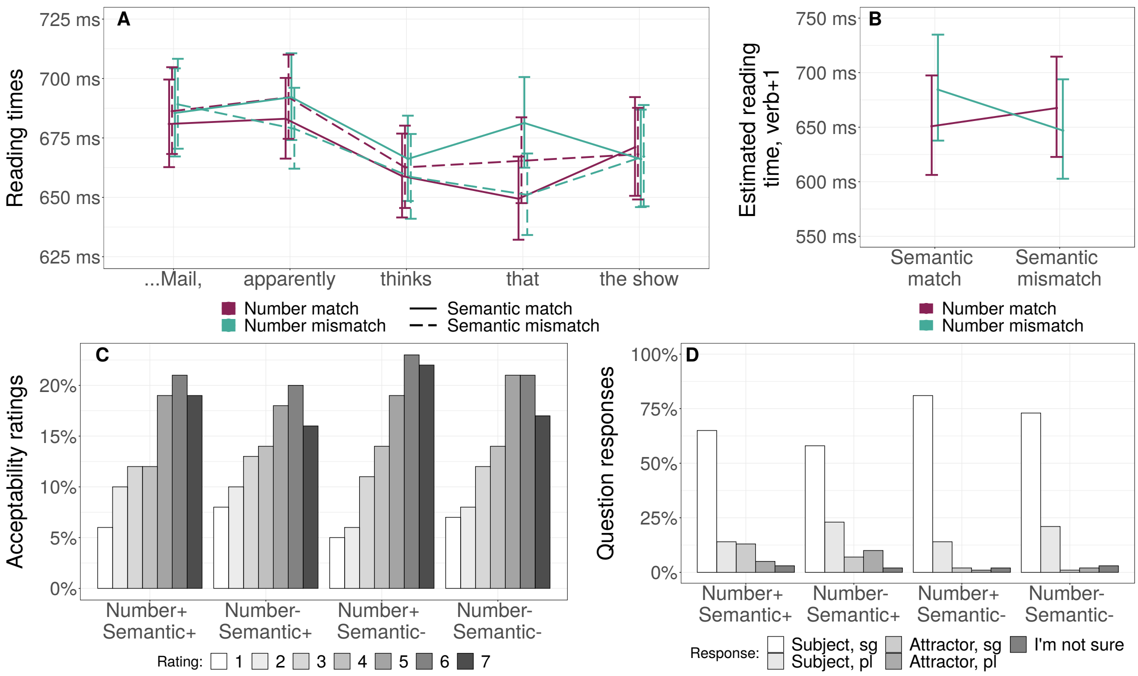

Figure 4.2: Results of Experiment 2. Panel A: geometric mean reading times across sentence regions. Panel B: Estimated reading times at the spillover after the verb (*that*) with 95% credible intervals. Panel C: acceptability ratings across conditions. Panel D: proportions of question responses across conditions. In panels C and D, *Number+* stands for number match, *number-* for number mismatch; similarly, *semantic+* stands for semantic match, and *semantic-* for semantic mismatch.

c. The admirer of the play, according to the Daily Mail, apparently thinks

d. The admirer of the plays, according to the Daily Mail, apparently thinks

... the show was a big success.

The full set of experimental items and comprehension questions is presented in Appendix 6.2.

4.2.4 Results

Summaries of reading times, acceptability ratings, and question response accuracies are presented in Figure 4.2.

Planned analyses

Reading times. Introducing parenthetical phrases successfully eliminated the plural complexity effect: in four regions preceding the critical verb, no main effects or interactions were detected, so we proceeded to the planned analyses (refer to Table 4.4).

On the region following the verb we found an interaction between the number and semantic match/mismatch. Nested comparisons showed that within semantic match conditions, number mismatch condition (b) was read more slowly than number match condition (a) (Est. = 33 ms, CrI: [4, 63] ms). There was no difference between the semantic mismatch conditions. As in Experiment 1, no interference effects at or after the verb were observed. As we did not find evidence that agreement attraction in grammatical sentences is masked by interference, we did not run the pre-registered mixture-modeling analysis.

Table 4.4: Experiment 2. Statistical modeling of reading times at the region following the verb.

Predictor	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$
Intercept	6.50	6.43 – 6.56	>0.999
Number mismatch	0.00	-0.01 – 0.02	0.746
Semantic match	0.01	-0.01 – 0.02	0.872
Number mismatch × Semantic match	0.02	0.01 – 0.04	0.995

Acceptability ratings. We found that semantic match conditions had lower acceptability ratings (see Table 4.5). There was a tendency for the sentences in the number mismatch condition to receive lower ratings as well. There was no interaction between the main effects.

Table 4.5: Experiment 2. Statistical modeling of acceptability ratings.

Predictor	Estimate	95%-CrI	$P(\beta < 0)$
Intercept[1]	-2.76	-3.07 – -2.45	>0.999
Intercept[2]	-1.76	-2.05 – -1.46	>0.999
Intercept[3]	-0.97	-1.25 – -0.67	>0.999
Intercept[4]	-0.32	-0.60 – -0.02	0.985

Intercept[5]	0.52	0.24 – 0.82	0.0007
Intercept[6]	1.65	1.37 – 1.95	<0.001
Number mismatch	-0.11	-0.23 – 0.01	0.963
Semantic match	-0.09	-0.19 – -0.00	0.975
Number mismatch × Semantic match	0.01	-0.05 – 0.08	0.334

Exploratory analyses

Question response accuracies. Mirroring the acceptability ratings results, both number mismatch and semantic match conditions decreased the probability of giving a correct response (see Table 4.6). There was no interaction between the main effects.

Table 4.6: Experiment 2. Statistical modeling of question response accuracies.

Predictor	Estimate (log-odds)	95%-CrI	$P(\beta < 0)$
Intercept	0.91	0.59 – 1.24	<0.001
Number mismatch	-0.23	-0.37 – -0.10	>0.999
Semantic match	-0.38	-0.51 – -0.25	>0.999
Number mismatch × Semantic match	0.03	-0.05 – 0.11	0.233

4.2.5 Discussion

We found an interaction between number and semantic match/mismatch conditions, but it went in an unexpected direction: a slowdown due to plural attractor arose in the semantic match conditions, i.e. the typical conditions extensively tested in the previous literature. To briefly remind the reader, we expected that in this condition the illusion of ungrammaticality would be masked by the inhibitory interference.

Unlike in Experiment 1, the plural complexity effect was detected only on the plural attractor region, and did not spill over to the following words. We will defer discussing the possible causes of such dramatic difference until after Experiment 3.

As in Experiment 1, we found that acceptability ratings tended to be lower in the number mismatch conditions. We also found semantic interference effect in acceptability ratings — ratings were lower for the semantic match conditions. In the comprehension question response accuracies, there was no interaction between the main effects. Given that both main effects are uninformative, nothing can be concluded from these results.

4.3 Experiment 3

The hypotheses, number of participants, and analyses planned for Experiment 3 were pre-registered on OSF together with Experiment 2 (doi:10.17605/OSF.IO/VM5BW). The motivation for the Experiment 3 was twofold: on the one hand, introducing more material between the attractor and the verb provided another way to mitigate the plural complexity effect found in Experiment 1 (such design was used in many previous studies: Lago et al., 2015; Wagers et al., 2009, to name just a few). On the other hand, it served to test the prediction of the faulty encoding accounts: no attraction effects are expected in the object relative clause configuration since the interfering noun is not a part of the subject noun phrase and the plural feature cannot percolate downwards the syntactic tree to the subject of the relative clause. If, in accordance with the predictions of the faulty encoding accounts, we do not observe any illusion of ungrammaticality, we should still observe the main effect of inhibitory interference predicted by the Lewis and Vasishth model. In fact, according to the extension of the model proposed in (Jäger et al., 2017b), interference effects might be even more pronounced in this configuration since the interfering noun is the subject of its own clause and therefore highly prominent.

4.3.1 Methods

Procedure and analysis were the same as in Experiment 1, except for the differences in the number of participants and experimental materials described below.

Participants

Participant recruitment and exclusion procedure followed that of Experiments 1 and 2. Participation in Experiment 3 was open, among others, for those who took part in the previous experiments, as the experimental materials were different and the experiments were separated by at least a week. 3,800 participants took part in the experiment. After applying exclusion criteria, 3,559 participants entered the analysis[6].

Materials

The 16 items from Experiment 1 were restructured to form sentences with object relative clauses[7], where the interfering noun is the head of the main clause, see (8):

(8) a. The singer that the actor openly admires, apparently
 b. The singers that the actor openly admires, apparently
 c. The play that the actor openly admires, apparently
 d. The plays that the actor openly admires, apparently
 ... received broad international recognition.

Within the sentence, the interfering noun was followed by an object relative clause containing the subject, an adverb, and a verb with correct (singular) number marking, the same across all conditions. The verb was followed by a region that did not differ across conditions and in no way indicated the number of the subject noun.

The comprehension questions from Experiment 1 were modified to match the sentences (see (9)).

(9) Who felt admiration? — Actor/Actors/(Play/Plays or Singer/Singers)/I'm not sure.

The full set of experimental items and comprehension questions is presented in

[6] We first pre-registered N=1,956 based on the limit on available funding. We found no effects and decided to collect more data based on the power analysis to be able to at least demonstrate the predicted interference effect.

[7] Seven items were originally misformed as sentences with possessive clauses. We corrected this mistake and collected data for the redesigned object relative clause sentences at a later point in time. This could have introduced some specific time-of-day or day-of-the-week effects during online data collection, but even if that were case, they should be absorbed by by-item random effects.

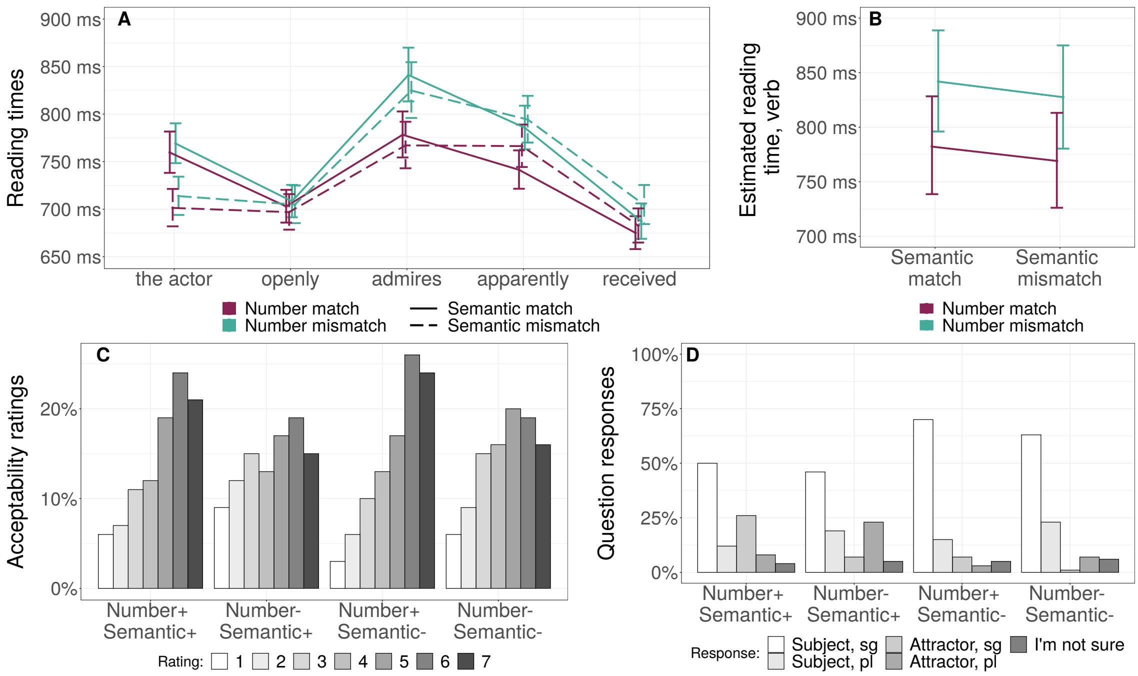

Figure 4.3: Results of Experiment 3. Panel A: geometric mean reading times across sentence regions. Panel B: Estimated reading times at the verb with 95% credible intervals. Panel C: acceptability ratings across conditions. Panel D: proportions of question responses across conditions. In panels C and D, *Number+* stands for number match, *number-* for number mismatch; similarly, *semantic+* stands for semantic match, and *semantic-* for semantic mismatch.

Appendix 6.2.

4.3.2 Results

Summaries of reading times, acceptability ratings, and question response accuracies are presented in Figure 4.3. Mean question response accuracy in Experiment 3 was lower than in previous experiments and comprised 57%. We do not think, however, that lower accuracy compromises the outcomes of the experiment. Firstly, participants were not guessing: with five response options, guessing would be represented by 20% accuracy or by majority of the "I'm not sure" responses. Secondly, recall that we a set of exploratory analyses adding trial accuracy as a predictor of reading times, and the results reported below hold in this additional analysis.

Planned analyses

Reading times. In the two regions preceding the verb, we found no main effect of number match, so we proceeded to the planned analysis. The results of statistical comparisons are presented in Table 4.7. At the verb, we found a main inhibitory effect of number mismatch: the verb was read slower in conditions with plural attractors (slowdown of 59 ms, CrI:[12, 105] ms). Again, no interference effects at or after the verb were observed. As we did not find evidence that agreement attraction in grammatical sentences is masked by interference, we did not run the pre-registered mixture-modeling analysis.

Table 4.7: Experiment 3. Statistical modeling of reading times at the verb region.

Predictor	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$
Intercept	6.69	6.65 – 6.73	>0.999
Number mismatch	0.04	0.01 – 0.07	0.993
Semantic match	0.01	-0.01 – 0.03	0.826
Number mismatch × Semantic match	0.00	-0.02 – 0.02	0.503

Acceptability ratings. Number mismatch conditions received lower ratings (see Table 4.8). There was a tendency towards lower ratings for the semantic match conditions. The effects did not interact.

Table 4.8: Experiment 3. Statistical modeling of acceptability ratings.

Predictor	Estimate	95%-CrI	$P(\beta < 0)$
Intercept[1]	-2.96	-3.31 – -2.58	>0.999
Intercept[2]	-1.89	-2.23 – -1.54	>0.999
Intercept[3]	-1.01	-1.35 – -0.66	>0.999
Intercept[4]	-0.34	-0.67 – 0.01	0.972
Intercept[5]	0.48	0.15 – 0.83	0.002
Intercept[6]	1.67	1.33 – 2.02	<0.001
Number mismatch	-0.27	-0.37 – -0.17	>0.999

Semantic match	-0.09	-0.19 – 0.01	0.962
Number mismatch × Semantic match	0.01	-0.05 – 0.08	0.336

Exploratory analyses

Question response accuracies. Mirroring the acceptability ratings, sentences in the number mismatch and in the semantic match conditions had lower probability of a correct response, but there was no interaction between the effects, see Table 4.9.

Table 4.9: Experiment 3. Statistical modeling of question response accuracies.

Predictor	Estimate (log-odds)	95%-CrI	$P(\beta < 0)$
Intercept	0.34	0.06 – 0.63	0.0095
Number mismatch	-0.13	-0.20 – -0.05	>0.999
Semantic match	-0.42	-0.57 – -0.27	>0.999
Number mismatch × Semantic match	0.03	-0.05 – 0.11	0.229

4.3.3 Discussion

Results of Experiment 3 represent a pronounced illusion of ungrammaticality in reading grammatical sentences: a lowdown in reading the number mismatch conditions right on the critical verb, not compromised by either preceding plural complexity effect or uninterpretable interaction. Interestingly, the faulty encoding accounts do not predict the illusion of ungrammaticality in the object relative clause configuration. We address the implications of this finding in the General discussion section, where it could be reviewed in the context of the findings of Experiments 1 and 2.

Acceptability judgments mirror reading times in demonstrating a clear decrease in ratings for the number mismatch conditions. Comprehension question response accuracies were lower than in Experiments 1 and 2, which reflects a well-established difficulty of processing object relative clauses (Gibson, 2000; Gordon et al., 2001). Crucially, in the comprehension question response accuracies, we found no interaction between the main effects. Given that both main effects are uninformative, nothing can be concluded from these results.

4.4 General discussion

The motivation for the three experiments presented here was to test whether parsing processes postulated by several faulty encoding (Bock & Eberhard, 1993a; Eberhard et al., 2005) and similarity-based interference accounts (Lewis & Vasishth, 2005; McElree, 2000) might be deployed simultaneously. If this is the case, the lack of both the predicted agreement attraction and number interference effects in grammatical sentences is due to these effects canceling each other out. However, this hypothesis received no support. Across the three experiments, reading times patterns were compatible only with the faulty encoding accounts: we found consistent slowdowns in reading grammatical sentences with plural interfering nouns. In Experiments 1 and 3, a main effect was detected on the critical verb (but in Experiment 1, we had to statistically correct for the spillover effect from the plural attractor), in Experiment 2, the slowdown was detected on the word following the critical verb within semantic match conditions. Acceptability ratings mirrored the illusion of ungrammaticality found in reading times: in Experiments 1 and 3, number mismatch conditions received lower ratings, in Experiment 2, there was a numerical tendency towards lower ratings. These results are in line with the predictions of the faulty encoding accounts, and we conclude that the illusion of ungrammaticality exists in grammatical sentences, and is not concealed by interference effect in the control condition.

4.4.1 Why didn't previous studies find these effects?

In such a clear-cut case, the illusion of ungrammaticality should have been repeatedly observed in the previous studies, whose experimental designs we closely followed. So why wasn't it the case? We suggest that several factors might have played a role, all tightly connected to the single-trial experimental procedure. In what follows, we entertain several possibilities, all of which must, however, remain a speculation until our results could be directly compared to a repeated-measures experiment using the same materials.

The first potential reason for the striking difference between the current and the previous results is that under the single-trial procedure a participant sees only one experimental sentence, and has no opportunity to adapt to the experimental manipulation. There is some evidence that adaptation decreases the effect size over the course of the experiment (Demberg & Sayeed, 2016; Fine et al., 2013). If this is true, the single-trial procedure allows us to detect the biggest possible effect size in each measurement we collect.

Another feature of our design is that in contrast to almost all other experiments (with an exception of Nicenboim et al., 2018), participants were not exposed to ungrammatical sentences, except for one sentence with an apparent agreement mistake in the training phase. We know that exposure to ungrammatical sentences over the course of the experiment shifts acceptability judgments (Hammerly et al., 2019). It is possible that not only the acceptability judgments, but also reading times are affected by repeated presentation of ungrammatical sentences: mistakes (including perceived mistakes, such as the illusion of ungrammaticality) may become less surprising over the course of the experiment and cause less slowdown. Participants' belief about the probability the upcoming structure being ungrammatical may get stronger, and as a result, participants might learn to rely less on agreement markers or even completely ignore these.

