Tel Aviv University
The Lester and Sally Entin Faculty of Humanities
The School of Philosophy, Linguistics and Science Studies

Misanalysis and Reanalysis of Structural Dependencies in Sentence Processing

Thesis submitted for the degree "Doctor of Philosophy"

by

Maayan Keshev

Prepared under the guidance of

Prof. Aya Meltzer-Asscher

November 2020

Acknowledgements

Getting a PhD is a bumpy ride. You find yourself in the driver's seat, leading some research project, and only after you've been going for a while you realize there's no reliable roadmap. The MA has probably already taught you that the path of effort and dedication does not directly lead to success in research the way it did with tests. But as a PhD student you are carving your way in an uncharted territory of your subfield, which means a roadmap is inconceivable. Still, my journey was nonetheless fascinating and invigorating! So I was constantly experiencing a peculiar combination of dread and excitement.

Now as I get to a natural stopping point I look back and see how far I've come. So, at this point, I am delighted to have the opportunity to extend my gratitude to all of you who modelled their ways, who helped me take this path (and carve through it), and who accompanied me in this journey. Thank you for making me the researcher and person I am today.

First goes my advisor, Aya Meltzer-Asscher, to whom I owe my deepest gratitude. Aya, as my professor you introduced me to the field which is now my homecourt. You opened my eyes to a research framework which was different from anything I met before. You pointed me to areas of language research where I could find some answers for my relentless questions. Then you took me in as a grad student, and we started looking for more answers together. For six year now, we've been driving this bumpy road together, backing up in some dead ends, discovering new questions and yet newer places for answer seeking. Your support has always been incredible. You've been there for me, beaming with pride at every success (no matter how small), but also comforting and consoling when I was going through a rough patch. I am so grateful that I got to be your student – I would never have developed the way I did without you. I wish you many more sparkling eyes to guide, you deserve the best of them.

Appreciation is also due to my academic homes over the years: The Linguistics Department at Tel-Aviv University, the Arian de Rothschild program for women doctoral students, and the Adi Lautman interdisciplinary program for excellent students.

The department of Linguistics was a home to me even before I was officially enrolled there and I would like to extend my thanks to my professors here - especially Tali Siloni, Mira Ariel, Julia Horvath, Irena Botvinik, Roni Katzir, and Einat Shetreet - who welcomed, encouraged, and inspired me on various stages of my studies. I would like to thank Mira Ariel also for pushing me to enrol for a PhD in the department, and to Roni Katzir for organizing a joint UMD-TAU workshop way back when (and to all the participants on both sides) - this workshop had more impact than I could imagine.

I also owe many thanks to Naama Friedmann and the professors of the Adi Lautman program, for giving me the opportunity to be part of this amazing group and for granting us a wonderful learning environment. Looking back, this truly transformed me. The people I met there, and program's core values of the integration, independence and in-depth thought, will stay with me for a long time.

The Arian de Rothschild program for women doctoral students has sponsored me in my PhD years. The research presented here was largely enabled thanks to this program's fellowship and travel grants. But this was more than a funding program. The program also set to support us, the fellows, by supplementing our training with time management tips, speech/presentation workshops etc. I benefited a great deal from these activities, and I am proud to have been a part of this community of exquisite women scholars.

On this note, I would also like to thank all those who helped me smoothly cross the bridge to the postdoc jurisdiction – thanks to all the letter writers, application commenters, and interview coaches. And thank you, Brian Dillon, for giving me a new (virtual but very real) academic home even before my postdoc could officially start. The exciting prospect of joining you at UMass helped me to bring about this last lag of the road, and our long discussions of future projects helped me see my old research in a new exciting light..

I have been incredibly lucky to present different parts of this work, on different stages, to many people I look up to. Thus, the work in this dissertation has benefitted from comments of many insightful researchers. First, I would like to thank Whitt Tabor and one anonymous reviewer for evaluating my dissertation and providing numerous thoughtful and perceptive comments. For comments on conference and colloquium presentation of the study presented Chapter 3, I would like to thank Idan Landau, Dave Kush, and Sakshi Bhatia. The study in Chapter 4 has highly benefitted from perceptive, suggestions and outlooks provided by Brian Dillon and Jesse Harris. Lastly, the final version of Chapter 5 is a product of exchanges, with Roger Levy, Ted Gibson, Jesse Harris, Fernanda Ferreira, Adrian Staub, Rachel Ryskin, Martin Van Schijndel, Michal Ben-Sachar, and Gabi Danon. Thank you all for your helpful comments and for your inspiring work cited in this dissertation!

I would also like to use this stage to publicly thank our reviewers over the years. When you work outside the big hubs of your field, reviewers can be an extremely valuable source of expertise. I had the privilege of having some thorough and patient reviewers who pointed me to the right track, without belittling me. Anonymous reviewers forced me to deal with nuances in existing literature that I was less familiar with and have been consistently pushing my science onwards and upwards. We submitted our first paper in 2015, with a simple ANOVA – because after doing all the relevant courses this was still the most advanced method I knew. Since then, anonymous reviewers motivated me to take up R and mixed effects models, to invest more thought in contrast coding, to engage in simulated power analyses, and to move to Bayesian statistics. If it weren't for their suggestions, I don't think I would have gotten this far. On this note, Shravan Vasishth and Bruno Nicenboim also deserve many thanks for the important lessons the taught me at the SMLP summer schools. All errors and faults remain my own.

I owe a great deal to the huge help of research assistants Hila Davidivich, Radan Nassralla, amd Nitzan Trainin, who helped me run vast amounts of participants and code their insane productions. Thanks for making this happen – you are truly amazing!

Additional thanks are also due to the partly overlapping group of my fellow graduate students: Hila Davidivich, Hezi Shabanov, Nitzan Trainin, and Si Berabi. Thank you for taking this non-coffee-drinker/non-smoker on some well needed breaks (in our pre-quarantine life), for your support-group functions (especially you Hila!), and for keeping me in touch with reality. Matan Mazor deserves a special note here: Thank you for your friendship for afar – you somehow manage saying just that right thing every time I get stressed. I would also like to thank my friends from before my University years. Thank you Lee and thank you Amir - for enduring my rants and sharing your own rants with me. You are in my heart even when we are apart for long periods of time.