The last property of the single-trial procedure that might have enhanced attraction effects in our experiments is that we might have unintentionally encouraged participants to adhere to superficial processing mode (Ferreira et al., 2009; Karimi & Ferreira, 2016). Participants had to rate the acceptability of every sentence they saw, but a difficult comprehension question was asked only as the very last task in the experiment and may come as a surprise. Acceptability judgment is a relatively easy task that does not necessarily require full sentence parsing: judging whether a sentence is grammatical does not require one to fully parse it and resolve the dependencies, merely noticing that there are no apparent conflicts is enough. Repeatedly encountering acceptability judgment tasks might have set participants into good-enough superficial processing mode, which might be the key to the emergence of attraction effects. Superficial processing mode is more difficult to achieve in a repeated-measures experiment: it is possible only by avoiding comprehension question probes entirely, as repeated exposure to comprehension questions targeting the critical dependency would promote deeper processing (Swets et al., 2008).

Finally, another difference between our study and those that did not find attraction effects in grammatical sentences lies in the number of observations. With roughly 3,900 samples per experiment, we have around 975 samples for each of four experimental conditions. To collect 975 samples per condition in a typical repeated-measures experiment with 40 experimental items, at least 97 participants are needed. This exceeds the average number of participants in a typical experiment targeting either agreement attraction or interference effects (but note the recent increase in larger-sample studies, such as Avetisyan et al., 2020; Brehm et al., 2019; Jäger et al., 2020; Mertzen et al., 2020; Nicenboim et al., 2018). But even an equal number of probes in a repeated-measures design might not ensure the statistical power similar to that of the single-trial experiment if the effect size within a single participant diminishes over the course of the experiment.

Due to any of the outlined factors or to all of them combined, we found an illusion of ungrammaticality predicted by the faulty encoding accounts in every experiment. But in Experiment 1 it was masked by a surprisingly long-lasting plural complexity effect. Before discussing the implications of our results for the sentence processing theories, we want to briefly discuss what might have caused the effect.

4.4.2 The plural complexity effect

The plural complexity effects (slower reading times following the plural interfering noun) were never reported to exceed one region in the standard design of experimental materials that we used. The plural complexity effect is believed to arise due to several properties of the the plural word form itself, such as length and frequency, and due to the difficulty of meaning construction and semantic integration associated with prepositional phrases with singular head and plural dependent noun. But why was the effect so long-lasting? It spanned for at least two regions in Experiment 1 (the plural attractor and the buffer adverb). Again, we can only speculate, but we suggest that this prolonged effect might also be a consequence of the single-trial design. If all processing-related effects are magnified by the single-trial procedure, the side-effects would be affected, too. This would be useful to keep in mind when designing materials for single-trial experiments.

Notably, there was no plural complexity effect in Experiment 2: the slowdown was detected only on the plural interfering noun itself, but not on any of the following words. We suggest that it is the design of experimental materials — a parenthetical structure intervening between the attractor and the verb — that made Experiment 2 special. Dillon et al. (2017) claim that parenthetical phrases are processed independently of their embedding structures. Our results support their claim: when the parser processes the parenthetical structure, spillover effects from processing the embedding clause do not cross over to the parenthetical.

4.4.3 Attraction effects in grammatical sentences

A consistent illusion of ungrammaticality — a slowdown during reading the verb or the word following the verb in the grammatical sentences with a plural non-subject interfering noun — was found across three experiments. The illusion arose both in the structures where the interfering plural noun was part of the subject noun phrase and in the structures where it was not. The mere presence of the plural noun seems to be enough to cause the illusion. The lack of structure effects is inconsistent with the predictions of both the feature percolation and the marking and morphing accounts; they predict the illusion only if the interfering noun is a part of the subject noun phrase. Unlike the feature percolation, the marking and morphing account can in principle be extended to cover the observed effects (to cite Eberhard, Cutting, & Bock, 2005, p. 544):

> Because SAP [number information] may flow unobstructed throughout a structural network, number information bound anywhere within a structure has the potential to influence agreement processes. For this reason, even number information outside a subject or antecedent noun phrase (as in Hartsuiker et al., 2001) can affect agreement, to a degree

that is negatively correlated with its structural distance from the locus of agreement control.

In contradiction to this prediction, we have no evidence that greater structural distance decreases the magnitude of the attraction effect: the illusion of ungrammaticality is even bigger numerically in case of greater structural distance (59 ms in Experiment 3 vs. 25 ms and 34 ms in Experiments 1 and 2), but our data set might be insufficient for a precise comparison.

Our results seem to favor a far less intricate parsing system, similar to the Kahneman's System 2 that gets easily sidetracked by superficial properties of the sentence, such as any plural noun being potentially able to derail subject-verb agreement computation. This system should be activated probabilistically and/or under certain circumstances only, or normal language comprehension would turn out to be nearly impossible. One of the factors activating the system could be the good-enough or shallow processing mode. In this mode, sentences with number match should be read faster and rated higher on the acceptability scale than sentences with number mismatch: a sentence is definitely well-formed if it has two singular nouns and a singular verb, one does not need to complete subject-verb dependency to elicit that judgment. But when confronted with a comprehension question, participants should experience greater difficulties in the number match conditions, as they would need to build a precise representation of the sentence relying only on memory. Unfortunately, we cannot evaluate this proposal on our data set: while number match conditions are indeed read faster and receive higher acceptability ratings, they also have higher comprehension question accuracies, which seemingly contradicts the predicted difficulty in answering comprehension questions. The caveat is that the direct comparison of accuracies between the number match and number mismatch conditions is uninformative in our design: in the number match conditions, only three answer options out of five are viable (singular subject, singular attractor, "I'm not sure"), while in the number mismatch conditions, plural nouns should receive more consideration as potential responses, and accuracy might be lower just because there are more answer options to choose from.

However, the proposal we sketched creates a testable prediction: if we can encourage deep processing that requires building syntactic structure (for example, by asking difficult comprehension questions after each training sentence), we should no longer observe number attraction and might observe interference instead. In addition, under deep processing requirements, number match sentences should also receive lower ratings than their counterparts: when participants make an attempt at processing, number match sentences should be more difficult to process due to similarity-based interference.

Although the precise nature of the mechanism underlying attraction effects in grammatical sentences is unclear, our results persuasively demonstrate that agreement attraction effects cannot be reduced to repair of ungrammatical sentences not only in offline grammaticality judgments (Hammerly et al., 2019), but also in self-paced reading, which reflects more immediate processing. This poses a challenge for the similarity-base interference accounts, such as Lewis and Vasishth model: they need to be extended to cover attraction effects both in grammatical and ungrammatical sentences. One form this extension could take is a hybrid account that combines processes postulated by both the retrieval accounts and expectation-based accounts. The prerequisite for the emergence of attraction errors would be that expectation-based accounts would probabilistically make 'encoding errors' in the form of incorrect expectations (predicting plural verb after seeing a plural attractor). No such formal hybrid account currently exists, but the interplay between retrieval and prediction processes is being studied (Schoknecht et al., 2019).

Another account that might be able to cover attraction effects in grammatical sentences is lossy-context surprisal (Futrell et al., 2020; Futrell & Levy, 2017). It postulates that the processing cost of a word is defined by word's surprisal given a noisy representation of the preceding context. For the case of agreement attraction, the noisy representation of the subject and the attractor nouns' number marking can lead to probabilistic erroneous attribution of the plural number feature to the subject. If this happens, surprisal at the verb, and hence the reading times, will be greater than in the control condition, where erroneous number encoding is impossible. Whether the lossy-context surprisal account indeed predicts this slowdown, and whether it predicts any differences in effect sizes between various syntactic configurations of subject and attractor nouns, can only be confirmed via modeling.

4.4.4 Interference effects

Another outcome of our experiments, as important as the presence of the illusion of ungrammaticality in grammatical sentences, is the absence of interference effects, either semantic or morphosyntactic (number), in reading times. With roughly 3900 participants per experiment, we should have ~80% power to detect a 13-ms effect (a mean estimate for the interference effect in reading subject-verb non-agreement dependencies, e.g., semantic interference, obtained by Jäger et al., 2017b).

Number interference has already been proven difficult to observe in earlier studies (Jäger et al., 2017b; Nicenboim et al., 2018). As suggested by Wagers et al. (2009), lack of number interference effects could be explained by privative number marking. If only plural number feature is marked, while the singular is the default and has no explicit marking (as independently claimed in theoretic letrature, e.g. Harley & Ritter, 2002), then singular nouns cannot cause number interference. The lack of number marking on singular nouns would explain the absence of interference effects in all those agreement attraction studies that explored the processing of grammatical sentences with singular subjects and singular attractors. If we accept this explanation, the theoretical premise of our study renders itself incorrect: if singular nouns create no interference, then number interference cannot lead to slowdowns, and therefore, cannot mask the illusion of ungrammaticality. This is well in line with our findings, as we found no support for interference concealing the illusion of ungrammaticality across three higher powered experiments.

The lack of semantic interference, however, is not as easy to explain. Semantic interference effects in grammatical sentences are believed to be well-established, although several recent studies failed to detect the effect (Cunnings & Sturt, 2018; Mertzen et al., 2020). One potential explanation for the lack of the effect could be that interference effects arise as a function of processing depth: present when deep processing is encouraged and absent when shallow processing is sufficient. As stated earlier, we might have unintendedly encouraged shallow processing, which could conceal interference effects.

At the same time, both in acceptability ratings and question response accuracies, we detected some effects compatible with semantic interference. Semantic match conditions elicited lower ratings in Experiment 2, similar tendencies being present in

Experiments 1 and 3. In question response accuracies, we also found an interaction that was compatible with semantic, but not number interference (Experiment 1). Taken together, these findings suggest that semantic interference is not fully absent, but, crucially, is only detected in 'late' measures, which might reflect not the structure built during online processing, but rather a post-hoc interpretation (Bader & Meng, 2018). Under that assumption, late emergence of semantic interference also suggests that participants engaged in good-enough shallow processing during reading, and started building a full representation only when confronted with subsequent tasks.

4.5 Experiment 4

We set out to directly test the hypothesis that if participants engage in deeper processing, no more illusions of ungrammaticality, but rather inhibitory interference effects predicted by the Lewis and Vasishth model will be observed in reading times. Experiment 4 used the same experimental materials as Experiment 3, but aimed to induce deep processing strategies in participants by employing more complex training sentences.

4.5.1 Methods

Procedure and analysis were the same as in Experiment 3, except for the differences in the number of participants and experimental materials described below.

Participants

Participant recruitment and exclusion procedure followed that of Experiment 3. Participation was open, among others, for those who took part in the previous experiments, as the experiments were separated by at least several months. Due to high number of reading times above three seconds per word in the experimental items, we had to collect data from 4,576 participants to be able to use data from 3,702 individuals in the analysis. We report the analysis of the whole data set in the Exploratory analysis section.

Materials

We used the same experimental items as in Experiment 3, but the practice sentences were more complex: each sentence contained three animate nouns that could potentially perform the action denoted by the verb. The interfering nouns were embedded either in a subject-extracted or in an object-extracted relative clause. Each practice sentence was followed by a comprehension question with five response options. The practice sentences and their respective comprehension questions are presented in Examples (10) through (12):

(10) The priest who had privately advised the lawyer of the art dealer, is accused of withholding information.
Who was accused? — The priest/The lawyer/The art dealer/The art dealers/ I'm not sure.

(11) The personal assistant who the bodyguard of the delegate does not trust attracts great public attention.
Who attracted public attention? — The personal assistant/The bodyguard/ The delegate/The bodyguards/I'm not sure.

(12) The philanthropist who had greeted the secretary of the director, later participated in the fundraising committee.
Who took part in the committee? — The philanthropist/The secretary/The director/The secretaries/I'm not sure.

Note that in the examples, the correct answer is presented first, while in the experiment the order of response options was randomized. In contrast to Experiment 3, practice sentences were not followed by acceptability judgments. The experimental sentence was followed first by the comprehension question, and after that, by the acceptability judgment task.

4.5.2 Results

We first verified whether manipulating the difficulty of practice sentences did lead to deeper processing. Several metrics can be diagnostic of deeper processing: slower reading times, higher question response accuracies and lower ratings than in Experiment 3. The reading times on the experimental sentences were indeed slower in Experiment 4 than in Experiment 3 in the beginning of the sentence, which suggests that participants were affected by the depth-of-processing manipulation. Exclusion of as many as 874 participants who had reading times on some word in the experimental item exceeding three seconds also points in that direction. Ratings were also lower across the board in Experiment 4 (see Table 4.14). However, question response accuracies did not differ from those of Experiment 3 (55% vs. 57% in Experiment 3; for the results of statistical comparison, refer to Table 4.15). Mean question response accuracies for the three practice sentences comprised 53%, 55%, and 80%, respectively. The practice sentences were always presented in the same order, and increase in the proportion of correct responses suggests that participants got better during the practice. It is unclear why the accuracy of question responses in the experimental items was not higher than in Experiment 3. One possible explanation is that in the training sentences, the question always targeted the subject of the matrix clause, while in the experimental sentence, the question targeted the subject of the relative clause. Although data is somewhat contradictory, we suggest that slower reading times on the experimental item from the first word in the sentence as well as lower acceptability ratings indicate that participants at least tried to engage in deep processing.

Summaries of reading times, acceptability ratings, and question response accuracies are presented in Figure 4.4.

Planned analyses

Reading times. In the two regions preceding the verb, we found no main effect of number match, so we proceeded to the planned analysis. No main effects or interactions were detected at the verb or on the region following the verb.

Acceptability ratings. We observed no influence of experimental manipulations on the acceptability ratings (see Table 4.10).

Table 4.10: Experiment 4. Statistical modeling of acceptability ratings.

Predictor	Estimate	95%-CrI	$P(\beta < 0)$

Intercept[1]	-3.11	-3.39 – -2.83	>0.999
Intercept[2]	-1.72	-1.96 – -1.47	>0.999
Intercept[3]	-0.72	-0.95 – -0.48	>0.999
Intercept[4]	0.12	-0.12 – 0.37	0.151
Intercept[5]	1.26	1.03 – 1.51	<0.001
Intercept[6]	2.69	2.42 – 2.95	<0.001
Number mismatch	-0.04	-0.13 – 0.06	0.797
Semantic match	0.04	-0.10 – 0.18	0.297
Number mismatch × Semantic match	0.00	-0.07 – 0.07	0.458

Exploratory analyses

Reading times. As we pre-registered the analysis of RTs only on the critical region and the following region, we report analyses of reading times on other regions in this section. On the second region following the verb, we observed a main facilitatory effect of number mismatch (speedup of -23 ms, CrI:[-48, 0.55] ms, see also Table 4.11). This speedup contradicts the predictions of the faulty encoding accounts, and is in line with the predictions of the similarity-based interference accounts.

Table 4.11: Experiment 4. Statistical modeling of reading times at the second region after the verb.

Predictor	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$
Intercept	6.58	6.55 – 6.62	>0.999
Number mismatch	-0.016	-0.033 – 0.00	0.0268
Semantic match	0.009	-0.040 – 0.022	0.278
Number mismatch × Semantic match	0.011	-0.005 – 0.027	0.914

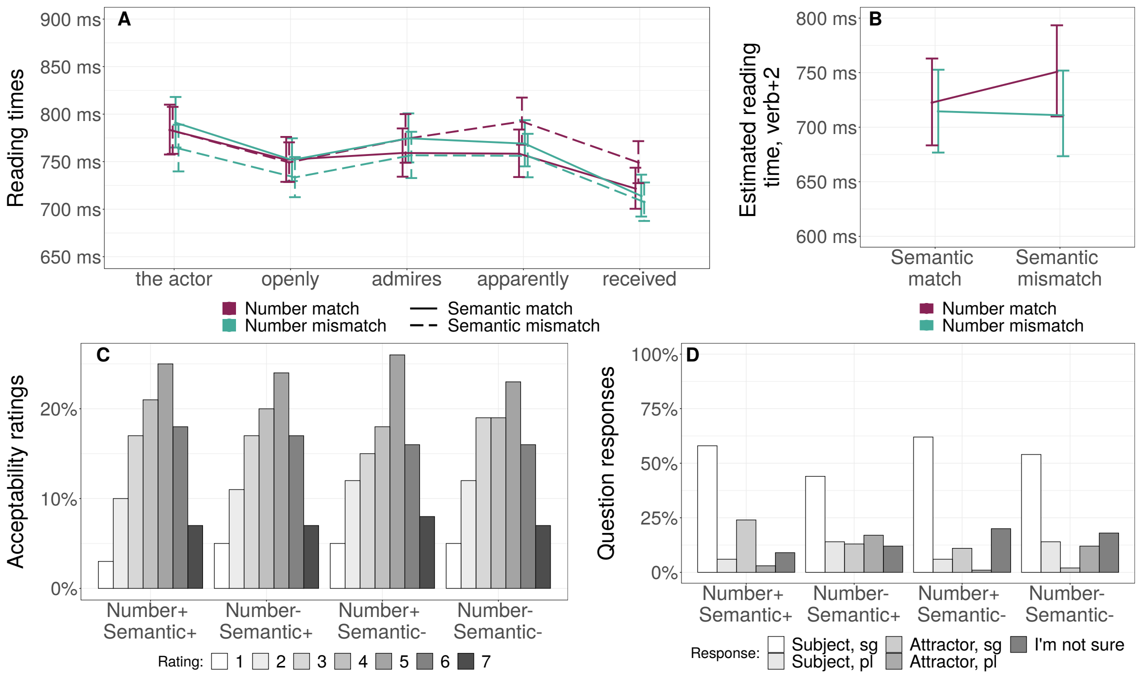

Figure 4.4: Results of Experiment 4. Panel A: geometric mean reading times across sentence regions. Panel B: Estimated reading times at the second region after the verb (*received*) with 95% credible intervals. Panel C: acceptability ratings across conditions. Panel D: proportions of question responses across conditions. In panels C and D, *Number+* stands for number match, *number-* for number mismatch; similarly, *semantic+* stands for semantic match, and *semantic-* for semantic mismatch.

Question response accuracies. Sentences in the number mismatch conditions had lower probability of a correct response, see Table 4.12.

Table 4.12: Experiment 4. Statistical modeling of question response accuracies.

Predictor	Estimate (log-odds)	95%-CrI	$P(\beta < 0)$
Intercept	-0.06	-1.10 – 0.95	0.532
Number mismatch	-0.28	-0.38 – -0.19	>0.999
Semantic match	-0.11	-0.46 – 0.23	0.755
Number mismatch × Semantic match	-0.04	-0.13 – 0.04	0.830

Analysis of the whole data set from Experiment 4

In this analysis, we still excluded data from self-reported non-native speakers and from participants who read any word in the experimental item faster than for 180 ms, but retained data from those participants who read any word in the experimental sentence longer than three seconds (resulting N=4,633). On the critical verb, we observed some indication of an interaction between number mismatch and semantic match conditions (Est.=0.02 log-ms, CrI:[-0.002, 0.04], $P(\beta > 0) = 0.964$). Nested comparisons showed that the interaction is likely driven by a slowdown in the semantic match (vs. mismatch) within the number mismatch conditions (53 ms, CrI:[-3.25, 105] ms). Average question response accuracy in this data set comprised 56%, which suggests that longer reading times do not necessarily result in more accurate processing.

Analysis of pooled data from Experiments 3 and 4

To be able to claim that deep processing blocks the illusion of ungrammaticality, we need to directly test the interaction between processing depth and the number match/mismatch conditions. To do that, we analyzed the pooled data set from Experiments 3 and 4; the processing depth in Experiment 3 with assumed superficial processing was coded as -1, in Experiment 4 with induced deep processing, as 1. The model included the interaction between the number and semantic match/mismatch conditions and the interaction between the number match/mismatch condition and the processing depth, as well as all the main effects.[8] The random effects structure included random intercepts for items as well as by-item random slopes for all the main effects and interactions.

Reading times. At the region of the critical verb, and in the two following regions, we observed an interaction between the number match condition and the processing depth (see Table 4.16). The nested comparisons showed that at the verb and the following region, the interaction was driven by the slowdown in number mismatch conditions in the superficial processing mode (the verb: 59 ms, CrI:[15, 103] ms; the

[8]Since we did not observe any interaction between the number and the semantic match/mismatch conditions in either of the experiments, we did not test whether this interaction depends on processing mode, i.e. did not include the three-way interaction. In case the reader is wondering, all results hold if we include the three-way interaction.

following region: 34 ms, CrI:[9, 59] ms). At the next region, nested comparisons showed the opposite effect: a speedup in number mismatch conditions in the deep processing mode (-25 ms, CrI:[-49, -2] ms).[9]

We additionally conducted a Bayes factor analysis to quantify the evidence in favor of the interaction between number match/mismatch conditions and processing depth as well as in favor of the effects in nested comparisons. Bayes factor quantifies how much more likely is the model that includes the predictor in question to have generated the data as compared to the model that does not include it. As Bayes factor is sensitive to priors, we computed Bayes factors for a small range of plausible priors: the regularizing priors ($Normal(0, 0.3)$), two increasingly more informative priors ($Normal(0, 0.1)$, $Normal(0, 0.01)$), and an even wider regularizing prior ($Normal(0, 1)$). For each model, we ran four chains with 20000 iterations each, the first 2000 samples were discarded as warm-up samples. The resulting Bayes factor values can be seen in Table 4.13.

Table 4.13: Analysis of pooled reading times from Experiments 3 and 4. Bayes factor values quantify evidence in favor of the presence of the effect. Slowdown and speedup refer to the effect observed in the nested comparison between number mismatch and number match conditions.

	Verb		**Verb+1**		**Verb+2**	
Prior SD	Interaction	Slowdown	Interaction	Slowdown	Interaction	Speedup
0.01	6.09	1.26	5.71	2.06	4.89	1.56
0.1	2.12	5.85	1.64	5.16	1.09	1.56
0.3	0.72	2.56	0.55	1.85	0.38	0.56
1	0.43	1.55	0.33	1.16	0.22	0.34

Informative priors that best correspond to the scale of the observed effects ($Normal(0, 0.01)$ for the smaller interaction effects, $Normal(0, 0.1)$ for the bigger

[9] Note that since more information on item-level variability is available in this pooled analysis, the estimated credible intervals for the effects got slightly tighter, and we even detect a slowdown on the spillover after the verb that we did not detect in the separate analysis of Experiment 3.

effects in nested comparisons) provide moderate support for all tested effects except the speedup in the number mismatch conditions in the second region following the verb. Nothing can be concluded with respect to this effect. With wider priors, evidence for all evaluated effects becomes anecdotal and inconclusive, as wider priors generally favor the null model.

We additionally analyzed the pooled question response accuracies and sentence acceptability ratings from both experiments. In these models, we included the three-way interaction, as well as all possible two-way interactions and main effects. The random effects structure included random intercepts for items as well as by-item random slopes for all the main effects and interactions.

Acceptability ratings. Number mismatch conditions received lower ratings across the board (see Table 4.14). We also observed a main effect of processing depth: the same experimental sentences received lower ratings in the deep processing condition, i.e. when preceded by complex training sentences. There was an interaction between number match/mismatch and processing depth: within the deep processing condition, number mismatch conditions received higher ratings, i.e. the general decrease in ratings due to number mismatch was much less pronounced under deep processing. We interpret this as an indication that participants did not experience the illusion of ungrammaticality as much as in the shallow processing condition. Finally, there was some indication for an interaction between semantic match/mismatch and deep processing mode: decrease in ratings due to deep processing tended to be smaller within semantic match conditions.

Table 4.14: Statistical modeling of acceptability ratings on the data pooled from Experiments 3 and 4.

Predictor	Estimate	95%-CrI	$P(\beta < 0)$
Intercept[1]	-3.06	-3.33 – -2.78	>0.999
Intercept[2]	-1.83	-2.08 – -1.57	>0.999
Intercept[3]	-0.88	-1.13 – -0.62	>0.999
Intercept[4]	-0.11	-0.37 – 0.15	0.816
Intercept[5]	0.87	0.62 – 1.13	<0.001

Intercept[6]	2.14	1.88 – 2.39	<0.001
Number mismatch	-0.16	-0.24 – -0.09	>0.999
Semantic match	-0.03	-0.11 – 0.06	0.757
Deep processing	-0.28	-0.47 – -0.07	0.993
Number mismatch × Semantic match	0.01	-0.04 – 0.05	0.392
Number mismatch × Deep processing	0.14	0.07 – 0.20	<0.001
Semantic match × Deep processing	0.08	-0.01 – 0.16	0.033
Number mismatch × Semantic match × Deep processing	-0.01	-0.06 – 0.04	0.65

Question response accuracies. As expected, probability of a correct response was lower in the number mismatch an in semantic match conditions (see Table 4.15), but as we have already discussed, these effects are trivial and cannot be interpreted. Interestingly, deep processing mode further decreased the probability of a correct response in number mismatch conditions, but increased it in the semantic match conditions.

Table 4.15: Statistical modeling of question response accuracies on the data pooled from Experiments 3 and 4.

Predictor	Estimate (log-odds)	95%-CrI	$P(\beta < 0)$
Intercept	0.18	-0.28 – 0.60	0.21
Number mismatch	-0.20	-0.27 – -0.15	>0.999
Semantic match	-0.27	-0.43 – -0.10	0.998
Deep processing	-0.11	-0.40 – 0.19	0.764
Number mismatch × Semantic match	-0.01	-0.08 – 0.06	0.573

Number mismatch × Deep processing	-0.07	-0.13 – -0.00	0.98
Semantic match × Deep processing	0.14	0.01 – 0.27	0.019
Number mismatch × Semantic match × Deep processing	-0.04	-0.10 – 0.01	0.934

Table 4.16: Statistical modeling of pooled data from experiments 3 and 4.

	Verb			**Region following the verb**			**Second region following the verb**		
Predictor	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$	Estimate (log-ms)	95%-CrI	$P(\beta > 0)$
Intercept	6.66	6.64 – 6.70	>0.999	6.649	6.60 – 6.70	>0.999	6.558	6.53 – 6.59	>0.999
Number mismatch	0.017	-0.003 – 0.038	0.95	0.006	-0.005 – 0.018	0.86	-0.004	-0.017 – 0.009	0.28
Semantic match	0.004	-0.009 – 0.017	0.74	-0.008	-0.021 – 0.005	0.097	-0.009	-0.025 – 0.006	0.12
Deep processing	-0.024	-0.053 – 0.006	0.06	-0.004	-0.030 – 0.023	0.38	0.025	0.008 – 0.042	0.995
Number mismatch × Semantic match	0.007	-0.006 – 0.019	0.86	0.010	-0.002 – 0.022	0.95	0.003	-0.009 – 0.016	0.71
Number mismatch × Deep processing	-0.020	-0.035 – -0.005	0.0052	-0.016	-0.027 – -0.003	0.0065	-0.014	-0.026 – -0.003	0.0068

4.5.3 Discussion

Experiment 4 demonstrates that the illusion of ungrammaticality can be switched off if participants engage in deep processing. This finding can potentially shed light on why the illusion was so rarely observed in previous studies and consistently found in Experiments 1 through 3 reported in this chapter. At the same time, this outcome is difficult to reconcile with the faulty encoding accounts: both accounts postulate that the illusion of ungrammaticality arises due to probabilistic errors in normal computation of number assignment. It is unclear why number assignment, an automated process that participants have no conscious control of, should be led astray less frequently when participants pay more attention to the linguistic input. According to both the feature percolation and the marking and morphing accounts, the illusion of ungrammaticality has nothing to do with participants being unsure of which particular noun has plural marking or not being able to assemble syntactic structure, so deeper processing should not play any role in agreement attraction. We suggest that our findings are more compatible with a simple heuristic tracking the instances of plural features, a heuristic that might kick in when parsing is not the main priority.

The hypothesis that interference effects should surface in the deep processing mode did not receive full support: there was no indication of semantic interference in reading times, but we observed a slowdown in the number match conditions, which is consistent with the predictions of the Lewis and Vasishth model. Acceptability ratings for the number match conditions were also lower in deep than in shallow processing conditions, which supports the proposal that number match between the subject and the interfering noun causes interference in the deep processing mode. However, the slowdown in reading times appeared quite late, on the second region after the critical verb, and was not supported by the Bayes factor analysis. It is possible that inhibitory interference effects occur relatively late, or that this slowdown is due to inhibitory interference arising during processing the matrix clause verb, which was exactly the second region after the critical verb in the majority of items. To conclude, the absence of predicted semantic interference, rather late manifestation of number interference, and the lack of conclusive evidence in favor of number interference effect in reading times all suggest that the slowdown in number match condition should be interpreted with caution.

We can offer no explanation for the absence of semantic interference at present. In the question response accuracies, semantic interference seems even to be diminished in the deep processing condition: decrease in accuracy in the semantic match conditions was much less pronounced in Experiment 4 than in Experiment 3. We can only say that our results mirror the recent failed replication of semantic interference in three languages reported in Mertzen et al. (2020).

4.6 Conclusion

This chapter demonstrates the illusion of ungrammaticality, and therefore, agreement attraction effects, in reading grammatical sentences across three experiments. The consistent presence of the effect suggests that the predictions of the faulty encoding accounts — the feature percolation and the marking and morphing — are applicable not only to production, but to sentence comprehension as well. The slowdown caused by the illusion of ungrammaticality is exactly the opposite of the slowdown that the similarity-based interference accounts predict due to number interference. This might pose a problem for the similarity-based interference accounts, however, we further show that the illusion of ungrammaticality arises only in the superficial processing mode; in the deep processing mode, a (delayed) slowdown consistent with inhibitory interference is observed. But our results still pose a challenge to the similarity-based interference accounts: we observed no semantic interference in reading times in any of the four experiments, even when deep processing was encouraged. At present we have no explanation to offer for the lack of semantic inhibitory interference.

Chapter 5

General discussion and conclusions

We will first briefly summarize the results of the studies comprising this dissertation and then discuss the implications of our findings.

The aim of the experiments presented in Chapter 2 was to disentangle encoding and retrieval similarity-based interference. In Experiment 1, no effects in reading times consistent with the predictions of any similarity-based interference account were found. In Experiments 2A (only in those participants who answered most of the comprehension questions accurately) and 2B, we found effects consistent with encoding, but not retrieval interference: gender-unmarked reflexives were read slower when the interfering noun shared the gender of the antecedent. Given the combined evidence from our findings and the lack of support for retrieval interference in a recent larger-sample replication of the Van Dyke and McElree study reported in (Mertzen et al., 2020), it is at present not clear whether similarity-based interference indeed originates at the stage of memory retrieval, as proposed by the Lewis and Vasishth model and the direct access model by McElree. We do not contest, however, that similarity-based interference affects language processing, if only under limited circumstances. In particular, effects consistent with the facilitatory interference predicted by the similarity-based interference accounts were overwhelmingly found in processing ill-formed sentences.

In Chapter 3, we directly compared morphosyntactic and semantic facilitatory interference effects in ill-formed sentences. The similarity-based interference accounts predict the effects to be of the same magnitude, as they are driven by the same processing mechanism. In contrast, the faulty encoding accounts that provide an alternative explanation for morphosyntactic facilitatory interference effects, do not predict semantic attraction effects (that is, facilitatory semantic interference). The faulty encoding accounts would be supported if we found only morphosyntactic attraction, or if morphosyntactic attraction effect was greater than semantic attraction. However, across three experiments, we found that the morphosyntactic and semantic attraction (facilitatory interference) effects were similar in size, both in the analysis of acceptability judgments and reaction times. We conclude that for the processing of ill-formed sentences, similarity-based interference is necessary and sufficient to explain attraction-like effects both in morphosyntactic and in semantic domains.

Chapter 4 explores whether the same is true for processing well-formed sentences. The inhibitory effect predicted by similarity-based interference accounts is much more elusive than the facilitatory effect predicted in ill-formed sentences (Jäger et al., 2017b; Jäger et al., 2020; Mertzen et al., 2020). At the same time, the slowdown complementary to the inhibitory interference effect predicted by the faulty encoding accounts is also observed only rarely (Cunnings & Sturt, 2018; Jäger et al., 2017b; Lago et al., 2015; Nicol et al., 1997; Patson & Husband, 2016; Thornton & MacDonald, 2003; Tucker et al., 2015; Wagers et al., 2009). We tested whether the absence of both predicted effects can be explained by both effects being present at the same time and canceling each other out. This turned out to not be the case. Across three experiments, we found no indication of the inhibitory interference effect predicted by the similarity-based interference accounts. On the contrary, we observed the illusion of ungrammaticality partially consistent with the predictions of the faulty encoding accounts. Partially, because the illusion was also observed in object relative clause configuration where it is not predicted to appear.

To condense the outcomes even more, we found that to explain how ill-formed sentences are processed, the predictions of the Lewis and Vasishth model, but not those of the faulty encoding accounts, are sufficient. But neither account can fully explain the processing of well-formed sentences. Predictions of the similarity-based interference accounts were partially supported in experiments reported in Chapter 2, but not supported in three experiments reported in Chapter 4. Instead, in the first three experiments reported in Chapter 4, we observed only the reverse effects consistent with the broad predictions of the faulty encoding accounts.

5.1 Implications

Our findings pose a challenge to the similarity-based interference accounts since they aim to cover language processing in all syntactic configurations. At first sight, it seems that whether the predictions of the interference accounts are fulfilled or not, depends on input well-formedness: consistent support for similarity-based interference was found only in the processing of ungrammatical sentences (Chapter 3). In the processing of grammatical sentences, partial support was found in Chapter 2. Note that despite appearances, there is no inherent conflict between the outcomes of experiments reported in Chapters 2 and 4. In Chapter 2, effects consistent with inhibitory interference were found in participants who performed many experimental trials (in one of the experiments, only in the subset of accurate participants). In Chapter 4, each participant saw only one experimental probe, and we have no way of knowing whether the same pattern as in Chapter 2 would emerge.

We would like to briefly remind the reader that the pattern of the results we observed is consistent with previous findings: facilitatory effects predicted by the interference accounts are universally found in ungrammatical sentences, but inhibitory effects predicted to arise in grammatical sentences are much harder to detect (Jäger et al., 2017b; Jäger et al., 2020; Mertzen et al., 2020).

From the point of view of the Lewis and Vasishth model, the dichotomy between interference effects in the ill- and well-formed sentences is surprising: the Lewis and Vasishth model does not postulate that interference effect size should depend on whether the effect arises during the processing of well- vs. ill-formed structures. The model does not even have a way to identify the structure as well- or ill-formed. And yet sentence well-formedness seems to matter to human participants. Interestingly, the direct access model by McElree assumes that ill-formedness of the mentally assembled structure is detected, although the mechanism enabling this detection is not specified. Detection of ill-formedness triggers reanalysis, i.e., the second attempt at retrieval, the hallmark feature of the model. However, the model does not specify what happens when the input itself is ill-formed: whether several additional retrieval attempts are executed, or parsing fails after a time-out. For that reason, we cannot assess how the model fits the observed data.

One way to align the predictions of the Lewis and Vasishth model with the observed data would be to assume that retrieval from memory, and the similarity-based interference effects associated with it, arise during reanalysis, after sentence ill-formedness had been detected by some mechanism external to the model, as proposed by Wagers et al. (2009) and McElree (2000). This proposal is supported by the conclusions of Lago et al. (2015) who report that in ungrammatical sentences, facilitatory interference (i.e. agreement attraction) effects are observed only following the detection of ungrammaticality.

Recall, however, that we also observed inhibitory interference effects (although not in all configurations where they were predicted) in grammatical sentences in two experiments reported in Chapter 2, which suggests that retrieval from memory and the similarity-based interference associated with it cannot be reduced to the processing of ill-formed structures. Recent proposals by Stone et al. (2020) and Schoknecht et al. (2019) suggest that similarity-based interference exists in a tight interplay with prediction and might be only deployed when prediction fails. In a similar vein, Nicenboim et al. (2016) and Mertzen et al. (2020) suggest that interference might arise only in people with high working memory capacity or under deep processing mode. Following these proposals, we hypothesized that interference effects might depend on the depth of processing: retrieval from memory is initiated when participants are engaged in deep processing (either following their internal intention, or when their predictions about the upcoming structure are violated, as in, but not limited to, ungrammatical structures).

Indirect support for this hypothesis comes from the fact that effects compatible with similarity-based interference are regularly observed in the measures of comprehension. In this dissertation as well, despite having found no or limited inhibitory interference effects in reading times, we still found compatible effects in comprehension question responses (lower accuracy in the gender-match conditions in Chapter 2) and acceptability judgments (lower ratings in the number-match and semantic-match conditions in Chapter 4).