I'm also take this opportunity to apologize to my family - my mother, my father, and my siblings - for being so unavailable in the months of writing this dissertation. I know these were also though times for you with the ongoing pandemic, and I am very sorry I wasn't there for you as much as you needed. Thank you for understanding and supporting me even though I never managed to fully explain what it is exactly this psycholinguistic research I am doing.

Lastly, Ofir – we are closing this chapter of eight years together at Tel-Aviv and together we will soon begin a new one. You have been my best friend through everything. Thank you for engaging conversations, for your goofy puns and for your loving support (oh and for delivering snacks to my "home office").

Abstract

This dissertation investigates how readers deal with linguistic errors in interpretation, memory, and perception or production of sentences. I explore this topic in three separate studies focusing on main issues in the sentence processing literature: the difficulty of amending an initial erroneous interpretation and the cost-effectiveness of predictive processing; the vulnerability of memory retrieval or encoding processes to interference from similar constituents; and the mechanisms available for compensation for possibly misleading or noisy input.

In the first study, I investigate how readers balance the benefits of predictive processing and the risk of prediction errors. When we take (calculated) guesses in every processing step, we are bound to fail sometimes. Many times, we place our bets on an interpretation which eventually turns out to be the correct one. But other times we will have to retrace our steps and modify our initial interpretation. How do readers overcome the difficulty in initial misinterpretation? How well do they manage to recover from incorrect expectations?

To explore this, I investigate the costs of failed predictions motivated by different considerations - syntactic licensing, rapid semantic integration, or alignment with pragmatic preferences. I explore this within fully grammatical constrictions to examine prediction failure rather than grammatical violations. I observe prediction for the resolution of a long-distance dependency, even in the absence of syntactic licensing pressures. However, I find that when it is motivated by syntactic pressures, predictive dependency formation has earlier and higher costs at disconfirmation sites. Moreover, disconfirmation of syntactic predictions exhibits higher semantic persistence of the initial interpretation relative to predictions motivated by pragmatic preferences. I argue that the difference in reanalysis costs can be attributed to degrees of commitment to the prediction. Specifically, I propose that active dependency formation is faster or more binding in syntactically-motivated predictions relative to pragmatically-motivated ones.

In the second study, I investigate memory interference in sentence processing, focusing on agreement attraction, as well as the mechanisms which may compensate such agreement errors. Representational or retrieval interference commonly cause errors in processing of agreement relations. Thus, it might be beneficial for the parser to try and minimize the role of accessing distant agreement information. To examine this possibility, I tested the processing of reflexive pronouns, which require identifying a feature-matched subject. Potentially, consulting the agreement features on the (linearly proximal) verb can provide a "shortcut" or updating point, as the verb agrees with the subject.

Using gender agreement on Hebrew verbs and reflexives, I exhibit that the computation of agreement features at the reflexive pronoun highly depends on the availability of agreement cues on the verb. I find that (i) when the verb mismatches the subject (i.e. following ungrammatical verbs), a reflexive which matches the verb is preferred over one which matches the subject; and (ii). Agreement attraction is less prominent when the verb manifests agreement cues. I interpret this as evidence that verbal agreement either alters the representation of the subject, or makes retrieval of the subject's features unnecessary (i.e. the reflexive does not re-access the subject when verb agreement is available, but verifies verb-reflexive agreement). These strategies might be beneficial for processing as they allow the parser to reduce the reliance on the representation of the early noun phrase.

In the last study, I present novel evidence for rational noisy-channel inference in incremental processing. I use Hebrew relative clauses in which the subject position is vacant, and the verb mismatches the filler in number. Thus, at the verb, the parser may construct a subject relative, ignoring the agreement mismatch, or choose a less frequent structure of an object relative (either assuming a subject-verb inversion or a null impersonal subject).

I show that the effect of agreement mismatch is modulated by the prior probability of the alternative analysis. I find evidence that a corrupted subject relative parse is preferred over assuming a rare word order (a subject-verb inversion). This suggests that readers are willing to ignore the agreement information. Yet, I also find evidence that readers attempt to remain faithful to the input, when the grammatical analysis is not extremely rare. Readers prefer constructing a structure which is grammatically pristine and relatively common (a null impersonal subject), over the more frequent yet corrupted subject relative reading. These findings suggest that readers rapidly apply elaborate probabilistic knowledge regarding the distribution of syntactic structures in their language. Specifically, readers do not neglect grammatical information due to shallow processing, but as a rational strategy for avoiding highly improbable utterances.

Over all, the findings in the dissertation present a complex system of "checks and balances" which can help comprehenders achieve successful interpretation given the prospect of various errors. Thus, the reported studies contribute to our understanding of the human comprehension mechanism and suggest several ways in which the system might compensate (during incremental processing) for different pitfalls in language processing like prediction errors, memory slips and faulty input.

Table of Contents

Preface

Parts of the research presented in this dissertation were submitted and published as peer-reviewed journal articles. Experiments 1-3 were published as a paper in the journal *Language, Cognition, and Neuroscience* (Volume 35, Issue 2). A different version of Chapter 5 was published in *Cognitive Psychology* (Volume 124, 101359). On both papers I am the first author, and they were written and submitted with my advisor, Dr. Aya Meltzer-Asscher, as a co-author.

Chapter 1

Introduction

1.1 On language and errors

Human communication is magnificent yet treacherous. It is right in front of us day by day, and yet its underpinnings are mysterious. How do we manage to transfer ideas from one mind to another using a dense string of visual symbols or a convoluted series of sounds? On both the producer's side and the comprehender's, language is a complex task to navigate. From visual/auditory stimuli, through linguistic form, to the intended message, language is full of ambiguities and pitfalls. Some mistakes are bound to happen. Indeed, errors in speech and hearing, or in typing and reading, are dangerously common. Yet, most of the time we manage to understand what our interlocutor is trying to convey. Errors are thus one of the most intriguing phenomena in human communication, exactly because we manage to move beyond them with seemingly little effort.