What is the alternative to deep processing? Ferreira et al. (2009) and Karimi and Ferreira (2016) introduced the concept of shallow or good-enough processing, which can be seen as parsing in the default functioning mode that sustains basic comprehension when comprehension is not the main priority. If only deep parsing can give rise to interference effects, maybe only the superficial good-enough processing mode can give rise to agreement attraction effects, given that we observed the illusion of ungrammaticality under conditions inviting shallow parsing. Under the assumption that processing depth determines, which parsing mechanism is deployed, similarity-based interference and agreement attraction effects, which are both predicted to arise in grammatical sentences, cannot arise simultaneously. The effects should be mutually exclusive, not because the predictions of only one theory are correct, but because only one mechanism can be deployed at one point in time.

We directly tested this hypothesis in Experiment 4 reported in Chapter 4 by using the experimental materials of Experiment 3 and manipulating depth of processing: we asked participants difficult comprehension questions during the training phase so that they would engage in deep processing by the time they encounter the experimental sentence. The results are mixed: we found that the illusion of ungrammaticality predicted by the faulty encoding accounts indeed disappears in the deep processing mode. Moreover, on the second region following the critical verb we detect the slowdown in the number match conditions consistent with the predictions of the similarity-based accounts. At the same time, even in the deep processing mode, we find no semantic interference.

Independently of whether similarity-based interference arises under limited circumstances or not, our results suggest that the faulty encoding accounts cannot be dismissed since they make unique predictions with regard to processing grammatical sentences, which are not shared by any other account. At the same time, the predictions of the faulty encoding accounts were also not fully supported as illusions of ungrammaticality arose in syntactic structures where they were not predicted to occur, and only in the superficial processing mode. Our results might therefore favor a much simpler system that superficially tracks the number features on nouns preceding the verb and is distracted by every plural feature. Whether such system or a more elaborate faulty encoding account better fits the human data, still needs to be evaluated.

Bibliography

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in spanish. *Lingua*, *143*, 108–128.

Anderson, J. R. (1996). Act: A simple theory of complex cognition. *American psychologist*, *51*(4), 355.

Anderson, J. R. (2014). *Rules of the mind*. Psychology Press.

Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought lawrence erlbaum. *Mathway, NJ*.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In *Memory* (pp. 237–313). Elsevier.

Antón-Méndez, I., Nicol, J., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, *5*(1), 1–25.

Arehalli, S., & Linzen, T. (2020). Neural language models capture some, but not all, agreement attraction effects.

Autry, K. S., & Levine, W. H. (2014). A fan effect in anaphor processing: Effects of multiple distractors. *Frontiers in Psychology*, *5*, 818.

Avetisyan, S., Lago, S., & Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension? *Journal of Memory and Language*, *112*, 104087.

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in slovak. *Journal of Memory and Language*, *56*(1), 65–85.

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

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

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumann, P., & Yoshida, M. (2015). A psycholinguist asking who binds himself: Interference effects in the processing of reflexives. *28th Annual CUNY Conference on Human Sentence Processing (Los Angeles, CA: University of Southern California)*, 162.

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of memory and language*, *31*(1), 99–127.

Bock, K., & Eberhard, K. M. (1993a). Meaning, sound and syntax in english number agreement. *Language and Cognitive Processes*, *8*(1), 57–99.

Bock, K., & Eberhard, K. M. (1993b). Meaning, sound and syntax in english number agreement. *Language and Cognitive Processes*, *8*(1), 57–99. https://doi.org/10.1080/01690969308406949

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, *43*, 83–128.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93. https://doi.org/10.1016/0010-0285(91)90003-7

Boyce, V., Futrell, R., & Levy, R. P. (2019). Maze made easy: Better and easier measures of incremental processing difficulty.

Brasoveanu, A., & Dotlačil, J. (2019). Formal linguistics and cognitive architecture. *Language, Cognition, and Mind (LCAM) Series. The pyactr library (Python3 ACT-R) is available here: https://github. com/jakdot/pyactr. Dordrecht: Springer.*

Brehm, L., Jackson, C. N., & Miller, K. L. (2019). Speaker-specific processing of anomalous utterances. *Quarterly Journal of Experimental Psychology*, *72*(4), 764–778.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12–21.

Bürkner, P.-C. Et al. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Chen, Z., Jäger, L. A., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. *Empirical Approaches to Linguistic Theory: Studies of Meaning and Structure, Studies in Generative Grammar*, 43–62.

Chomsky, N. (1981). Lectures on government and binding. Dordrecht, Foris, 1981. *Studies in Generative Grammar*, *9*.

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

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*(2), 128–144.

Clackson, K., & Heyer, V. (2014). Reflexive anaphor resolution in spoken language comprehension: Structural constraints and beyond. *Frontiers in Psychology*, *5*, 904.

Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Language and cognitive processes*, *4*(3-4), SI77–SI103.

Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension: 2703. *Italian journal of linguistics*, *11*(1), 11–40.

Cole, P., Hermon, G., & Sung, L.-M. (1993). Feature percolation. *Journal of East Asian Linguistics*, *2*(1), 91–118.

Cowper, E. A. (1987). Pied piping, feature percolation and the structure of the noun phrase. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *32*(4), 321–338.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1-2), 188–219.

Cunnings, I., Patterson, C., & Felser, C. (2015). Structural constraints on pronoun binding and coreference: Evidence from eye movements during reading. *Frontiers in Psychology*, *6*.

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

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, *102*, 16–27.

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PloS one*, *11*(1), e0146194.

Deutsch, A., & Dank, M. (2011). Symmetric and asymmetric patterns of attraction errors in producing subject–predicate agreement in hebrew: An issue of morphological structure. *Language and Cognitive Processes*, *26*(1), 24–46.

Dillon, B., Clifton, C., Sloggett, S., & Frazier, L. (2017). Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language*, *96*, 93–109.

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

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

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*(3), 531.

Engelmann, F., Jäger, L. A., & Vasishth, S. (2015). Cue confusion and distractor prominence explain inconsistent effects of retrieval interference in human sentence processing. *Proceedings of the 13th International Conference on Cognitive Modeling (The Netherlands, Groningen)*, 192–193.

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

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.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in psychological science*, *11*(1), 11–15.

Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satisficing approach, In *Proceedings of the 31st annual conference of the cognitive science society. austin: Cognitive science society.*

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, *8*(10), e77661.

Foote, R., & Bock, K. (2012). The role of morphology in subject–verb number agreement: A comparison of mexican and dominican spanish. *Language and Cognitive Processes*, *27*(3), 429–461. https://doi.org/10.1080/01690965.2010.550166

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in psychology*, *6*, 349.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, *101*(1), 173–216.

Franck, J., Soare, G., Frauenfelder, U. H., & Rizzi, L. (2010). Object interference in subject–verb agreement: The role of intermediate traces of movement. *Journal* *of Memory and Language*, *62*(2), 166–182. https://doi.org/10.1016/j.jml.2009.11.001

Franck, J., Vigliocco, G., & Nicol, J. (2002a). Subject-verb agreement errors in french and english: The role of syntactic hierarchy. *Language and cognitive processes*, *17*(4), 371–404.

Franck, J., Vigliocco, G., & Nicol, J. (2002b). Subject-verb agreement errors in french and english: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371–404. https://doi.org/10.1080/01690960143000254

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. *Doctoral dissertation, University of Connecticut.*

Frazier, L. (1987). Theories of sentence processing. *Modularity in knowledge representation and natural-language understanding.*

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178–210.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive science*, *44*(3), e12814.

Futrell, R., & Levy, R. (2017). Noisy-context surprisal as a human sentence processing cost model, In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 1, long papers.*

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

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of experimental psychology: learning, memory, and cognition*, *27*(6), 1411.

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.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model, In *Second meeting of the north american chapter of the association for computational linguistics*.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive psychology*, *110*, 70–104.

Harley, H., & Ritter, E. (2002). Person and number in pronouns: A feature-geometric analysis. *Language*, *78*(3), 482–526. https://doi.org/10.1353/lan.2002.0158

Hartsuiker, R. J., Antón-Méndez, I., & Van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language*, *45*(4), 546–572.

Hartsuiker, R. J., Kolk, H. H., & Huinck, W. J. (1999). Agrammatic production of subject–verb agreement: The effect of conceptual number. *Brain and Language*, *69*(2), 119–160. https://doi.org/10.1006/brln.1999.2059

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, *31*(8), 1316–1326.

Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: Evidence from subject-verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 891–904. https://doi.org/10.1037/0278-7393.31.5.891

Häussler, J. (2009). *The emergence of attraction errors during sentence comprehension* (Doctoral dissertation).

Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin Review*, *12*(4), 689–695. https://doi.org/10.3758/bf03196759

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Frontiers in psychology*, *6*, 506.

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

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

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

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2019). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study.

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.

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly journal of experimental psychology*, *69*(5), 1013–1040.

Kay, M. (2019). *tidybayes: Tidy data and geoms for Bayesian models* [R package version 1.1.0]. R package version 1.1.0. https://doi.org/10.5281/zenodo.1308151

Kennison, S. M. (2003). Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language*, *49*(3), 335–352.

Kimball, J., & Aissen, J. (1971). I think, you think, he think. *Linguistic Inquiry*, *2*(2), 241–246. http://www.jstor.org/stable/4177629

King, J., Andrews, C., & Wagers, M. (2012). Do reflexives always find a grammatical antecedent for themselves? *25th Annual CUNY Conference on Human Sentence Processing (New York, NY: The CUNY Graduate Center)*, 67.

Konieczny, L., Schimke, S., & Hemforth, B. (2004). An activation-based model of agreement errors in production and comprehension, In *Proceedings of the annual meeting of the cognitive science society*.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology*, *5*.

Lago, S., & Felser, C. (2018). Agreement attraction in native and nonnative speakers of german. *Applied Psycholinguistics*, *39*(3), 619–647.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in spanish comprehension. *Journal of Memory and Language*, *82*, 133–149.

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.

Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R. (2008b). A noisy-channel model of human sentence comprehension under uncertain input, In *Proceedings of the 2008 conference on empirical methods in natural language processing*.

Levy, R. (2011). Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results, In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*, *59*(2), 200–222.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 1–45.

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.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

Linzen, T., & Leonard, B. (2018). Distinct patterns of syntactic agreement errors in recurrent networks and humans. *arXiv preprint arXiv:1807.06882*.

Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in russian. *Language and cognitive processes*, *23*(6), 769–799.

Luegi, P., Leitão, M., Carvalho, M., & Costa, A. (2016). Retrieving and encoding during the processing of gender-marked and unmarked European Portuguese reflexives. *22nd AMLaP conference, Architectures and Mechanisms for Language Processing (Bilbao, Spain: Basque Center on Cognition, Brain and Language)*, 98.

Lyashevskaya, O., & Sharov, S. (2009). The frequency dictionary of modern Russian language. *Azbukovnik, Moscow*.

Lyutikova, E. A. (1997). Reflexives and emphasis. *Voprosy Jazykoznanija (Topics in the study of language)*, (6), 49–74.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. https://doi.org/10.1080/01690965.2012.728232

von der Malsburg, T., Lago, S., & Schäfer, R. (2020). The role of recovery processes in agreement attraction – an ERP investigation using German SOV constructions [Manuscript in preparation].

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1231.

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, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: Erp evidence from ellipsis. *Brain and language*, *135*, 42–51.

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.

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Ranton, Florida, USA, CRC Press.

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.

Mertzen, D., Laurinavichyute, A., Dillon, B. W., Engbert, R., & Vasishth, S. (2020). *A cross-linguistic investigation of proactive, similarity-based retrieval interference in sentence comprehension: No support from English, German and Russian eye-tracking data* [submitted]. submitted.

Meyer, A. S., Huettig, F., & Levelt, W. J. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, *89*, 1–7.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users.

Nairne, J. S. (2002a). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389–395.

Nairne, J. S. (2002b). Remembering over the short-term: The case against the standard model. *Annual review of psychology*, *53*(1), 53–81.

Nicenboim, B., Engelmann, F., Suckow, K., & Vasishth, S. (2015). Fail fast or succeed slowly: Good-enough processing can mask interference effects. *Proceedings of the 13th International Conference on Cognitive Modeling (The Netherlands, Groningen)*, 196–197.

Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in psychology*, *7*, 280.

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

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

Nicol, J., Forster, K., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569–587. https://doi.org/10.1006/jmla.1996.2497

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic research*, *18*(1), 5–19.

Nicol, J., Swinney, D., & Barss, A. (2003). The psycholinguistics of anaphora. *Anaphora: A reference guide*, 72–104.

Nieuwenhuis, R., Te Grotenhuis, M., & Pelzer, B. (2012). Influence.ME: Tools for detecting influential data in mixed effects models. *R Journal*, *4*(2), 38–47.

Noam, C. Et al. (1957). Syntactic structures. *The Hague: Mouton.*

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of memory and language*, *55*(4), 601–626.

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339.

Paspali, A., & Marinis, T. (2020). Gender agreement attraction in greek comprehension. *Frontiers in Psychology*, *11*, 717.

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

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology*, *69*(5), 950–971.

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology*, *5*.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999a). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427–456. https://doi.org/10.1006/jmla.1999.2653

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999b). Agreement processes in sentence comprehension. *Journal of Memory and language*, *41*(3), 427–456.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*(4), 329–347.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org

Rappaport, G. C. (1986). On anaphor binding in Russian. *Natural Language & Linguistic Theory*, *4*(1), 97–120.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior*, *22*(3), 358–374.

Reifegerste, J., Jarvis, R., & Felser, C. (2020). Effects of chronological age on native and nonnative sentence processing: Evidence from subject-verb agreement in german. *Journal of Memory and Language*, *111*, 104083.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, *14*, 611–653.

Rohde, D. (2005). Linger experiment presentation software. *h ttp://tedlab. mit. edu/dr/Linger*.

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., 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.

Schäfer, R., Lago, S., & von der Malsburg, T. (2019). ERP evidence of object agreement attraction in comprehension (E. Colunga, A. Kim, L. Michaelis, & B. Narasimhan, Eds.). In E. Colunga, A. Kim, L. Michaelis, & B. Narasimhan (Eds.), *Proceedings of the 32th annual cuny conference on human sentence processing*, Boulder, CO, USA, University of Colorado Boulder.

Schlueter, Z., Parker, D., & Lau, E. F. (2019). Error-driven retrieval in agreement attraction rarely leads to misinterpretation. *Frontiers in psychology*, *10*, 1002.

Schoknecht, P., Roehm, D., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2019). Looking forward does not mean forgetting about the past: Erp evidence for the interplay of predictive coding and interference during language processing. *BioRxiv*, 567560.

Sekerina, I. A., Campanelli, L., & Van Dyke, J. A. (2016). Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in psychology*, *7*, 873.

Slioussar, N., & Malko, A. (2016). Gender agreement attraction in russian: Production and comprehension evidence. *Frontiers in psychology*, *7*, 1651.

Slioussar, N., Stetsenko, A., & Matyushkina, T. (2015). Producing case errors in russian, In *Formal approaches to slavic linguistics: The first new york meeting*.

Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject-verb number agreement. *Cognitive Science*. https://doi.org/10.1111/cogs.12591

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, *49*(1), 1–46. https://doi.org/10.1016/j.cogpsych.2003.10.001

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of memory and language*, *60*(2), 308–327.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, *114*(3), 447–454.

Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, *1*(4), 281–313.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors.

Stone, K., Oltrogge, E., Vasishth, S., & Lago, S. (2020). The real-time application of grammatical constraints to prediction: Timecourse evidence from eye tracking, In *Talk at cuny, amherst, massachusetts, march 19–21.*

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370.

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

Tanner, D., & Bulkes, N. Z. (2015). Cues, quantification, and agreement in language comprehension. *Psychonomic bulletin & review*, *22*(6), 1753–1763.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of memory and language*, *76*, 195–215.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*(4), 740–759.

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from arabic. *Frontiers in psychology*, *6*, 347.

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., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

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

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

Vasishth, S. (2006). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories, In *Proceedings of the international conference on linguistic evidence*.

Vasishth, S., Jäger, L. A., & Nicenboim, B. (2017). Feature overwriting as a finite mixture process: Evidence from comprehension data. *arXiv preprint arXiv:1703.04081*.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, *71*, 147–161.

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125–134.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186–215.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. (1996). One or more labels on the bottles? Notional concord in dutch and french. *Language and Cognitive Processes*, *11*(4), 407–442. https://doi.org/10.1080/016909696387169

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, *68*(1), B13–B29. https://doi.org/10.1016/s0010-0277(98)00041-9

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in psychology*, *9*, 2.

Von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*(2), 105–143. https://doi.org/10.1016/s0010-0277(00)00063-9

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

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40–55.