The psycholinguistic study of sentence processing examines step-by-step the interpretation of language and aims to create a model of the obstacles, and the strategies adopted to overcome them, in this process. These topics are studied though experimental measures like reading times, comprehension questions, speeded decision making, and physiological responses. A prominent line of research within this framework is concerned with the analysis of sentence structure and word-by-word construction of linguistic dependencies. Dependency formation means linking words which rely on each other for grammatical licensing (e.g. the verbal form must agree in number with its subject) and for creating an interpretation (e.g. understanding who did what to whom requires linking the verb with its arguments). This process in one of the most basic building blocks of sentence comprehension as it enables creating a syntactic and semantic representation of the sentence by connecting phrases which belong together linguistically.

Sentence processing, and especially the processing of dependencies, requires quick application of knowledge of different types of information (e.g. grammatical constraints, language probabilities, and real-world plausibility), usage of different levels of representation (e.g. syntactic structure, semantic meaning and pragmatic inference), and simultaneous analysis of different parts of the utterance (e.g. retrieving phrases which were previously

processed). This complex operation requires accommodation of different strategies to allow for fast language communication. Such heuristics and rapid processing mechanism are inclined to some errors.

Errors are a great resource for cognitive research and for uncovering the working mechanisms of our language system. Linguistic errors are like missteps of a magician, invisible threads caught in the act, a glimpse of the hidden door. They expose the fragile and intricate nature of the show and tell us how the magic happens, and they thus found a home in the minds and hearts of many psycholinguists. This dissertation is concerned with errors in language comprehension – errors resulting from disconfirmed predictions, those resulting from memory faults, and rational "errors" which aim to correct for other possible mistakes.

In the rest of this introduction chapter, I present three types of "errors" in sentence processing and how they are studied in the field of sentence processing. I focus on: (i) Reanalysis processes, which occur due to inaccurate predictive processing (erroneous initial interpretation, based on partial or ambiguous input). (ii); Breakdowns in memory representations or operations which arise during word-by-word sentence interpretation; (iii) Readers online rational adaptation to the possibility of errors in the text. These three overviews lay the ground for the three studies reported in this dissertation (Chapters 4-6, correspondingly).

1.2 Reanalysis: Predictive processing and ambiguity resolution

Imagine that the we could take as much time as we needed to thoroughly analyze each utterance. In such a world, the task of inferring the sentence structure and assigning a meaning to it could be accomplished by applying the grammatical rules of the language in a systematic way. Clearly, however, human communication is time-bound. Effective communication and interaction require responding in a timely manner. Moreover, since our short-term memory cannot hold much information simultaneously, incoming material rapidly decays unless it is quickly structured (Atkinson & Shiffrin, 1968; Cowan, 1988, 2001; among others). Thus, our restricted memory capacity highly constrains the parser's function and running time.

One strategy used by comprehenders to speed up sentence interpretation is predictive processing. Readers and listeners do not wait to the end of a sentence in order to construct its interpretation. On the contrary, they extract information from their current (partial) input, and actively aim at predicting the next constituents. Predictive processing is manifested in diverse realms of human cognition, including perception, action, attention, and learning. It was even suggested that the human brain is designed to constantly anticipate the upcoming sensory states of the organism and minimize surprise as to changes in the environment (Friston 2009; 2010).

Thus, predictions and anticipatory processes are common in various levels of linguistic analysis. As we read or listen to a sentence unfold, we constantly make assumptions based on partial and ambiguous input. We decide how to interpret what we have heard, and what we believe will come next, even though we do not have enough information to confidently make the right decision. We manage this so quickly and effortlessly, that we can finish each other's sentences, or cut into an ongoing speech turn (at least in some cultures). Similarly, as speakers, we sometimes count on this predictive processing, leaving our sentence not entirely…

Sometimes the comprehender's predictions are borne out, but other times they would have to change their initial interpretation as the sentence unfolds. When our predictions are disconfirmed, we have to backtrack our steps and amend the analysis we formed (or re-rank the various interpretations we considered, in a parallel processing framework). This process is called reanalysis. Reanalysis processes are necessary for accurate language comprehension and are employed frequently. Altering processing commitments may be challenging in terms of linguistic analysis, working memory and inhibition processes. Yet, reanalysis is often automatic and invisible to the naked eye. Like most errors, it goes unnoticed.

We consciously experience reanalysis when it produces a humorous effect. Consider the following example, taken from an Israeli satirical television show. One of their sketches visits a fictional hair salon which notoriously overprices everything. As the hair dresser's assistant tries to sell an expensive product to the consumers, he utters the sentence in (1). To the amusement of the viewer, the essence of the conning salesperson is embodied in the contrast between the initial interpretation, where the shampoo is custom-made, and the final meaning, where only the label is switched to allow overpricing. This is a great example of premature resolution of structural ambiguity (as to the relation between the shampoo and the embedded verb, 'make').

(1) *Ze shampo she-mexinim bimyuxad bishvileno et ha-mdbekot shelo*

'This is a shampoo which they make especially for us the stickers of'

Eretz Nehederet, Season 16, episode 11

Example (2) is another nice example of failure at predicting the resolution of a sentence, and specifically, failed lexical prediction. This line, form the Disney film Frozen, captures the situation of a snowman, Olaf, who naively expects to have a great time as the weather gets warmer. He sings, as he jumps above a small pond, and the contrast between the predicted word (puddle) and the actual ending of the sentence gives a sense of irony.

(2) Winter is great time to stay in and cuddle, but in the summer, I'll be a…. happy snowman!

Frozen, from the song "In Summer"

What makes reanalysis easy or difficult? How do people manage to balance the benefits in stipulating the sentence interpretation, with the possible costs of reanalysis? Is there a difference between interpretations based on ambiguous input and those based on active prediction? These questions can be answered only under the magnifying glass of experimental research.