Chapter 6

Appendix

6.1 Additional materials for Chapter 3

6.1.1 Materials of Experiments 1 and 2

Item	Attraction	Violation	Head	PP	Verb	Cond
1	none	morph.	The radio	by the desk	play	a
1	morph.	morph.	The radio	by the desks	play	b
1	none	semantic	The radio	by the desk	glows	c
1	semantic	semantic	The radio	by the lamp	glows	d
1	none	double	The radio	by the desk	glow	e
1	double	double	The radio	by the lamps	glow	f
1	semantic	double	The radio	by the lamp	glow	g
1	morph.	double	The radio	by the desks	glow	h
2	none	morph.	The camera	near the entrance	record	a
2	morph.	morph.	The camera	near the entrances	record	b
2	none	semantic	The camera	near the entrance	swings	c
2	semantic	semantic	The camera	near the door	swings	d
2	none	double	The camera	near the entrance	swing	e
2	double	double	The camera	near the doors	swing	f
2	semantic	double	The camera	near the door	swing	g
2	morph.	double	The camera	near the entrances	swing	h
3	none	morph.	The sign	at the information desk	say	a

Item	Attraction	Violation	Head	PP	Verb	Cond
3	morph.	morph.	The sign	at the information desks	say	b
3	none	semantic	The sign	at the information desk	descends	c
3	semantic	semantic	The sign	at the elevator	descends	d
3	none	double	The sign	at the information desk	descend	e
3	double	double	The sign	at the elevators	descend	f
3	semantic	double	The sign	at the elevator	descend	g
3	morph.	double	The sign	at the information desks	descend	h
4	none	morph.	The microphone	in the cell phone	hiss	a
4	morph.	morph.	The microphone	in the cell phones	hiss	b
4	none	semantic	The microphone	in the cell phone	commences	c
4	semantic	semantic	The microphone	for the ceremony	commences	d
4	none	double	The microphone	in the cell phone	commence	e
4	double	double	The microphone	for the ceremonies	commence	f
4	semantic	double	The microphone	for the ceremony	commence	g
4	morph.	double	The microphone	in the cell phones	commence	h
5	none	morph.	The vent	above the window	blow	a
5	morph.	morph.	The vent	above the windows	blow	b
5	none	semantic	The vent	above the window	stands	c
5	semantic	semantic	The vent	near the table	stands	d
5	none	double	The vent	above the window	stand	e
5	double	double	The vent	near the tables	stand	f
5	semantic	double	The vent	near the table	stand	g
5	morph.	double	The vent	above the windows	stand	h
6	none	morph.	The turn	after the junction	lead	a

Item	Attraction	Violation	Head	PP	Verb	Cond
6	morph.	morph.	The turn	after the junctions	lead	b
6	none	semantic	The turn	after the junction	stands	c
6	semantic	semantic	The turn	near the village	stands	d
6	none	double	The turn	after the junction	stand	e
6	double	double	The turn	near the villages	stand	f
6	semantic	double	The turn	near the village	stand	g
6	morph.	double	The turn	after the junctions	stand	h
7	none	morph.	The kiosk	near the theater	sell	a
7	morph.	morph.	The kiosk	near the theaters	sell	b
7	none	semantic	The kiosk	near the theater	descends	c
7	semantic	semantic	The kiosk	near the escalator	descends	d
7	none	double	The kiosk	near the theater	descend	e
7	double	double	The kiosk	near the escalators	descend	f
7	semantic	double	The kiosk	near the escalator	descend	g
7	morph.	double	The kiosk	near the theaters	descend	h
8	none	morph.	The flower stall	near the subway exit	smell	a
8	morph.	morph.	The flower stall	near the subway exits	smell	b
8	none	semantic	The flower stall	near the subway exit	illuminates	c
8	semantic	semantic	The flower stall	near the street lamp	illuminates	d
8	none	double	The flower stall	near the subway exit	illuminate	e
8	double	double	The flower stall	near the street lamps	illuminate	f
8	semantic	double	The flower stall	near the street lamp	illuminate	g
8	morph.	double	The flower stall	near the subway exits	illuminate	h
9	none	morph.	The bakery	near the office building	smell	a

Item	Attraction	Violation	Head	PP	Verb	Cond
9	morph.	morph.	The bakery	near the office buildings	smell	b
9	none	semantic	The bakery	near the office building	bubbles	c
9	semantic	semantic	The bakery	near the fountain	bubbles	d
9	none	double	The bakery	near the office building	bubble	e
9	double	double	The bakery	near the fountains	bubble	f
9	semantic	double	The bakery	near the fountain	bubble	g
9	morph.	double	The bakery	near the office buildings	bubble	h
10	none	morph.	The gas station	near the church	offer	a
10	morph.	morph.	The gas station	near the churches	offer	b
10	none	semantic	The gas station	near the church	leads	c
10	semantic	semantic	The gas station	near the freeway	leads	d
10	none	double	The gas station	near the church	lead	e
10	double	double	The gas station	near the freeways	lead	f
10	semantic	double	The gas station	near the freeway	lead	g
10	morph.	double	The gas station	near the churches	lead	h
11	none	morph.	The baggage carousel	with the defect	move	a
11	morph.	morph.	The baggage carousel	with the defects	move	b
11	none	semantic	The baggage carousel	with the defect	contains	c
11	semantic	semantic	The baggage carousel	with the bag	contains	d

Item	Attraction	Violation	Head	PP	Verb	Cond
11	none	double	The baggage carousel	with the defect	contain	e
11	double	double	The baggage carousel	with the bags	contain	f
11	semantic	double	The baggage carousel	with the bag	contain	g
11	morph.	double	The baggage carousel	with the defects	contain	h
12	none	morph.	The car	without a license plate	brake	a
12	morph.	morph.	The car	without license plates	brake	b
12	none	semantic	The car	without a license plate	inflates	c
12	semantic	semantic	The car	with the airbag	inflates	d
12	none	double	The car	without a license plate	inflate	e
12	double	double	The car	with the airbags	inflate	f
12	semantic	double	The car	with the airbag	inflate	g
12	morph.	double	The car	without license plates	inflate	h
13	none	morph.	The page	with the map	crash	a
13	morph.	morph.	The page	with the maps	crash	b
13	none	semantic	The page	with the map	sells	c
13	semantic	semantic	The page	with the advertisement	sells	d
13	none	double	The page	with the map	sell	e
13	double	double	The page	with the advertisements	sell	f
13	semantic	double	The page	with the advertisement	sell	g
13	morph.	double	The page	with the maps	sell	h
14	none	morph.	The video	of the crash	play	a
14	morph.	morph.	The video	of the crashes	play	b
14	none	semantic	The video	of the crash	works	c
14	semantic	semantic	The video	with the recipe	works	d
14	none	double	The video	of the crash	work	e
14	double	double	The video	with the recipes	work	f
14	semantic	double	The video	with the recipe	work	g
14	morph.	double	The video	of the crashes	work	h

Item	Attraction	Violation	Head	PP	Verb	Cond
15	none	morph.	The commercial	by the TV network	air	a
15	morph.	morph.	The commercial	by the TV networks	air	b
15	none	semantic	The commercial	by the TV network	cures	c
15	semantic	semantic	The commercial	about the pill	cures	d
15	none	double	The commercial	by the TV network	cure	e
15	double	double	The commercial	about the pills	cure	f
15	semantic	double	The commercial	about the pill	cure	g
15	morph.	double	The commercial	by the TV networks	cure	h
16	none	morph.	The article	about the marathon	appear	a
16	morph.	morph.	The article	about the marathons	appear	b
16	none	semantic	The article	about the marathon	accepts	c
16	semantic	semantic	The article	about the animal shelter	accepts	d
16	none	double	The article	about the marathon	accept	e
16	double	double	The article	about the animal shelters	accept	f
16	semantic	double	The article	about the animal shelter	accept	g
16	morph.	double	The article	about the marathons	accept	h
17	none	morph.	The plan	for the restructure	meet	a
17	morph.	morph.	The plan	for the restructures	meet	b
17	none	semantic	The plan	for the restructure	towers	c
17	semantic	semantic	The plan	for the skyscraper	towers	d
17	none	double	The plan	for the restructure	tower	e
17	double	double	The plan	for the skyscrapers	tower	f
17	semantic	double	The plan	for the skyscraper	tower	g
17	morph.	double	The plan	for the restructures	tower	h

Item	Attraction	Violation	Head	PP	Verb	Cond
18	none	morph.	The drawer	with the handle	open	a
18	morph.	morph.	The drawer	with the handles	open	b
18	none	semantic	The drawer	with the handle	cuts	c
18	semantic	semantic	The drawer	with the knife	cuts	d
18	none	double	The drawer	with the handle	cut	e
18	double	double	The drawer	with the knives	cut	f
18	semantic	double	The drawer	with the knife	cut	g
18	morph.	double	The drawer	with the handles	cut	h
19	none	morph.	The bakery	with the cake	take	a
19	morph.	morph.	The bakery	with the cakes	take	b
19	none	semantic	The bakery	with the cake	brews	c
19	semantic	semantic	The bakery	with the coffee machine	brews	d
19	none	double	The bakery	with the cake	brew	e
19	double	double	The bakery	with the coffee machines	brew	f
19	semantic	double	The bakery	with the coffee machine	brew	g
19	morph.	double	The bakery	with the cakes	brew	h
20	none	morph.	The medication	for the allergy	contain	a
20	morph.	morph.	The medication	for the allergies	contain	b
20	none	semantic	The medication	for the allergy	spreads	c
20	semantic	semantic	The medication	for the infection	spreads	d
20	none	double	The medication	for the allergy	spread	e
20	double	double	The medication	for the infections	spread	f
20	semantic	double	The medication	for the infection	spread	g
20	morph.	double	The medication	for the allergies	spread	h

Item	Attraction	Violation	Head	PP	Verb	Cond
21	none	morph.	The treatment	for the infection	sterilize	a
21	morph.	morph.	The treatment	for the infections	sterilize	b
21	none	semantic	The treatment	for the infection	closes	c
21	semantic	semantic	The treatment	for the wound	closes	d
21	none	double	The treatment	for the infection	close	e
21	double	double	The treatment	for the wounds	close	f
21	semantic	double	The treatment	for the wound	close	g
21	morph.	double	The treatment	for the infections	close	h
22	none	morph.	The house	near the mountain	overlook	a
22	morph.	morph.	The house	near the mountains	overlook	b
22	none	semantic	The house	near the mountain	blossoms	c
22	semantic	semantic	The house	near the tree	blossoms	d
22	none	double	The house	near the mountain	blossom	e
22	double	double	The house	near the trees	blossom	f
22	semantic	double	The house	near the tree	blossom	g
22	morph.	double	The house	near the mountains	blossom	h
23	none	morph.	The boat	without the engine	drift out	a
23	morph.	morph.	The boat	without the engines	drift out	b
23	none	semantic	The boat	without the engine	collapses	c
23	semantic	semantic	The boat	near the pier	collapses	d
23	none	double	The boat	without the engine	collapse	e
23	double	double	The boat	near the piers	collapse	f
23	semantic	double	The boat	near the pier	collapse	g
23	morph.	double	The boat	without the engines	collapse	h

Item	Attraction	Violation	Head	PP	Verb	Cond
24	none	morph.	The fence	around the garden	block	a
24	morph.	morph.	The fence	around the gardens	block	b
24	none	semantic	The fence	around the garden	teaches	c
24	semantic	semantic	The fence	around the school	teaches	d
24	none	double	The fence	around the garden	teach	e
24	double	double	The fence	around the schools	teach	f
24	semantic	double	The fence	around the school	teach	g
24	morph.	double	The fence	around the gardens	teach	h
25	none	morph.	The pond	with the bridge	freeze	a
25	morph.	morph.	The pond	with the bridges	freeze	b
25	none	semantic	The pond	with the bridge	produces	c
25	semantic	semantic	The pond	near the factory	produces	d
25	none	double	The pond	with the bridge	produce	e
25	double	double	The pond	near the factories	produce	f
25	semantic	double	The pond	near the factory	produce	g
25	morph.	double	The pond	with the bridges	produce	h

6.1.2 Analysis of plausibility ratings for Experiment 3

Ratings were analyzed using a ordinal regression model (Liddell & Kruschke, 2018). Factor 'plausible' encodes the difference between the two sentence preambles we constructed as plausible and the three preambles constructed as implausible (plausible preambles coded as 1, implausible as -1). Based on the outcomes of the analysis, we excluded four items for which the 95% credible interval for this estimated difference contained 0.

6.1.3 Materials of Experiment 3

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
1	none	morph.	The museum	of art	soon	open	a

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
1	morph.	morph.	The museum	of arts	soon	open	b
1	none	semantic	The museum	of art	soon	shoots	c
1	semantic	semantic	The museum	of the photographer	soon	shoots	d
1	none	double	The museum	of art	soon	shoot	e
1	double	double	The museum	of the photographers	soon	shoot	f
1	semantic	double	The museum	of the photographer	soon	shoot	g
1	morph.	double	The museum	of arts	soon	shoot	h
2	none	morph.	The shredder	near the table	usually	squeal	a
2	morph.	morph.	The shredder	near the tables	usually	squeal	b
2	none	semantic	The shredder	near the table	usually	scans	c
2	semantic	semantic	The shredder	near the copier	usually	scans	d
2	none	double	The shredder	near the table	usually	scan	e
2	double	double	The shredder	near the copiers	usually	scan	f
2	semantic	double	The shredder	near the copier	usually	scan	g
2	morph.	double	The shredder	near the tables	usually	scan	h
3	none	morph.	The car	with the dent	silently	approach	a

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
3	morph.	morph.	The car	with the dents	silently	approach	b
3	none	semantic	The car	with the dent	loudly	transmits	c
3	semantic	semantic	The car	with the walkie talkie	loudly	transmits	d
3	none	double	The car	with the dent	loudly	transmit	e
3	double	double	The car	with the walkie talkies	loudly	transmit	f
3	semantic	double	The car	with the walkie talkie	loudly	transmit	g
3	morph.	double	The car	with the dents	loudly	transmit	h
4	none	morph.	The shelf	with the jar	softly	creak	a
4	morph.	morph.	The shelf	with the jars	softly	creak	b
4	none	semantic	The shelf	with the jar	delicately	blooms	c
4	semantic	semantic	The shelf	with the plant	delicately	blooms	d
4	none	double	The shelf	with the jar	delicately	bloom	e
4	double	double	The shelf	with the plants	delicately	bloom	f
4	semantic	double	The shelf	with the plant	delicately	bloom	g
4	morph.	double	The shelf	with the jars	delicately	bloom	h
5	none	morph.	The boat	with the flag	silently	glide	a
5	morph.	morph.	The boat	with the flags	silently	glide	b
5	none	semantic	The boat	with the flag	silently	rows	c
5	semantic	semantic	The boat	with the contestant	silently	rows	d
5	none	double	The boat	with the flag	silently	row	e
5	double	double	The boat	with the contestants	silently	row	f

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
5	semantic	double	The boat	with the contestant	silently	row	g
5	morph.	double	The boat	with the flags	silently	row	h
6	none	morph.	The exit	next to the traffic light	apparently	lead	a
6	morph.	morph.	The exit	next to the traffic lights	apparently	lead	b
6	none	semantic	The exit	next to the traffic light	apparently	dries up	c
6	semantic	semantic	The exit	next to the creek	apparently	dries up	d
6	none	double	The exit	next to the traffic light	apparently	dry up	e
6	double	double	The exit	next to the creeks	apparently	dry up	f
6	semantic	double	The exit	next to the creek	apparently	dry up	g
6	morph.	double	The exit	next to the traffic lights	apparently	dry up	h
7	none	morph.	The tram stop	next to the fire hydrant	usually	shelter	a
7	morph.	morph.	The tram stop	next to the fire hydrants	usually	shelter	b
7	none	semantic	The tram stop	next to the fire hydrant	usually	sells	c
7	semantic	semantic	The tram stop	next to the shop	usually	sells	d
7	none	double	The tram stop	next to the fire hydrant	usually	sell	e
7	double	double	The tram stop	next to the shops	usually	sell	f

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
7	semantic	double	The tram stop	next to the shop	usually	sell	g
7	morph.	double	The tram stop	next to the fire hydrants	usually	sell	h
8	none	morph.	The restaurant	with the chocolate fountain	admittedly	go bankrupt	a
8	morph.	morph.	The restaurant	with the chocolate fountains	admittedly	go bankrupt	b
8	none	semantic	The restaurant	with the chocolate fountain	admittedly	withers	c
8	semantic	semantic	The restaurant	with the winter garden	admittedly	withers	d
8	none	double	The restaurant	with the chocolate fountain	admittedly	wither	e
8	double	double	The restaurant	with the winter gardens	admittedly	wither	f
8	semantic	double	The restaurant	with the winter garden	admittedly	wither	g
8	morph.	double	The restaurant	with the chocolate fountains	admittedly	wither	h
9	none	morph.	The highrise	with the loft	proudly	stand	a
9	morph.	morph.	The highrise	with the lofts	proudly	stand	b
9	none	semantic	The highrise	with the loft	silently	descends	c
9	semantic	semantic	The highrise	with the elevator	silently	descends	d
9	none	double	The highrise	with the loft	silently	descend	e

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
9	double	double	The highrise	with the elevators	silently	descend	f
9	semantic	double	The highrise	with the elevator	silently	descend	g
9	morph.	double	The highrise	with the lofts	silently	descend	h
10	none	morph.	The wall calendar	with the landscape	always	hang	a
10	morph.	morph.	The wall calendar	with the landscapes	always	hang	b
10	none	semantic	The wall calendar	with the landscape	always	smiles	c
10	semantic	semantic	The wall calendar	with the lady	always	smiles	d
10	none	double	The wall calendar	with the landscape	always	smile	e
10	double	double	The wall calendar	with the ladies	always	smile	f
10	semantic	double	The wall calendar	with the lady	always	smile	g
10	morph.	double	The wall calendar	with the landscape	always	smile	h
11	none	morph.	The washer	by the dryer	sometimes	leak	a
11	morph.	morph.	The washer	by the dryers	sometimes	leak	b
11	none	semantic	The washer	by the dryer	sometimes	cooks	c
11	semantic	semantic	The washer	by the stove	sometimes	cooks	d
11	none	double	The washer	by the dryer	sometimes	cook	e
11	double	double	The washer	by the stoves	sometimes	cook	f
11	semantic	double	The washer	by the stove	sometimes	cook	g
11	morph.	double	The washer	by the dryers	sometimes	cook	h
12	none	morph.	The newsstand	near the bench	usually	sell	a

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
12	morph.	morph.	The newsstand	near the coffee shops	usually	sell	b
12	none	semantic	The newsstand	near the bench	usually	smells	c
12	semantic	semantic	The newsstand	near the coffee shop	usually	smells	d
12	none	double	The newsstand	near the bench	usually	smell	e
12	double	double	The newsstand	near the coffee shops	usually	smell	f
12	semantic	double	The newsstand	near the coffee shop	usually	smell	g
12	morph.	double	The newsstand	near the benches	usually	smell	h
13	none	morph.	The fireplace	near the shelf	soothingly	crackle	a
13	morph.	morph.	The fireplace	near the shelves	soothingly	crackle	b
13	none	semantic	The fireplace	near the shelf	soothingly	rocks	c
13	semantic	semantic	The fireplace	near the chair	soothingly	rocks	d
13	none	double	The fireplace	near the shelf	soothingly	rock	e
13	double	double	The fireplace	near the chairs	soothingly	rock	f
13	semantic	double	The fireplace	near the chair	soothingly	rock	g
13	morph.	double	The fireplace	near the shelves	soothingly	rock	h

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
14	none	morph.	The bakery	near the office building	rarely	smell	a
14	morph.	morph.	The bakery	near the office buildings	rarely	smell	b
14	none	semantic	The bakery	near the office building	rarely	sprays	c
14	semantic	semantic	The bakery	near the fire hydrant	rarely	sprays	d
14	none	double	The bakery	near the office building	rarely	spray	e
14	double	double	The bakery	near the fire hydrants	rarely	spray	f
14	semantic	double	The bakery	near the fire hydrant	rarely	spray	g
14	morph.	double	The bakery	near the office buildings	rarely	spray	h
15	none	morph.	The cinema	near the playground	sometimes	advertise	a
15	morph.	morph.	The cinema	near the playgrounds	sometimes	advertise	b
15	none	semantic	The cinema	near the playground	sometimes	sheds	c
15	semantic	semantic	The cinema	near the old tree	sometimes	sheds	d
15	none	double	The cinema	near the playground	sometimes	shed	e

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
15	double	double	The cinema	near the old trees	sometimes	shed	f
15	semantic	double	The cinema	near the old tree	sometimes	shed	g
15	morph.	double	The cinema	near the playgrounds	sometimes	shed	h
16	none	morph.	The blender	next to the breadmaker	loudly	whirr	a
16	morph.	morph.	The blender	next to the breadmakers	loudly	whirr	b
16	none	semantic	The blender	next to the breadmaker	loudly	hisses	c
16	semantic	semantic	The blender	next to the coffee machine	loudly	hisses	d
16	none	double	The blender	next to the breadmaker	loudly	hiss	e
16	double	double	The blender	next to the coffee machines	loudly	hiss	f
16	semantic	double	The blender	next to the coffee machine	loudly	hiss	g
16	morph.	double	The blender	next to the breadmakers	loudly	hiss	h
17	none	morph.	The display	next to the plant	suddenly	flicker	a
17	morph.	morph.	The display	next to the plants	suddenly	flicker	b
17	none	semantic	The display	next to the plant	suddenly	clicks	c

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
17	semantic	semantic	The display	next to the keyboard	suddenly	clicks	d
17	none	double	The display	next to the plant	suddenly	click	e
17	double	double	The display	next to the keyboards	suddenly	click	f
17	semantic	double	The display	next to the keyboard	suddenly	click	g
17	morph.	double	The display	next to the plants	suddenly	click	h
18	none	morph.	The keyboard	next to the display	apparently	click	a
18	morph.	morph.	The keyboard	next to the displays	apparently	click	b
18	none	semantic	The keyboard	next to the display	apparently	withers	c
18	semantic	semantic	The keyboard	next to the plant	apparently	withers	d
18	none	double	The keyboard	next to the display	apparently	wither	e
18	double	double	The keyboard	next to the plants	apparently	wither	f
18	semantic	double	The keyboard	next to the plant	apparently	wither	g
18	morph.	double	The keyboard	next to the displays	apparently	wither	h
19	none	morph.	The water meter	next to the towel hook	regularly	tick	a
19	morph.	morph.	The water meter	next to the towel hooks	regularly	tick	b
19	none	semantic	The water meter	next to the towel hook	regularly	leaks	c

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
19	semantic	semantic	The water meter	next to the sink	regularly	leaks	d
19	none	double	The water meter	next to the towel hook	regularly	leak	e
19	double	double	The water meter	next to the sinks	regularly	leak	f
19	semantic	double	The water meter	next to the sink	regularly	leak	g
19	morph.	double	The water meter	next to the towel hooks	regularly	leak	h
20	none	morph.	The pipe	below the light switch	usually	dribble	a
20	morph.	morph.	The pipe	below the light switches	usually	dribble	b
20	none	semantic	The pipe	below the light switch	usually	swings out	c
20	semantic	semantic	The pipe	above the window	usually	swings out	d
20	none	double	The pipe	below the light switch	usually	swing out	e
20	double	double	The pipe	above the windows	usually	swing out	f
20	semantic	double	The pipe	above the window	usually	swing out	g
20	morph.	double	The pipe	below the light switches	usually	swing out	h
21	none	morph.	The radio	by the desk	usually	play	a
21	morph.	morph.	The radio	by the desks	usually	play	b
21	none	semantic	The radio	by the desk	usually	glows	c
21	semantic	semantic	The radio	by the lamp	usually	glows	d
21	none	double	The radio	by the desk	usually	glow	e
21	double	double	The radio	by the lamps	usually	glow	f

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
21	semantic	double	The radio	by the lamp	usually	glow	g
21	morph.	double	The radio	by the desks	usually	glow	h
22	none	morph.	The car	without license plate	suddenly	slow down	a
22	morph.	morph.	The car	without license plates	suddenly	slow down	b
22	none	semantic	The car	without license plate	suddenly	inflates	c
22	semantic	semantic	The car	with the faulty airbag	suddenly	inflates	d
22	none	double	The car	without license plate	suddenly	inflate	e
22	double	double	The car	with the faulty airbags	suddenly	inflate	f
22	semantic	double	The car	with the faulty airbag	suddenly	inflate	g
22	morph.	double	The car	without license plates	suddenly	inflate	h
23	none	morph.	The medication	for the allergy	obviously	help	a
23	morph.	morph.	The medication	for the allergies	obviously	help	b
23	none	semantic	The medication	for the allergy	obviously	spreads	c
23	semantic	semantic	The medication	for the infection	obviously	spreads	d
23	none	double	The medication	for the allergy	obviously	spread	e
23	double	double	The medication	for the infections	obviously	spread	f
23	semantic	double	The medication	for the infection	obviously	spread	g

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
23	morph.	double	The medication	for the allergies	obviously	spread	h
24	none	morph.	The treatment	for the infection	finally	work	a
24	morph.	morph.	The treatment	for the infections	finally	work	b
24	none	semantic	The treatment	for the infection	finally	closes	c
24	semantic	semantic	The treatment	for the wound	finally	closes	d
24	none	double	The treatment	for the infection	finally	close	e
24	double	double	The treatment	for the wounds	finally	close	f
24	semantic	double	The treatment	for the wound	finally	close	g
24	morph.	double	The treatment	for the infections	finally	close	h
25	none	morph.	The fence	around the garden	supposedly	conceal	a
25	morph.	morph.	The fence	around the gardens	supposedly	conceal	b
25	none	semantic	The fence	around the garden	supposedly	teaches	c
25	semantic	semantic	The fence	around the school	supposedly	teaches	d
25	none	double	The fence	around the garden	supposedly	teach	e
25	double	double	The fence	around the schools	supposedly	teach	f
25	semantic	double	The fence	around the school	supposedly	teach	g

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
25	morph.	double	The fence	around the gardens	supposedly	teach	h
26	none	morph.	The pond	with the bridge	clearly	dry up	a
26	morph.	morph.	The pond	with the bridges	clearly	dry up	b
26	none	semantic	The pond	with the bridge	clearly	sells	c
26	semantic	semantic	The pond	near the farm stand	clearly	sells	d
26	none	double	The pond	with the bridge	clearly	sell	e
26	double	double	The pond	near the farm stands	clearly	sell	f
26	semantic	double	The pond	near the farm stand	clearly	sell	g
26	morph.	double	The pond	with the bridges	clearly	sell	h
27	none	morph.	The mansion	near the hill	allegedly	cost	a
27	morph.	morph.	The mansion	near the hills	allegedly	cost	b
27	none	semantic	The mansion	near the hill	allegedly	dries up	c
27	semantic	semantic	The mansion	near the pond	allegedly	dries up	d
27	none	double	The mansion	near the hill	allegedly	dry up	e
27	double	double	The mansion	near the ponds	allegedly	dry up	f
27	semantic	double	The mansion	near the pond	allegedly	dry up	g

Item	Attraction	Violation	Head	PP	Adverb	Verb	Cond
27	morph.	double	The mansion	near the hills	allegedly	dry up	h

6.1.4 Reaction times analysis

For the sake of brevity, we present only the analysis of reaction times (RTs) performed on the data set from Experiment 3, which provides the most precise and unbiased estimates. Recall that reaction times in our study incorporate not only the decision times, but also the time it took participants to read the sentence preamble, and preambles in different conditions were of varying lengths. We aim to account for that variation by including the length of sentence preamble as a covariate.

RTs were modeled assuming lognormal distribution; we used default *brms* priors. Models had the same structure and contrast coding as those used for the analysis of accuracy on the pooled dataset, except that we added two more predictors of reaction times, the trial response accuracy and the centered length of sentence preamble. Accuracy was coded as 0 for the incorrect and 1 for the correct responses in the model, and was included both as a main effect and an interaction term. Preamble length was only included as a main effect.

Analysis of conditions a–d The estimated RTs for sentence preambles of average length are presented on Figure 6.1. In trials with correct responses, we found slowdowns in both the morphosyntactic and semantic attraction conditions. Average RT in condition (a) with correct responses was 4,781 ms, and did not differ from RTs in trials with incorrect responses ($\hat{\beta} = 0.05$, 95%-CrI: $[-0.08, 0.18]$). The baseline for semantic attraction (c) did not differ from the morphosyntactic baseline (a) ($\hat{\beta} = -0.07$, 95%-CrI: $[-0.29, 0.14]$). In condition (b) with morphosyntactic attraction RTs were 1,043 ms slower than the baseline (a) ($\hat{\beta} = 0.32$, 95%-CrI: $[0.14, 0.49]$). In condition (d) with semantic attraction RTs tended to be slower (928 ms) than in the baseline (c), but the 95%-CrI included 0 ($\hat{\beta} = 0.18$, 95%-CrI: $[-0.06, 0.43]$, $P(\beta > 0) = 0.927$).

In trials with incorrect responses, no differences between conditions were found. Average RT in condition (a) was 4,537 ms ($\hat{\beta} = 8.42$, 95%-CrI: $[8.32, 8.53]$). The baseline for semantic attraction (c) did not differ from the morphosyntactic baseline (a) (c vs. a: $\hat{\beta} = 0.03$, 95%-CrI: $[-0.13, 0.20]$). RTs in condition (b) with morphosyntactic attraction tended to be lower (513 ms) than in the baseline (a) ($\hat{\beta} = -0.12$, 95%-CrI: $[-0.25, 0.01]$, $P(\beta < 0) = 0.964$);

Predictor	Log-Odds Estimate	95%-CrI
Rating 1	-2.26	-2.57 – -1.94
Rating 2	-1.26	-1.55 – -0.95
Rating 3	-0.61	-0.90 – -0.30
Rating 4	0.04	-0.25 – 0.34
Rating 5	0.81	0.53 – 1.12
Rating 6	1.90	1.60 – 2.22
Plausible	1.58	1.27 – 1.89

Table 6.2: Analysis of plausibility ratings for Experiment 3 items.

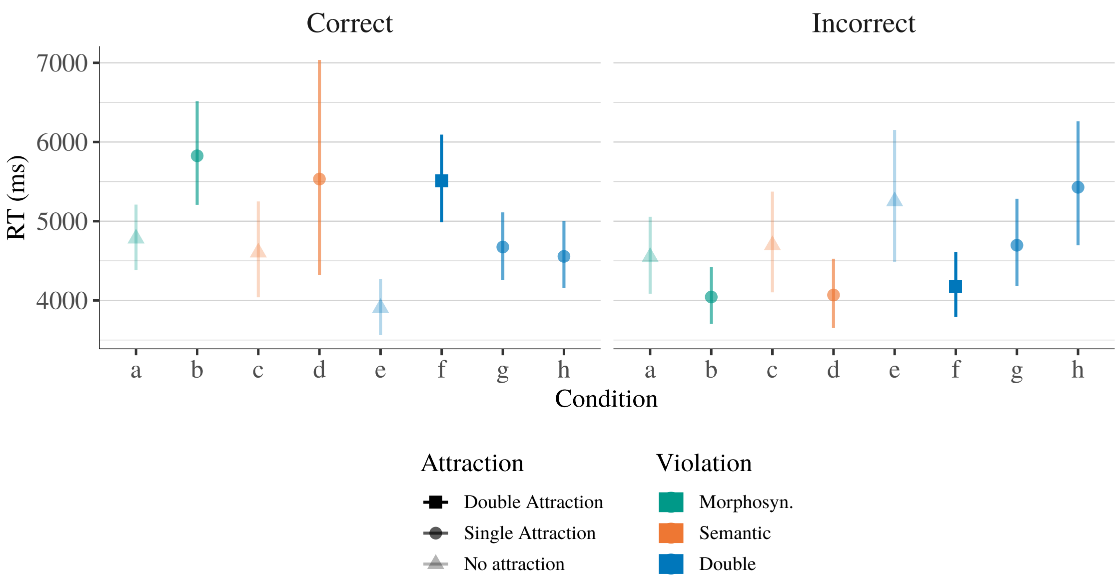

Figure 6.1: Reaction times in Experiment 3 depending on the trial response: Estimated condition means with 95% credible intervals.

similarly, RTs in condition (d) with semantic attraction tended to be lower (628 ms) than in its respective baseline (c) ($\hat{\beta} = -0.14$, 95%-CrI: $[-0.30, 0.01]$, $P(\beta < 0) = 0.966$).

Preamble length did not affect reaction times ($\hat{\beta} = 0.01$, 95%-CrI: $[-0.001, 0.02]$).

Overall, we found that slowdowns of similar magnitudes were present in correct trials with morphosyntactic and semantic attraction, but absent in incorrect trials. Instead, in incorrect trials, both conditions with attraction seem to lead to a speedup, but the estimate of the speedup included 0 for both attraction effects.

Analysis testing the interaction of morphosyntactic and semantic attraction (conditions e–h) The estimated RTs are presented on Figure 6.1. In correct trials, both morphosyntactic and semantic attraction caused a slowdown in RTs (morphosyntactic attraction, 654 ms: $\hat{\beta} = 0.20$, 95%-CrI: $[0.06, 0.34]$; semantic attraction, 766 ms: $\hat{\beta} = 0.37$, 95%-CrI: $[0.21, 0.53]$). We also found a slowdown of 1,603 ms due to double attraction in correct trials ($\hat{\beta} = 0.34$, 95%-CrI: $[0.21, 0.48]$). Morphosyntactic and semantic attraction effects did not interact ($\hat{\beta} = 0.16$, 95%-CrI: $[-0.12, 0.44]$).

In incorrect trials, the estimated RTs in condition (e) without attraction comprised 4,865 ms ($\hat{\beta} = 8.49$, 95%-CrI: $[8.42, 8.56]$). There was no speedup due to morphosyntactic attraction ($\hat{\beta} = -0.04$, 95%-CrI: $[-0.17, 0.09]$), but RTs in the semantic attraction condition were 628 ms shorter than in the corresponding control condition ($\hat{\beta} = -0.19$, 95%-CrI: $[-0.32, -0.05]$). We also found a speedup of 1,069 ms due to double attraction ($\hat{\beta} = -0.23$, 95%-CrI: $[-0.41, -0.04]$). Attraction effects did not interact ($\hat{\beta} = -0.15$, 95%-CrI: $[-0.40, 0.11]$).

Longer preambles increased reaction times ($\hat{\beta} = 0.01$, 95%-CrI: $[0.001, 0.02]$).

Discussion With the preamble length factored out, we still found clear effects of attraction. There is no difference between the morphosyntactic and semantic attraction effects, which is consistent with both effects having a common underlying source. Both morphosyntactic and semantic attraction lead to slowdowns of similar magnitudes in trials that received correct responses, and to speedups in trials that received incorrect responses. This echoes previous findings of faster processing times in the probes that received incorrect responses (Laurinavichyute et al., 2017; von der Malsburg & Vasishth, 2013; Nicenboim et al., 2016; Nicenboim & Vasishth, 2018). In general, slowdown in reaction times is interpreted as the additional time taken to notice or reanalyze sentence ill-formedness. But if sentence ill-formedness goes unnoticed, there is no need for additional processing time. Our results align with this picture: correct trials had longer RTs for morphosyntactic, semantic, and double attraction conditions. This suggests that noticing and mentally correcting sentence ill-formedness is harder and takes more time in attraction than in control conditions.

6.2 Additional materials for Chapter 4

6.2.1 Materials of Experiment 1

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
1	Number match, semantic match	The_admirer of_the singer supposedly thinks the_show was_a big_success.	Who considered the show a success? — Admirer/Singer/Admirers/Singers
1	Number mismatch, semantic match	The_admirer of_the singers supposedly thinks the_show was_a big_success.	Who considered the show a success? — Admirer/Singer/Admirers/Singers
1	Number match, semantic mismatch	The_admirer of_the play supposedly thinks the_show was_a big_success.	Who considered the show a success? — Admirer/Play/Admirers/Plays
1	Number mismatch, semantic mismatch	The_admirer of_the plays supposedly thinks the_show was_a big_success.	Who considered the show a success? — Admirer/Play/Admirers/Plays
2	Number match, semantic match	The_supervisor of_the trainee informally recommends regular breaks.	Who encouraged regular breaks? — Supervisor/Trainee/Supervisors/Trainees
2	Number mismatch, semantic match	The_supervisor of_the trainees informally recommends regular breaks.	Who encouraged regular breaks? — Supervisor/Trainee/Supervisors/Trainees
2	Number match, semantic mismatch	The_supervisor of_the building_site informally recommends regular breaks.	Who encouraged regular breaks? — Supervisor/Building/Supervisors/Buildings
2	Number mismatch, semantic mismatch	The_supervisor of_the building_sites informally recommends regular breaks.	Who encouraged regular breaks? — Supervisor/Building/Supervisors/Buildings

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
3	Number match, semantic match	The_opponent of_the legislator secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Legislator/Opponents/Legislators
3	Number mismatch, semantic match	The_opponent of_the legislators secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Legislator/Opponents/Legislators
3	Number match, semantic mismatch	The_opponent of_the bill secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Bill/Opponents/Bills
3	Number mismatch, semantic mismatch	The_opponent of_the bills secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Bill/Opponents/Bills
4	Number match, semantic match	The_supporter of_the politician hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Politician/Supporters/Politicians
4	Number mismatch, semantic match	The_supporter of_the politicians hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Politician/Supporters/Politicians
4	Number match, semantic mismatch	The_supporter of_the regulation hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Regulation/Supporters/Regulation
4	Number mismatch, semantic mismatch	The_supporter of_the regulations hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Regulation/Supporters/Regulation

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
5	Number match, semantic match	The_visitor of_the poet graciously admires the_spacious room.	Who praised the room? — Visitor/Poet/Visitors/Poets
5	Number mismatch, semantic match	The_visitor of_the poets graciously admires the_spacious room.	Who praised the room? — Visitor/Poet/Visitors/Poets
5	Number match, semantic mismatch	The_visitor of_the gallery graciously admires the_spacious room.	Who praised the room? — Visitor/Gallery/Visitors/Galleries
5	Number mismatch, semantic mismatch	The_visitor of_the galleries graciously admires the_spacious room.	Who praised the room? — Visitor/Gallery/Visitors/Galleries
6	Number match, semantic match	The_observer of_the monkey frantically gesticulates to_call for_attention.	Who used pantomime? — Observer/Monkey/Observers/Monkeys
6	Number mismatch, semantic match	The_observer of_the monkeys frantically gesticulates to_call for_attention.	Who used pantomime? — Observer/Monkey/Observers/Monkeys
6	Number match, semantic mismatch	The_observer of_the event frantically gesticulates to_call for_attention.	Who used pantomime? — Observer/Event/Observers/Events
6	Number mismatch, semantic mismatch	The_observer of_the events frantically gesticulates to_call for_attention.	Who used pantomime? — Observer/Event/Observers/Events