1.2.1 Ambiguity resolution and reanalysis

Sentences in natural language often present ambiguities that cannot be resolved immediately. Words can be ambiguous with respect to their meaning (e.g. the word *bank* may be interpreted as a financial institution or as a riverside) and/or their grammatical role in the sentence (e.g. the word *fall* may be analyzed as either a verb or a noun). Moreover, even if all lexical items are unambiguous, relations within the sentence structure can also be ambiguous. The sentence in (3) may mean that Mary told a story to a lawyer and that the judge likes this lawyer (3a), or that the judge likes the story and Mary told that to the lawyer (3b). In this case the sentence is globally ambiguous, but in other cases information which occurs later may resolve the ambiguity.

(3) a. Mary told the lawyer$_1$ [the judge liked $_1$] the story.
b. Mary told the lawyer [the judge liked the story].

The identification of strategies for resolving ambiguities gave rise to heated debates over the years and started off much of the psycholinguistic study of sentence processing. Thus, much research in psycholinguistics has been devoted to mapping parsing choices at points of local ambiguity in natural language. **Structural ambiguities** are resolved using top-down parsing, which may be guided by economy-based heuristics (Frazier, 1979; and later on - Gibson, 1998) or by weighing frequency and context biases (MacDonald, Pearlmutter & Seidenberg, 1994; and later on - Levy, 2008a). For example, on initially reading the verb *sent* in sentence (4), the structure is ambiguous between a main verb reading (where *the candidates* are the senders) and an embedded passive reading (where *the candidates* are the recipients). A preference for the main verb interpretation could reflect minimization of the syntactic structure (the embedded structure requires a more complex syntactic tree), or occur due to the low frequency of the passive form.

(4) The candidates sent the acceptance letters were happy.

Such parsing preferences are traditionally deduced from processing disruption in sentences where the dispreferred interpretation turns out to be correct. In (4) we need to create a new interpretation, or revive one that received less consideration, once we arrive at the main predicate *were happy*. This entails increased processing costs (e.g. slow reading times). Thus, in such cases, a disruption of reading is treated as an index for **structural reanalysis**, which is required if the wrong structure was chosen. Misleading sentences, like the one in (4) were termed **Garden Path** (GP) sentences and were extensively tested since the early days of sentence processing research (e.g. Frazier & Rayner, 1982; Rayner & Frazier 1987; Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994)

Another type of structural reanalysis occurs in **filler-gap dependencies**. Filler-gap dependencies are predominately manifested in wh-questions (5a) or relative clauses (5b), and present a constituent (the filler, underlined) dislocated from its canonical thematic position (the gap, marked by an underscore).

(5) a. Which student was the teacher waiting for __ at the office?
b. This is the student who the teacher was waiting for __ at the office.

Filler-gap dependencies invariably exhibit local ambiguity as to the gap's position. For example, at the first verb in sentence (6), comprehenders may assume a gap at the direct object position as in (6a), or wait for a later resolution as in (6b). This ambiguity (at least partly) originates form the fact that the gap is a null element. Therefore, readers and listeners might not know whether it had already appeared in their input or not. They are left with the choice of assuming a gap without unequivocal evidence, or waiting to a later stage where they will be able to review the sentence fully and locate unfulfilled positions or missing constituents (e.g. a missing optional/obligatory object).

(6) Which student did the teacher persuade { a) __ to speak up in class? / b) the principal to punish __ ?

When faced with such ambiguity, the parser tends to postulate a gap as soon as possible, prior to any direct evidence in the input. This was termed **the Active Filler strategy** by Frazier (1989). The Active Filler strategy was exhibited using a variety of experimental methods (Boland, Tanenhaus, Garnsey, & Carlson, 1995; Garnsey, Tanenhaus, & Chapman, 1989; Sussman & Sedivy, 2003; Traxler & Pickering, 1996) and in different languages (Aoshima, Phillips, & Weinberg, 2004; Bourdages, 1992; de Vincenzi, 1991; Frazier & Flores d'Aracais, 1989; Schlesewsky, Fanselow, Kliegl, & Krems, 2000). This predictive dependency formation may result in a need for reanalysis when another argument occupying this position is later encountered. Such a reanalysis was revealed in a self-paced reading study by Stowe (1986) and termed the **Filled Gap Effect** (FGE). Stowe observed elevated reading times on the word occupying the potential gap position (the embedded object, *us*, in example 7). Following this, many studies have replicated this reanalysis effect.

(7) My brother wanted to know who Ruth will bring us home to _ at Christmas.

1.2.2 Other types of predictive sentence processing

Active inference is not unique to structural ambiguities. It extends to cases in which the input is not ambiguous but has yet to be completed. Many times, we anticipate the continuation of the sentence, rather than process words as they appear. This is reflected in various psycholinguistic findings and theories.

One classic piece of evidence for next-word expectations comes from an eyetracking study by Altman and Kamide (1999). In this study participants looked at an array of objects while hearing a sentence (a method called the visual world paradigm). Upon hearing the verb, participants exhibited anticipatory eye movements, preferentially looking at potential arguments of the verb. For example, when hearing a verb with selectional restrictions like *eat* they looked at its only plausible object (e.g. a cake) at a higher rate than following verbs which allow more possible objects (e.g. *move*). This findings was also extended to combinatory restrictions, including restrictions due to the combination of the agent and the verb (e.g. *girl* and *drink* may favor milkshake while *man* and *drink* may favor beer; Kamide, Altmann, & Haywood, 2003), and due to the tense of the verb (e.g. looking at an empty glass following the preamble *the man has drunk*, but at a full glass after *the man will drink*; Altmann & Kamide, 2007). These studies show that listeners incrementally anticipate an upcoming constituent, or at least its semantic domain, based on rapid integration of linguistic input and context.

Another line of research on prediction exhibits the processing ease at predictable words. Without a visually constraining context, a given sentence may be completed in many possible ways. Yet those completions differ in their probability. Much research has shown that more probable words are associated with faster reading relative to less predictable or anomalous words (Ehrlich & Rayner, 1981; Rayner & Well, 1996; 1999; McDonald & Shillcock 2003; Kliegl, Nuthmann, & Engbert, 2006; Smith & Levy, 2013). Similarly, studies tracking readers' eye movement found that predictable words are skipped more often (Balota, Pollatsek, & Rayner, 1985; Rayner & Well, 1996; Rayner, Slattery, Drieghe, & Liversedge, 2011). Lastly, electroencephalography studies (EEG, or ERP) reveal that the amplitude of the N400, an ERP component which was initially associated with semantic incongruency (Kutas & Hillyard, 1980), is sensitive to graded changes in predictability (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984; Delong, Urbach & Kutas, 2005; Wlotko & Federmeier, 2012). This component inversely correlates with the word's cloze probability, a measure which reflects the percentage of participants who produced this specific word at a preamble completion task.

Note, however, that these ease-of-processing effects are consistent with prediction but can also arise without predictive retrieval of the word itself. These effects could reflect processes that occur after reading the supposedly predictable word, rather than before it, like ease of integration. Integration of the current word into the preceding sentential context may passively benefit from overlapping activation, a residue of the previous words in the context (see e.g. Van Petten & Luka, 2012 for discussion of benefits for confirmed expectation and prediction commitment).

Yet, there are also evidence for prediction of a word's linguistic features which are not compatible with the above integration view. Anticipating a specific grammatical gender, for example, may indicate direct activation of the lexical entry because synonymous words, or

equally fitting words, may have different gender. Indeed, different studies exhibited ERP effects associated with articles and adjectives which mismatch the gender of the expected noun, before the noun itself is presented (In Spanish - Wicha, Bates, Moreno, & Kutas, 2003; Foucart, Martin, Moreno, & Costa, 2014; Martin, Branzi, & Bar, 2018; In Dutch – Otten, Nieuland, & Van Berkum, 2007; Otten & Van Berkum, 2009; In Italian – Ito, Gambi, Pickering, Fuellenbach, & Husband, 2020). The specific pattern of ERP correlates varies between these studies, and the effect of gender prediction failed to replicate in a large-scale study (Nieuwland et al., 2018). Yet, these effects may still imply prediction of word-form properties (see Nicenboim, Vasishth, & Rösler, 2020 for a recent meta-analysis).

1.2.3 Does disconfirmation of non-structural predictions elicit processing difficulty?

While there seems to be good evidence for prediction outside the realm of structural ambiguities, disconfirmation of these predictions might not be as costly. As many have noted, predictive processing should be reflected not only on ease of processing, but also in costs of failed predictions. If we commit to predicting a specific word, the processing of a different completion would become more straining, relatively to a case where no binding prediction was made. We should expect difficulty following a disconfirmed prediction, as a parallel to effects of structural reanalysis. Yet, evidence is rather mixed with regards to the costs of such lexical "reanalysis".

In the search of prediction costs, the **sentential constraint** comes into play. Sentences differ with regards to the probability of their most favored completion. The higher this probability is, the more constraining the sentence frame is (e.g. consider the difference between *we went to the grocery store to buy…* , and *since the kids wanted to sell some freshly squeezed juice, we went to the grocery store to buy*…). If the parser commits to high probability completions, a word which is congruent but unlikely (low cloze), should be harder to process in a constraining context (i.e. one for which there are very few plausible continuations) relative to a sentence which does not imply any specific continuation (i.e. a low constraint sentence which does not entail lexical prediction).

Response times and eyetracking studies seem to provide little or no evidence that making a wrong prediction causes a slowdown, relative to making no prediction at all. Different studies have failed to observe evidence for disruption in reading times correlating with the sentential constraint. For example, Frisson, Harvey, and Staub (2017) observed similar reading times for unpredictable words in high constraint contexts, which should generate prediction of a different word, and low constraint contexts, which do not support prediction of any specific word. Traxler and Foss (2000) obtained similar results with a naming task. Lastly, Luke and Christianson (2016) faced similar null effects in a large correlation study. These authors conducted a large-scale cloze survey, mapping predictability for every word of multiple naturalistic text passages. They tested the difficulty of all low cloze words using eye tracking while reading. This study failed to detect a correlation with the predictability of the most probable word, namely with the constraint of the sentence. These findings suggest that recovering from an incorrect prediction does not tend to cause a difficulty detectable in behavioral measures.

Most studies which demonstrate effects of the sentence's constraint (independent of the cloze probability of the word which actually appears in the sentence) come from the ERP literature. Within this framework, recent studies revealed a component associated with integrating a new plausible word into a constraining context which encourages prediction of a different word. As Van Petten and Luka (2012) note in their thorough review, the N400 component does not index costs of incorrect predictions. Rather, a frontal positivity, which frequently follows the N400 component, may reflect disconfirmed predictions. Federmeier, Wlotko, De Ochoa-Dewald, and Kutas (2007) crossed the predictability of the presented word and the sentence's constraint. They found that unexpected (but congruent) words were associated with a late frontal positivity, only when another word was predicted by the context (i.e. only in high constraining contexts). This finding was replicated and extended by Thornhill and Van Petten (2012), who found that the post-N400-positivity can target low cloze words which are related to the predicted one. This suggests that this ERP component reflects a failure of lexical prediction rather than a residue of conceptual pre-activation. If prediction would have targeted only the semantic features, without the word's form, low cloze completions which are semantically similar to a high cloze word should not be regarded as disconfirmation.

Yet some claim that these ERP effects, associated with making the wrong prediction, might not reflect processing difficulty per se (Pickering & Gambi, 2018). As the ERP component arising for low cloze words in high constraint sentences does not seem to be associated with additional effort in behavioral measures, we cannot unequivocally state that the cognitive operation this component reflects comes with a price (i.e. that more effort is needed for failed predictions, relative to no prediction conditions). Pickering and Gambi (2018) even argue that the relatively little evidence of processing difficulty in these cases compared to reanalysis costs may indicate that lexical predictions are highly cost-effective for the parser.