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
7	Number match, semantic match	The_favorite of_the investor openly boasts to_the girl about his_talents.	Who bragged to the girl? — Favorite/Investor/Favorites/Investors
7	Number mismatch, semantic match	The_favorite of_the investors openly boasts to_the girl about his_talents.	Who bragged to the girl? — Favorite/Investor/Favorites/Investors
7	Number match, semantic mismatch	The_favorite in_the race openly boasts to_the girl about his_talents.	Who bragged to the girl? — Favorite/Race/Favorites/Races
7	Number mismatch, semantic mismatch	The_favorite in_the races openly boasts to_the girl about his_talents.	Who bragged to the girl? — Favorite/Race/Favorites/Races
8	Number match, semantic match	The_advocate for_the teenager enthusiastically addresses the_audience in_court.	Who talked to the audience? — Advocate/Teenager/Advocates/Teenagers
8	Number mismatch, semantic match	The_advocate for_the teenagers enthusiastically addresses the_audience in_court.	Who talked to the audience? — Advocate/Teenager/Advocates/Teenagers
8	Number match, semantic mismatch	The_advocate for_the technology enthusiastically addresses the_audience in_court.	Who talked to the audience? — Advocate/Technology/Advocates/Technologies
8	Number mismatch, semantic mismatch	The_advocate for_the technologies enthusiastically addresses the_audience in_court.	Who talked to the audience? — Advocate/Technology/Advocates/Technologies

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
9	Number match, semantic match	The_fan of_the singer still dreams_of an_invite to_the private party.	Who thought about the party? — Fan/Singer/Fans/Singers
9	Number mismatch, semantic match	The_fan of_the singers still dreams_of an_invite to_the private party.	Who thought about the party? — Fan/Singer/Fans/Singers
9	Number match, semantic mismatch	The_fan of_the board game still dreams_of an_invite to_the private party.	Who thought about the party? — Fan/Board game/Fans/Board games
9	Number mismatch, semantic mismatch	The_fan of_the board games still dreams_of an_invite to_the private party.	Who thought about the party? — Fan/Board game/Fans/Board games
10	Number match, semantic match	The_interpreter of_the ambassador noticeably stumbles on_a difficult passage.	Who experienced difficulties? — Interpreter/Ambassador/Interpreters/Ambassadors
10	Number mismatch, semantic match	The_interpreter of_the ambassadors noticeably stumbles on_a difficult passage.	Who experienced difficulties? — Interpreter/Ambassador/Interpreters/Ambassadors
10	Number match, semantic mismatch	The_interpreter of_the speech noticeably stumbles on_a difficult passage.	Who experienced difficulties? — Interpreter/Speech/Interpreters/Speeches
10	Number mismatch, semantic mismatch	The_interpreter of_the speeches noticeably stumbles on_a difficult passage.	Who experienced difficulties? — Interpreter/Speech/Interpreters/Speeches

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
11	Number match, semantic match	The_heir to_the duke reverently preserves the_original state of_things.	Who protected the original state of things? — Heir/Duke/Heirs/Dukes
11	Number mismatch, semantic match	The_heir to_the dukes reverently preserves the_original state of_things.	Who protected the original state of things? — Heir/Duke/Heirs/Dukes
11	Number match, semantic mismatch	The_heir to_the painting reverently preserves the_original state of_things.	Who protected the original state of things? — Heir/Painting/Heirs/Paintings
11	Number mismatch, semantic mismatch	The_heir to_the paintings reverently preserves the_original state of_things.	Who protected the original state of things? — Heir/Painting/Heirs/Paintings
12	Number match, semantic match	The_painter of_the king really wishes_for another commission.	Who desired a commission? — Painter/King/Painters/Kings
12	Number mismatch, semantic match	The_painter of_the kings really wishes_for another commission.	Who desired a commission? — Painter/King/Painters/Kings
12	Number match, semantic mismatch	The_painter of_the landscape really wishes_for another commission.	Who desired a commission? — Painter/Landscape/Painters/Landscapes
12	Number mismatch, semantic mismatch	The_painter of_the landscapes really wishes_for another commission.	Who desired a commission? — Painter/Landscape/Painters/Landscapes

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
13	Number match, semantic match	The_manager of_the musician cheerfully signs a_contract for_the next two years.	Who put the signature on the contract? — Manager/Musician/Managers/Musicians
13	Number mismatch, semantic match	The_manager of_the musicians cheerfully signs a_contract for_the next two years.	Who put the signature on the contract? — Manager/Musician/Managers/Musicians
13	Number match, semantic mismatch	The_manager of_the estate cheerfully signs a_contract for_the next two years.	Who put the signature on the contract? — Manager/Estate/Managers/Estates
13	Number mismatch, semantic mismatch	The_manager of_the estates cheerfully signs a_contract for_the next two years.	Who put the signature on the contract? — Manager/Estate/Managers/Estates
14	Number match, semantic match	The_student of_the professor categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Professor/Students/Professors
14	Number mismatch, semantic match	The_student of_the professors categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Professor/Students/Professors
14	Number match, semantic mismatch	The_student in_the course categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Course/Students/Courses
14	Number mismatch, semantic mismatch	The_student in_the courses categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Course/Students/Courses

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
15	Number match, semantic match	The_critic of_the politician passionately protests the_results of_the_vote.	Who opposed the results of the vote? — Critic/Politician/Critics/Politicians
15	Number mismatch, semantic match	The_critic of_the politicians passionately protests the_results of_the_vote.	Who opposed the results of the vote? — Critic/Politician/Critics/Politicians
15	Number match, semantic mismatch	The_critic of_the proposal passionately protests the_results of_the_vote.	Who opposed the results of the vote? — Critic/Proposal/Critics/Proposals
15	Number mismatch, semantic mismatch	The_critic of_the proposals passionately protests the_results of_the_vote.	Who opposed the results of the vote? — Critic/Proposal/Critics/Proposals
16	Number match, semantic match	The_gardener of_the landlord heatedly insists_on waiting another week.	Who demanded a delay? — Gardener/Landlord/Gardeners/Landlords
16	Number mismatch, semantic match	The_gardener of_the landlords heatedly insists_on waiting another week.	Who demanded a delay? — Gardener/Landlord/Gardeners/Landlords
16	Number match, semantic mismatch	The_gardener of_the park heatedly insists_on waiting another week.	Who demanded a delay? — Gardener/Park/Gardeners/Parks
16	Number mismatch, semantic mismatch	The_gardener of_the parks heatedly insists_on waiting another week.	Who demanded a delay? — Gardener/Park/Gardeners/Parks

Table 6.4: Stimuli used in Experiment 1.

ID	Condition	Sentence	Question and response options
17	Number match, semantic match	The_coach of_the athlete skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Athlete/Coaches/Athletes
17	Number mismatch, semantic match	The_coach of_the athletes skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Athlete/Coaches/Athletes
17	Number match, semantic mismatch	The_coach with_the tattoo skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Tattoo/Coaches/Tattoos
17	Number mismatch, semantic mismatch	The_coach with_the tattoos skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Tattoo/Coaches/Tattoos

Note. The response options are always presented in the following order: correct response; wrong noun, correct number marking; correct noun, wrong number marking; wrong noun, wrong number marking. Option “I’m not sure” is omitted as it is the same in every item. Underscores mark the words that were presented as a single region during self-paced reading. Participants did not see the underscores.

6.2.2 Materials of Experiment 2

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
1	Number match, semantic match	The_admirer of_the singer, according to_the Daily Mail, apparently thinks that the_show was a_big success.	Who considered the show a success? — Admirer/Singer/Admirers/Singers
1	Number mismatch, semantic match	The_admirer of_the singers, according to_the Daily Mail, apparently thinks that the_show was a_big success.	Who considered the show a success? — Admirer/Singer/Admirers/Singers
1	Number match, semantic mismatch	The_admirer of_the play, according to_the Daily Mail, apparently thinks that the_show was a_big success.	Who considered the show a success? — Admirer/Play/Admirers/Plays
1	Number mismatch, semantic mismatch	The_admirer of_the plays, according to_the Daily Mail, apparently thinks that the_show was a_big success.	Who considered the show a success? — Admirer/Play/Admirers/Plays
2	Number match, semantic match	The_supervisor of_the trainee, as_far as_I can remember, informally recommends just a_few regular breaks.	Who spoke about breaks? — Supervisor/Trainee/Supervisors/Trainees
2	Number mismatch, semantic match	The_supervisor of_the trainees, as_far as_I can remember, informally recommends just a_few regular breaks.	Who spoke about breaks? — Supervisor/Trainee/Supervisors/Trainees
2	Number match, semantic mismatch	The_supervisor of_the building_site, as_far as_I can remember, informally recommends just a_few regular breaks.	Who spoke about breaks? — Supervisor/Building site/Supervisors/Building sites

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
2	Number mismatch, semantic mismatch	The_supervisor of_the building_sites, as_far as_I can remember, informally recommends just a_few regular breaks.	Who spoke about breaks? — Supervisor/Building site/Supervisors/Building sites
3	Number match, semantic match	The_opponent of_the legislator, according to_the unnamed sources, secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Legislator/Opponents/Legislators
3	Number mismatch, semantic match	The_opponent of_the legislators, according to_the unnamed sources, secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Legislator/Opponents/Legislators
3	Number match, semantic mismatch	The_opponent of_the bill, according to_the unnamed sources, secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Bill/Opponents/Bills
3	Number mismatch, semantic mismatch	The_opponent of_the bills, according to_the unnamed sources, secretly conspires against the_vote.	Who plotted against the vote? — Opponent/Bill/Opponents/Bills
4	Number match, semantic match	The_supporter of_the politician, in_line with_the general trend, hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Politician/Supporters/Politicians
4	Number mismatch, semantic match	The_supporter of_the politicians, in_line with_the general trend, hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Politician/Supporters/Politicians
4	Number match, semantic mismatch	The_supporter of_the regulation, in_line with_the general trend, hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Regulation/Supporters/Regulations

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
4	Number mismatch, semantic mismatch	The_supporter of_the regulations, in_line with_the general trend, hastily suggests forming a_new committee.	Who proposed to create a new committee? — Supporter/Regulation/Supporters/Regulations
5	Number match, semantic match	The_visitor of_the poet, as_far as_I can see, just admires the_spacious room.	Who praised the room? — Visitor/Poet/Visitors/Poets
5	Number mismatch, semantic match	The_visitor of_the poets, as_far as_I can see, just admires the_spacious room.	Who praised the room? — Visitor/Poet/Visitors/Poets
5	Number match, semantic mismatch	The_visitor of_the gallery, as_far as_I can see, just admires the_spacious room.	Who praised the room? — Visitor/Gallery/Visitors/Galleries
5	Number mismatch, semantic mismatch	The_visitor of_the galleries, as_far as_I can see, just admires the_spacious room.	Who praised the room? — Visitor/Gallery/Visitors/Galleries
6	Number match, semantic match	The_observer of_the monkey, as_far as_I can see, just frantically gesticulates to_call for_attention.	Who made signs agitatedly? — Observer/Monkey/Observers/Monkeys
6	Number mismatch, semantic match	The_observer of_the monkeys, as_far as_I can see, just frantically gesticulates to_call for_attention.	Who made signs agitatedly? — Observer/Monkey/Observers/Monkeys
6	Number match, semantic mismatch	The_observer of_the event, as_far as_I can see, just frantically gesticulates to_call for_attention.	Who made signs agitatedly? — Observer/Event/Observers/Events

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
6	Number mismatch, semantic mismatch	The_observer of_the events, as_far as_I can see, just frantically gesticulates to_call for_attention.	Who made signs agitatedly? — Observer/Event/Observers/Events
7	Number match, semantic match	The_favorite of_the investor, as_far as_I know, very openly boasts about his_talents.	Who bragged? — Favorite/Investor/Favorites/Investors
7	Number mismatch, semantic match	The_favorite of_the investors, as_far as_I know, very openly boasts about his_talents.	Who bragged? — Favorite/Investor/Favorites/Investors
7	Number match, semantic mismatch	The_favorite in_the race, as_far as_I know, very openly boasts about his_talents.	Who bragged? — Favorite/Race/Favorites/Races
7	Number mismatch, semantic mismatch	The_favorite in_the races, as_far as_I know, very openly boasts about his_talents.	Who bragged? — Favorite/Race/Favorites/Races
8	Number match, semantic match	The_advocate for_the teenager, as_far as_I'm concerned, overly enthusiastically addresses the_audience.	Who talked to the audience? — Advocate/Teenager/Advocates/Teenagers
8	Number mismatch, semantic match	The_advocate for_the teenagers, as_far as_I'm concerned, overly enthusiastically addresses the_audience.	Who talked to the audience? — Advocate/Teenager/Advocates/Teenagers
8	Number match, semantic mismatch	The_advocate for_the technology, as_far as_I'm concerned, overly enthusiastically addresses the_audience.	Who talked to the audience? — Advocate/Technology/Advocates/Technologies

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
8	Number mismatch, semantic mismatch	The_advocate for_the technologies, as_far as_I'm concerned, overly enthusiastically addresses the_audience.	Who talked to the audience? — Advocate/Technology/Advocates/Technologies
9	Number match, semantic match	The_fan of_the singer, as_far as_I know, still dreams of_an invite to_the private party.	Who thought about the party? — Fan/Singer/Fans/Singers
9	Number mismatch, semantic match	The_fan of_the singers, as_far as_I know, still dreams of_an invite to_the private party.	Who thought about the party? — Fan/Singer/Fans/Singers
9	Number match, semantic mismatch	The_fan of_the board_game, as_far as_I know, still dreams of_an invite to_the private party.	Who thought about the party? — Fan/Board game/Fans/Board games
9	Number mismatch, semantic mismatch	The_fan of_the board_games, as_far as_I know, still dreams of_an invite to_the private party.	Who thought about the party? — Fan/Board game/Fans/Board games
10	Number match, semantic match	The_interpreter of_the ambassador, as_far as_I'm informed, still noticeably stumbles on_difficult passages.	Who experienced difficulties? — Interpreter/Ambassador/Interpreters/Ambassadors
10	Number mismatch, semantic match	The_interpreter of_the ambassadors, as_far as_I'm informed, still noticeably stumbles on_difficult passages.	Who experienced difficulties? — Interpreter/Ambassador/Interpreters/Ambassadors
10	Number match, semantic mismatch	The_interpreter of_the speech, as_far as_I'm informed, still noticeably stumbles on_difficult passages.	Who experienced difficulties? — Interpreter/Speech/Interpreters/Speeches

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
10	Number mismatch, semantic mismatch	The_interpreter of_the speeches, as_far as_I'm informed, still noticeably stumbles on_difficult passages.	Who experienced difficulties? — Interpreter/Speech/Interpreters/Speeches
11	Number match, semantic match	The_gardener of_the landlords, as_far as_I know, still heatedly insists on_waiting another week.	Who wanted to wait? — Gardener/Landlord/Gardeners/Landlords
11	Number mismatch, semantic match	The_gardener of_the landlords, as_far as_I know, still heatedly insists on_waiting another week.	Who wanted to wait? — Gardener/Landlord/Gardeners/Landlords
11	Number match, semantic mismatch	The_gardener of_the park, as_far as_I know, still heatedly insists on_waiting another week.	Who wanted to wait? — Gardener/Park/Gardeners/Parks
11	Number mismatch, semantic mismatch	The_gardener of_the parks, as_far as_I know, still heatedly insists on_waiting another week.	Who wanted to wait? — Gardener/Park/Gardeners/Parks
12	Number match, semantic match	The_painter of_the king, as_far as_I've heard, really wishes for_another commission.	Who desired a commission? — Painter/King/Painters/Kings
12	Number mismatch, semantic match	The_painter of_the kings, as_far as_I've heard, really wishes for_another commission.	Who desired a commission? — Painter/King/Painters/Kings
12	Number match, semantic mismatch	The_painter of_the landscape, as_far as_I've heard, really wishes for_another commission.	Who desired a commission? — Painter/Landscape/Painters/Landscapes

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
12	Number mismatch, semantic mismatch	The_painter of_the landscapes, as_far as_I've heard, really wishes for_another commission.	Who desired a commission? — Painter/Landscape/Painters/Landscapes
13	Number match, semantic match	The_manager of_the musician, as_far as_I can see, very cheerfully signs a_contract for_the next_two years.	Who put the signature on the contract? — Manager/Musician/Managers/Musicians
13	Number mismatch, semantic match	The_manager of_the musicians, as_far as_I can see, very cheerfully signs a_contract for_the next_two years.	Who put the signature on the contract? — Manager/Musician/Managers/Musicians
13	Number match, semantic mismatch	The_manager of_the estate, as_far as_I can see, very cheerfully signs a_contract for_the next_two years.	Who put the signature on the contract? — Manager/Estate/Managers/Estates
13	Number mismatch, semantic mismatch	The_manager of_the estates, as_far as_I can see, very cheerfully signs a_contract for_the next_two years.	Who put the signature on the contract? — Manager/Estate/Managers/Estates
14	Number match, semantic match	The_student of_the professor, as_far as_I know, still categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Professor/Students/Professors
14	Number mismatch, semantic match	The_student of_the professors, as_far as_I know, still categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Professor/Students/Professors
14	Number match, semantic mismatch	The_student in_the course, as_far as_I know, still categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Course/Students/Courses

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
14	Number mismatch, semantic mismatch	The_student in_the courses, as_far as_I know, still categorically refuses to_take the_final exam.	Who withdrew from the exam? — Student/Course/Students/Courses
15	Number match, semantic match	The_critic of_the politician, as_far as_we know, still passionately protests the_results of_the vote.	Who opposed the results of the vote? — Critic/Politician/Critics/Politicians
15	Number mismatch, semantic match	The_critic of_the politicians, as_far as_we know, still passionately protests the_results of_the vote.	Who opposed the results of the vote? — Critic/Politician/Critics/Politicians
15	Number match, semantic mismatch	The_critic of_the proposal, as_far as_we know, still passionately protests the_results of_the vote.	Who opposed the results of the vote? — Critic/Proposal/Critics/Proposals
15	Number mismatch, semantic mismatch	The_critic of_the proposals, as_far as_we know, still passionately protests the_results of_the vote.	Who opposed the results of the vote? — Critic/Proposal/Critics/Proposals
16	Number match, semantic match	The_coach with_the tattoo, as_far as_I'm informed, extremely skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Athlete/Coaches/Athletes
16	Number mismatch, semantic match	The_coach with_the tattoos, as_far as_I'm informed, extremely skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Athlete/Coaches/Athletes
16	Number match, semantic mismatch	The_coach of_the athlete, as_far as_I'm informed, extremely skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Tattoo/Coaches/Tattoos

Table 6.5: Stimuli used in Experiment 2.