1.3 Unintended misanalysis: Linguistic illusions and memory fallibility

Language processing relies not only on predicting the upcoming input but also on revisiting previous parts of the input, and this process is vulnerable to errors too. Many times, we are required to access a representation of phrases we have already processed. This is crucial for successful language comprehension, as it allows establishing dependencies between non-adjacent words and phrases. Such dependencies are very common in natural language. For example, consider the sentence in (8). Here the head of the subject phrase *dissertation* has to be integrated with the corresponding verb *discusses*, across a whole embedded clause, consisting of six words. The process of integrating distant words, in both syntax and meaning, offers challenges due to the limitations of the human memory capacity.

(8) The dissertation which you are reading right now discusses various processing errors.

The constraints of our memory system are especially salient when it comes to noticing detail and individual elements, rather than forming a gist out of the information stream. For example, people largely fail to notice (small and large) changes to a busy visual scene. In one of the most prominent change detection studies, Rensink, O'Regan, and Clark (1997) presented participants each time with pair of flickering images, separated by a brief blank screen. The images continued to alternate rapidly until participants detected the change. Although, by the nature of the task, participants knew that something was changing right before their eyes, changes as large as the emergence/disappearance of a bulky building in the background, went unnoticed through numerous iterations. This change blindness demonstrates the "limits on our capacity to encode, retain and compare information" (Simons & Ambinder, 2005).

Similar to inspection of visual scenes, language comprehension requires extracting an overall interpretation from multi-featured input. To establish subject-verb relations, for example, we have to locate a noun which appears in the correct syntactic position, which matches the verb in number and person, and which is a plausible agent of the verb. Similar processes apply to other dependency types as well. Thus, each dependency hinges on different types of syntactic, semantic, or discourse constraints. The manifold information types mean that accurate sentence interpretation requires precise memory of various features regarding each word, and faithful access to it.

Moreover, as in the case of change detection, the only way to keep track of details in the input is by accessing memory representations. Words are uttered and gone in a conversation; text falls out of our foveal focus as we read the next words. Thus, tracking the dynamic unfolding of a sentence requires careful memory and attention management. The parser may need to recover the appropriate information after possibly lengthy processing of other intervening material. For example, in (8), the subject and the verb are separated by many additional words which require memory maintenance and various parsing computations. At the point of encountering the relevant verb, the representation of the subject head may have decayed, due to the intervening material and small number of items which can be active in working memory.

This reliance on memory encoding and retrieval can be a source of many errors. A constituent may be encoded ineptly. Thus, the parser may initially register wrong information regarding the target item. In addition, the target representation may endure accidental overwriting of some features or just gradually fade from memory. Lastly, retrieving the target from memory is also at risk, as the parser may pick the wrong constituent at the critical moment. How reliable are these memory mechanisms? How well do we manage to get to the end of a sentence with a precise memory of phrases that are no longer available in our direct input? A growing number of studies have suggested that language comprehension is often susceptible to memory fallibility. In particular, the role of interference in encoding and retrieval has drawn much attention, as elaborated on below.

1.3.1 Interference effects in memory mechanisms

Over the years, many researchers shared the intuition that intervening and/or similar constituents interfere with linguistic dependency formation. In his model, Gibson (1998, 2000) suggests that a main determinant of processing difficulty is integrating a new word into the structure of the sentence, and that such integration gets more complex when the distance between the two dependent elements is long. Gibson formalizes this distance in terms on the number of discourse referents in the intervening region which anchors the complexity of the dependency with crossing prominent information.

Another early intuition regarding interference was suggested by Lewis (1996). Lewis claimed that the working memory of the parser is able to hold only two items of the same syntactic relation/position. Thus, if, for example, three nouns occupy sentential subject positions, one of them will be compromised. This might create processing failure later on, if this noun has to accessed for some operation. This constraint, he suggests, account for the incomprehensibility of center embedded sentences as in (9). Thus, Lewis turns the focus from the complexity or abundance of intervening material to **similarity-based interference.**

(9) The grandmother [who the wolf [who little red riding hood met] ate] was later rescued.

Similarity-based interference was argued to apply also in relatively simple sentences. Gordon, Hendrick, and Johnson (2001, 2004) exhibited experimental evidence for the effect of noun phrase similarity on reading times of relative clauses. Their research revealed that having two noun phrases of the same type (lexical nouns, indexical pronouns, or proper names) incurred processing costs in object relative clauses. [1]

---

[1] Similarity-based interference manifests to some extent also in syntactic theory, within Rizzi's Relativized Minimality (1990, 2004). This theory states that syntactic relations should be formed in configurations where there is no intervening element which is in a similar syntactic position as the target antecedent (see i). This aims to derive, for example, the ill-formedness of island structures as in (ii) from the similarity between the two wh-phrases.

(i) Given the configuration: ... X ... Z ... Y ...
Y is in a minimal configuration with X, if and only if there is no Z such that:
Z is of the same structural type as X, and
Z intervenes between X and Y (c-commands Y and does not c-command X) (Rizzi, 2004)

(ii) What do you wonder who could solve _ ?
X Z Y

Later, Lewis and his colleagues further developed the notion of similarity-based interference into the **Cue-Based Retrieval** model (McElree, 2000 Van Dyke & Lewis, 2003; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Martin & McElree, 2008). The architecture of this model suggests that linguistic constituents are stored in memory as bundles of feature-value pairs. For example, in the example introduced at the beginning of this section (repeated in 10 for convenience), the subject headed by *the dissertation* will be represented as [nominal: yes, subject position: yes, singular: yes, animate: no]. During sentence processing, incoming words, which need to be integrated with the preceding structure, trigger a search for specific previous constituents. Thus, the verb will initiate a memory search for a constituent which can complete the required subject-verb dependency. This search is guided by feature-values that the current dependency requires. For example, in sentence (10) the verb *discusses* will aim to retrieve a target with features like [noun: yes, subject position: yes]. Such features are called retrieval cues.

(10) The dissertation which you are reading right now discusses various processing errors.

Since retrieval is based on various syntactic and semantic features, some items may present as partial matches in the search process. Such items are called **distractors**, and may resemble the target in category, structural position, agreement features or meaning. Distractors may pose a problem to the parser whether they appear before or after the target (i.e. they do not have to be positioned between the target and the retrieval trigger to create interference).

When a cue matches more than one item (i.e. when a distractor is present), retrieving the target gets more difficult as the distinctiveness of the target is dimmed. This can be manifested in slow reading times, and is called **inhibitory interference.** A large body of studies has detected such increased processing difficulty. Van Dyke (2007; also Van Dyke & McElree, 2011) exhibited disruption of processing when constituents intervening between a verb and its subject overlapped with the subject's syntactic or semantic features (with more prominent interference when it is based on syntactic position over animacy cues).

There is another type of interference that Cue-Based Retrieval predicts. When the sentence is ungrammatical sometimes no item matches all retrieval cues and their feature specifications. In this case, multiple partial matches may cause **facilitatory interference.** On the Cue-Based Retrieval model, such a process, involving only partial matches, acts as a race. One cue tries to access one item in memory, while access to another item is attempted by a different cue. The first item to be retrieved terminates the parallel search. This means that on average such a parallel search is faster than in cases where there is only one partial match. This is called statistical facilitation (Raab, 1962) and, in this context, it predicts speed-up in reading time in ungrammatical sentences if there is a suitable distractor.

An ungrammatical or incoherent sentence might even be mistaken as acceptable due to the architecture of this retrieval process. Indeed, various studies exhibited that during the early stages of processing comprehenders appear to accept sentences which they would judge as unacceptable given enough time. Such haphazard or deceptive licensing is called **linguistic** (or grammatical) **illusions** and may target different types of dependencies including agreement, semantic selectional restrictions, and negation.

1.3.2 Agreement attraction

The most well-studied case of linguistic illusion arises in number agreement of subject-verb dependencies. It has been repeatedly shown that the processing of verbal agreement is vulnerable to interference from structurally irrelevant distractors. Ungrammatical verbs may be perceived as well-formed depending on features of non-subject nouns. For example, the plural verb in (11b) may appear more acceptable than in (11a) even though in both cases it mismatches the head of the subject phrase (the singular *key*). This effect was first identified in production (Bock & Miller, 1991), where people use erroneous verb forms more frequently following a nominal configuration as in (11b) relative to (11a). Later on, such effects were observed also in comprehension. In the same environments where individuals are likely to produce an agreement error, comprehenders are also likely to miss such errors (Pearlmutter, Garnsey, & Bock, 1999). This is reflected in acceptability judgments and in relatively fast reading times of ungrammatical verbs (more similar to that of grammatical verbs than to ungrammatical ones as in 11a).

(11) a. The key to the cabinet <u>are</u> rusty.
b. The key to the cabinets <u>are</u> rusty.

Effects of agreement attraction in comprehension were replicated across different languages, including Spanish, Slovak, Russian, Turkish and Arabic (Badecker & Kuminiak, 2007; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Tucker, Idrissi, & Almeida, 2015; Slioussar, 2018; Lago, Gračanin-Yuksek, Şafak, Demir, Kırkıcı, & Felser, 2019;), and in various experimental measures, including self-paced reading, eye tracking, ERP, binary and scaled judgments (e.g. Pearlmutter, Garnsey, & Bock, 1999; Staub, 2009; Wagers, Lau, & Phillips, 2009; Tanner, Nicol, & Brehm, 2014). It even appears in jabberwocky sentences (semi-artificial language in which content words are replaced with pseudo-words, while function words and infliction are retained; Franck & Wagers, 2020).

A robust body of research has identified different factors which modulate the susceptibility of agreement to such illusions and errors. These include asymmetry in the "attractiveness" of singular and plural features (Eberhard, Cutting, & Bock, 2005), and the syntactic depth of the distractor (e.g. within a prepositional phrase, as in 11, or embedded in another relative clause like in 12a; Bock & Cutting, 1992). In addition, agreement attraction is found also in cases where the distractor precedes the target subject as in (12b). Namely, such interference effects can occur even when the distractor does not form a locally coherent string with the verb (Bock & Cutting, 1992; Solomon & Pearlmutter, 2004).

(12) a. The key that opened the cabinets <u>were</u> rusty
b. The cabinets that the key <u>open</u> will be useful

Overall, these advances suggest that agreement attraction is not a result of shallow processing, but reflects a principled fallibility of the memory mechanism. Different accounts have been suggested for agreement attraction in comprehension, targeting the process of accessing the correct constituent in memory, or the encoding of the linguistic feature representation.

The Cue-Based Retrieval model (Lewis & Vasishth, 2005), described above as a general mechanism of interference, has been proposed as an account for agreement attraction in

comprehension (Wagers, Lau, & Phillips, 2009). This approach unites agreement attraction effects in comprehension with other cases of interference in sentence comprehension (see Jäger, Engelmann, & Vasishth, 2017, for a recent review). On this account, the parser attempts at accessing the subject item in memory, using cues derived from the ungrammatical verb. These cues result in multiple partial matches but no full match. Therefore, sometimes the "wrong" partial match is chosen, and the parser retrieves the distractor rather than the target noun. This accounts for the relatively high acceptance rate in judgment data. Cue-Based Retrieval (and specifically its activation model implementation, Vasishth, Nicenboim, Engelmann, & Burchert, 2019) explains the speed-up in processing time using the "race" process described above. This account, however, does not straightforwardly capture agreement attraction effects in production, where the verb is generated by the speaker and not provided from the input along with its feature cues (But see Badecker & Kuminiak, 2007).