ID	Condition	Sentence	Question and response options
16	Number mismatch, semantic mismatch	The_coach of_the athletes, as_far as_I'm informed, extremely skillfully negotiates a_pay rise.	Who bargained about the salary? — Coach/Tattoo/Coaches/Tattoos

Note. The response options are always presented in the following order: correct response; wrong noun, correct number marking; correct noun, wrong number marking; wrong noun, wrong number marking. Option "I'm not sure" is omitted as it is the same in every item. Underscores mark the words that were presented as a single region during self-paced reading. Participants did not see the underscores.

6.2.3 Materials of Experiment 3

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
1	Number match, semantic match	The_singer that the_actor so_openly and profoundly admires apparently received some harsh criticism.	Who felt admiration? — Actor/Singer/Actors/Singers
1	Number mismatch, semantic match	The_singers that the_actor so_openly and profoundly admires apparently received some harsh criticism.	Who felt admiration? — Actor/Singer/Actors/Singers
1	Number match, semantic mismatch	The_play that the_actor so_openly and profoundly admires apparently received some harsh criticism.	Who felt admiration? — Actor/Play/Actors/Plays
1	Number mismatch, semantic mismatch	The_plays that the_actor so_openly and profoundly admires apparently received some harsh criticism.	Who felt admiration? — Actor/Play/Actors/Plays
2	Number match, semantic match	The_trainee that the_supervisor always strongly endorses in_public turns_out to_have been nominated for_an award.	Who provided support? — Supervisor/Trainee/Supervisors/Trainees
2	Number mismatch, semantic match	The_trainees that the_supervisor always strongly endorses in_public turn_out to_have been nominated for_an award.	Who provided support? — Supervisor/Trainee/Supervisors/Trainees
2	Number match, semantic mismatch	The_start-up that the_supervisor always strongly endorses in_public turns_out to_have been nominated for_an award.	Who provided support? — Supervisor/Start-up/Supervisors/Start-ups
2	Number mismatch, semantic mismatch	The_start-ups that the_supervisor always strongly endorses in_public turn_out to_have been nominated for_an award.	Who provided support? — Supervisor/Start-up/Supervisors/Start-ups

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
3	Number match, semantic match	The_legislator that the_opponent secretly and efficiently conspires against turns_out to_be not_so popular after_all.	Who did the plotting? — Opponent/ Legislator/Opponents/Legislators
3	Number mismatch, semantic match	The_legislators that the_opponent secretly and efficiently conspires against turn_out to_be not_so popular after_all.	Who did the plotting? — Opponent/ Legislator/Opponents/Legislators
3	Number match, semantic mismatch	The_bill that the_opponent secretly and efficiently conspires against turns_out to_be not_so popular after_all.	Who did the plotting? — Opponent/Bill/Opponents/Bills
3	Number mismatch, semantic mismatch	The_bills that the_opponent secretly and efficiently conspires against turn_out to_be not_so popular after_all.	Who did the plotting? — Opponent/Bill/Opponents/Bills
4	Number match, semantic match	The_politician that the_vice-president openly and enthusiastically supports in_the campaign enjoys broad international coverage.	Who demonstrated support? — Vice-president/Politician/Vice-presidents/ Politicians
4	Number mismatch, semantic match	The_politicians that the_vice-president openly and enthusiastically supports in_the campaign enjoy broad international coverage.	Who demonstrated support? — Vice-president/Politician/Vice-presidents/ Politicians
4	Number match, semantic mismatch	The_regulation that the_vice-president openly and enthusiastically supports in_the campaign enjoys broad international coverage.	Who demonstrated support? — Vice-president/Regulation/Vice-presidents/ Regulations

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
4	Number mismatch, semantic mismatch	The_regulations that the_vice-president openly and enthusiastically supports in_the campaign enjoy broad international coverage.	Who demonstrated support? — Vice-president/Regulation/Vice-presidents/Regulations
5	Number match, semantic match	The_poet that the_painter apparently regularly visits these days receives a_lot of_attention from the_media.	Who did the visiting? — Painter/Poet/Painters/Poets
5	Number mismatch, semantic match	The_poets that the_painter apparently regularly visits these days receive a_lot of_attention from the_media.	Who did the visiting? — Painter/Poet/Painters/Poets
5	Number match, semantic mismatch	The_gallery that the_painter apparently regularly visits these days receives a_lot of_attention from the_media.	Who did the visiting? — Painter/Gallery/Painters/Galleries
5	Number mismatch, semantic mismatch	The_galleries that the_painter apparently regularly visits these days receive a_lot of_attention from the_media.	Who did the visiting? — Painter/Gallery/Painters/Galleries
6	Number match, semantic match	The_customer that the_consultant unobtrusively but carefully observes probably deserves no_such attention.	Who did the observing? — Consultant/Customer/Consultants/Customers
6	Number mismatch, semantic match	The_customers that the_consultant unobtrusively but carefully observes probably deserve no_such attention.	Who did the observing? — Consultant/Customer/Consultants/Customers
6	Number match, semantic mismatch	The_event that the_consultant unobtrusively but carefully observes probably deserves no_such attention.	Who did the observing? — Consultant/Event/Consultants/Events

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
6	Number mismatch, semantic mismatch	The_events that the_consultant unobtrusively but carefully observes probably deserve no_such attention.	Who did the observing? — Consultant/Event/Consultants/Events
7	Number match, semantic match	The_lobbyist whose candidate openly and eloquently supports cannabis legalization will attract a_lot of_attention.	Who supported cannabis legalization? — Candidate/Lobbyist/Candidates/Lobbyists
7	Number mismatch, semantic match	The_lobbyists whose candidate openly and eloquently supports cannabis legalization will attract a_lot of_attention.	Who supported cannabis legalization? — Candidate/Lobbyist/Candidates/Lobbyists
7	Number match, semantic mismatch	The_party whose candidate openly and eloquently supports cannabis legalization will attract a_lot of_attention.	Who supported cannabis legalization? — Candidate/Party/Candidates/Parties
7	Number mismatch, semantic mismatch	The_parties whose candidate openly and eloquently supports cannabis legalization will attract a_lot of_attention.	Who supported cannabis legalization? — Candidate/Party/Candidates/Parties
8	Number match, semantic match	The_teenager whose advocate enthusiastically and convincingly addresses the_audience had_not enjoyed popularity in_the past.	Who talked to the audience? — Advocate/Teenager/Advocates/Teenagers
8	Number mismatch, semantic match	The_teenagers whose advocate enthusiastically and convincingly addresses the_audience had_not enjoyed popularity in_the past.	Who talked to the audience? — Advocate/Teenager/Advocates/Teenagers
8	Number match, semantic mismatch	The_technology whose advocate enthusiastically and convincingly addresses the_audience had_not enjoyed popularity in_the past.	Who talked to the audience? — Advocate/ Technology/Advocates/Technologies

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
8	Number mismatch, semantic mismatch	The_technologies whose advocate enthusiastically and convincingly addresses the_audience had_not enjoyed popularity in_the past.	Who talked to the audience? — Advocate/ Technology/Advocates/Technologies
9	Number match, semantic match	The_famous_singer whose fan apparently still dreams of_a private behind-the-scenes tour became vastly popular several years ago.	Who thought about the party? — Fan/Singer/Fans/Singers
9	Number mismatch, semantic match	The_famous_singers whose fan apparently still dreams of_a private behind-the-scenes tour became vastly popular several years ago.	Who thought about the party? — Fan/Singer/Fans/Singers
9	Number match, semantic mismatch	The_fashion_label whose fan apparently still dreams of_a private behind-the-scenes tour became vastly popular several years ago.	Who thought about the party? — Fan/Board game/Fans/Board games
9	Number mismatch, semantic mismatch	The_fashion_labels whose fan apparently still dreams of_a private behind-the-scenes tour became vastly popular several years ago.	Who thought about the party? — Fan/Board game/Fans/Board games
10	Number match, semantic match	The_ambassador whose interpreter occasionally noticeably stumbles on_difficult passages will still receive a_warm welcome.	Who experienced difficulties? — Interpreter/ Ambassador/Interpreters/Ambassadors
10	Number mismatch, semantic match	The_ambassadors whose interpreter occasionally noticeably stumbles on_difficult passages will still receive a_warm welcome.	Who experienced difficulties? — Interpreter/ Ambassador/Interpreters/Ambassadors
10	Number match, semantic mismatch	The_speech whose interpreter occasionally noticeably stumbles on_difficult passages will still receive a_warm welcome.	Who experienced difficulties? — Interpreter/Speech/Interpreters/Speeches

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
10	Number mismatch, semantic mismatch	The_speeches whose interpreter occasionally noticeably stumbles on_difficult passages will still receive a_warm welcome.	Who experienced difficulties? — Interpreter/Speech/Interpreters/Speeches
11	Number match, semantic match	The_landlord whose gardener apparently still works by_the_book receives lots_of visitors in_the summer.	Who worked by the book? — Gardener/Landlord/Gardeners/Landlords
11	Number mismatch, semantic match	The_landlords whose gardener apparently still works by_the_book receive lots_of visitors in_the summer.	Who worked by the book? — Gardener/Landlord/Gardeners/Landlords
11	Number match, semantic mismatch	The_park whose gardener apparently still works by_the_book receives lots_of visitors in_the summer.	Who worked by the book? — Gardener/Park/Gardeners/Parks
11	Number mismatch, semantic mismatch	The_parks whose gardener apparently still works by_the_book receive lots_of visitors in_the summer.	Who worked by the book? — Gardener/Park/Gardeners/Parks
12	Number match, semantic match	The_king whose painter always painstakingly strives for_perfection will_be depicted in_a_very flattering manner.	Who wanted to achieve perfection? — Painter/King/Painters/Kings
12	Number mismatch, semantic match	The_kings whose painter always painstakingly strives for_perfection will_be depicted in_a_very flattering manner.	Who wanted to achieve perfection? — Painter/King/Painters/Kings
12	Number match, semantic mismatch	The_landscape whose painter always painstakingly strives for_perfection will_be depicted in_a_very flattering manner.	Who wanted to achieve perfection? — Painter/Landscape/Painters/Landscapes

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
12	Number mismatch, semantic mismatch	The_landscapes whose painter always painstakingly strives for_perfection will_be depicted in_a_very flattering manner.	Who wanted to achieve perfection? — Painter/Landscape/Painters/Landscapes
13	Number match, semantic match	The_musician whose manager carelessly and extravagantly spends the_profits will soon face financial problems.	Who spent the money? — Manager/Musician/Managers/Musicians
13	Number mismatch, semantic match	The_musicians whose manager carelessly and extravagantly spends the_profits will soon face financial problems.	Who spent the money? — Manager/Musician/Managers/Musicians
13	Number match, semantic mismatch	The_estate whose manager carelessly and extravagantly spends the_profits will soon face financial problems.	Who spent the money? — Manager/Estate/Managers/Estates
13	Number mismatch, semantic mismatch	The_estates whose manager carelessly and extravagantly spends the_profits will soon face financial problems.	Who spent the money? — Manager/Estate/Managers/Estates
14	Number match, semantic match	The_fashion_blogger that the_student always passionately reads while commuting covers all_the latest trends.	Who read while commuting? — Student/Fashion blogger/Students/Fashion bloggers
14	Number mismatch, semantic match	The_fashion_bloggers that the_student always passionately reads while commuting cover all_the latest trends.	Who read while commuting? — Student/Fashion blogger/Students/Fashion bloggers

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
14	Number match, semantic mismatch	The_fashion_blog that the_student always passionately reads while commuting covers all_the latest trends.	Who read while commuting? — Student/Fashion blog/Students/Fashion blogs
14	Number mismatch, semantic mismatch	The_fashion_blogs that the_student always passionately reads while commuting cover all_the latest trends.	Who read while commuting? — Student/Fashion blog/Students/Fashion blogs
15	Number match, semantic match	The_politician that the_critic passionately but fruitlessly protests against will nevertheless win the_majority.	Who demonstrated disagreement? — Critic/Politician/Critics/Politicians
15	Number mismatch, semantic match	The_politicians that the_critic passionately but fruitlessly protests against will nevertheless win the_majority.	Who demonstrated disagreement? — Critic/Politician/Critics/Politicians
15	Number match, semantic mismatch	The_proposal that the_critic passionately but fruitlessly protests against will nevertheless win the_majority.	Who demonstrated disagreement? — Critic/Proposal/Critics/Proposals
15	Number mismatch, semantic mismatch	The_proposals that the_critic passionately but fruitlessly protests against will nevertheless win the_majority.	Who demonstrated disagreement? — Critic/Proposal/Critics/Proposals
16	Number match, semantic match	The_young_athlete that the_coach now enthusiastically recommends for_the team will soon become popular.	Who made recommendations? — Coach/Athlete/Coaches/Athletes

Table 6.6: Stimuli used in Experiment 3.

ID	Condition	Sentence	Question and response options
16	Number mismatch, semantic match	The_young_athletes that the_coach now enthusiastically recommends for_the team will soon become popular.	Who made recommendations? — Coach/Athlete/Coaches/Athletes
16	Number match, semantic mismatch	The_training_method that the_coach now enthusiastically recommends for_the team will soon become popular.	Who made recommendations? — Coach/Training method/Coaches/Training methods
16	Number mismatch, semantic mismatch	The_training_methods that the_coach now enthusiastically recommends for_the team will soon become popular.	Who made recommendations? — Coach/Training method/Coaches/Training methods

Note. The response options are always presented in the following order: correct response; wrong noun, correct number marking; correct noun, wrong number marking; wrong noun, wrong number marking. Option "I'm not sure" is omitted as it is the same in every item. Underscores mark the words that were presented as a single region during self-paced reading. Participants did not see the underscores.

6.2.4 Item norming

Mean by-item ratings are presented on Figure 6.2. To analyze Likert scale ratings, we used ordinal ordered logistic mixed-effects regression models. Results of statistical analysis are presented in Table 6.7.

We tested 17 items while only 16 were needed for the experiment, so we decided to exclude item 11 based on the lower mean ratings and personal judgment. The resulting set of experimental items and comprehension questions is presented in Appendix 6.2.

Table 6.7: Statistical modeling of plausibility norming.

	Truncated sentences		Full sentences	
Predictor	Estimate	95%-CrI	Estimate	95%-CrI
Intercept[1]	-4.14	-4.58 – -3.70	-3.84	-4.30 – -3.39
Intercept[2]	-2.97	-3.36 – -2.59	-2.72	-3.11 – -2.32
Intercept[3]	-2.13	-2.48 – -1.77	-2.06	-2.43 – -1.67
Intercept[4]	-1.35	-1.69 – -1.02	-1.05	-1.40 – -0.69
Intercept[5]	-0.53	-0.84 – -0.20	0.05	-0.31 – 0.39
Intercept[6]	0.89	0.58 – 1.23	1.58	1.22 – 1.94
Semantic match	0.06	-0.23 – 0.35	0.25	-0.11 – 0.57
Semantic mismatch	-0.99	-1.45 – -0.47	-1.21	-1.71 – -0.65

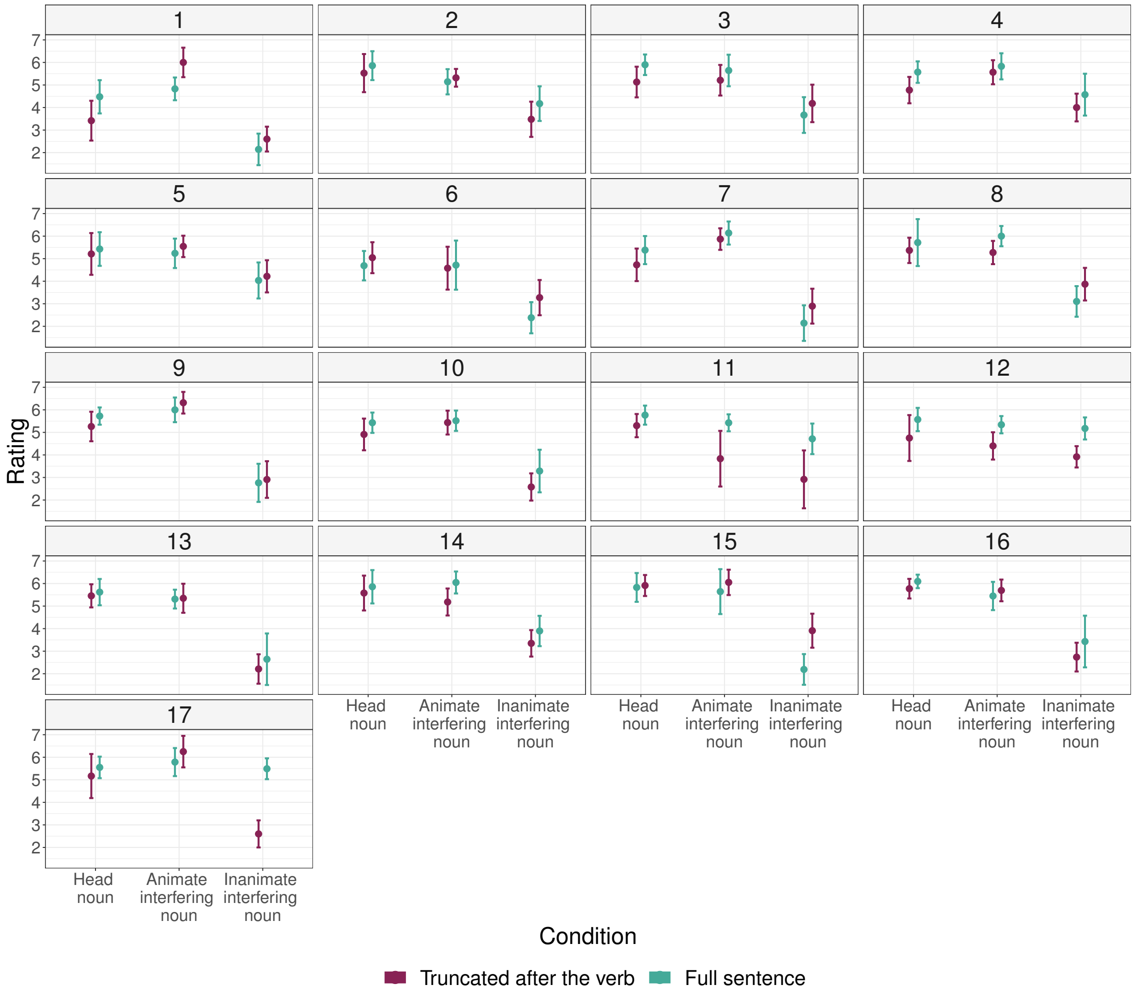

Figure 6.2: Mean rating for each condition across pretests and experimental items. Errorbars represent 95% confidence intervals.