The Marking and Morphing model (Eberhard, Cutting, & Bock, 2005), on the other hand, attributes attraction to the representation of agreement, rather than to the process of accessing it. In this model, attraction results from defective encoding of the subject's features. The number of the subject phrase is calculated not by the morphosyntactic number of the head noun, but based on two successive processes: (i) inferring the notional number of the subject phrase (Marking), and (ii) spreading activation associated with the number morphology in the sentence (Morphing). Therefore, the evaluation of the subject's morphological number might be disrupted by a distractor activating the plural morphology. The result is equivocal number marking in the subject's representation, and thus errors in production and illusions in comprehension. Agreement computation, on this account, fails to distinguish the sources of contradicting agreement information, during initial encoding of the subject rather than in the attempt at retrieving it. While this approach does not align agreement attraction with other effects in comprehension, it does allow a unified account for production and comprehension.

Lastly, the framework of Self-Organizing Sentence Processing (SOSP) offers another perspective on agreement attraction. In this model, linguistic constituents attempt all possible attachments and eventually the attachments that have better feature match should outcompete the ones that have poorer feature match (Tabor & Hutchins, 2004; Tabor, Galantucci, & Richardson, 2004). As a link grows stronger, features on both sides migrate toward the same values. On this account there is no difference between the mechanisms involved in encoding and retrieval (Smith, Franck, & Tabor, 2018; Villata, Tabor, & Franck, 2018). Agreement attraction may arise from unification of two nouns' features as they attach to their slots (as "encoding" interference), or from the agreement features of the verb which attract similarly marked nouns (as "retrieval" interference). Thus, within different frameworks, agreement attraction effects are largely attributed to fallibility of memory encoding or retrieval

1.3.3 Semantic attraction

Recent studies revealed that illusory effects in subject-verb dependencies are not limited to agreement relations. Integrating a verb with its subject obviously involves a lot more than choosing the correct conjugation form. The sentence's interpretation crucially depends on the semantic integration of the subject and the verb, for construal of an event representation. This semantic link is apparently also susceptible to interference.

Cunnings and Sturt (2018) found that an implausible verb (such as *shattered* in 13) creates a smaller processing disruption when a non-subject noun semantically fits it (e.g. *cup* in 13a). Cunnings and Sturt interpret this as evidence for semantic facilitatory interference caused by the partial match of this distractor. Namely, they suggest that the "shatterablity" of the intervening noun in (13a) obscures the semantic mismatch between the verb and its actual subject. This finding was extended by Laurinavichyutea and von der Malsburg (2020), who suggest, based on a judgment paradigm, that semantic attraction and agreement attraction errors occur at comparable rates and are additive.

(13) 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.

1.3.4 Negative polarity

Lastly, another type of linguistic illusion comes from negative polarity items. Negative polarity items (NPIs) are expressions like *ever* or *any* (or the more colloquial *lift a finger*, and *budge an inch*). Broadly speaking, these lexical items require a context of negation to be grammatically licensed (see 14). Moreover, the NPI has to scope under negation, rather than just appear next to it. The exact constraints on the licensing environments for NPIs are more intricate. However, for our purposes, the important issue is that the structural position and the semantic properties of two non-adjacent words have to be verified in these cases. This process, once again, has to rely on the memory mechanisms which support sentence processing.

(14) a. No bills that the senators voted for will ever become law.
b. *The bills that the senators voted for will ever become law.
c. *The bills that no senators voted for will ever become law.

While speakers declare sentences like (14c) as ungrammatical, experimental work revealed these sentences often give rise to an illusion of acceptability. Sentences which include negation but not in the correct structural position are experienced as acceptable, according to eyetracking measures (Vasishth, Brüssow, Lewis, & Drenhaus, 2008), self-paced reading data (Xiang, Grove, & Glannakidou 2013; Parker & Phillips 2016), speeded acceptability judgments (Parker & Phillips 2016; Menida, Poole, & Dillon, 2018), and ERPs (Xiang, Dillon, & Phillips, 2009). Similar effects do not arise in the absence a distractor (as in 14b).

There are several different accounts for this phenomenon. One account suggested for this takes the same path of faciliatory interference which was also suggested for agreement and semantic attraction. Vasishth and colleagues (2008) proposed that the parser accidentally retrieves the distractor when searching for a of licensor for the NPI. Parker and Phillips (2016) argued that the NPI grammaticality illusion reflects access to one internal stage of the encoding process, and that the illusion can be reliably switched on and off when we probe a different point of the encoding process. Namely, they suggested that the representation of semantic-pragmatic relations is not fixed, and that the distractor can become opaque as a candidate for illusory licensing with the passage of time. Yet other accounts are not based on memory retrieval or encoding, but on pragmatic inferences (Xiang, Dillon, & Phillips, 2009; Menida, Poole, & Dillon, 2018).

1.4 Intended misanalysis: Rational noisy-channel processing

So far, I have discussed errors of the comprehender. However, there is another type of errors that each reader and each listener have to deal with – errors inherent in their input. That is, errors whose source is the producer of the text or the speech. Since typing and speech errors are highly frequent, and may obscure the intention of the speaker, a rational comprehender should take these errors into account when interpreting language. Indeed, most errors go completely unnoticed by the speaker and listener. The listener or reader extract the intended message, dismissing the corrupted form or the literal content. These complex computations are barely noticed.

Errors are registered if you are a linguist plotting to collect them, or if these slips of the tongue create humorous effects. A classic example for the humorous effect of errors is Mrs. Malaprop, the pompous aunt, in the 1775 comic play The Rivals. Various amusing moments arise as Mrs. Malaprop repeatedly draws similar sounding but absurd words. For example, she compliments someone by stating he is “the very pineapple of politeness” (i.e. pinnacle). The modern version of Mrs. Malaprop might be our current autocorrect applications. Various nonsense phrases which these technological advances produce have made their way into viral lists of "hilarious autocorrect fails". The gap between our skillful comprehension and the shorthand of these applications continues to surprise and humor people, but it also means that human typing errors have not yet been mastered by algorithms of natural language processing. The software seems to transform small and understandable errors (as perceived from the human perspective) into much bigger outlandish ones.

How come human errors are so frequent yet (mostly) leave our communicative ability untouched? Hofstadter and Moser (1989, p. 186) offer the following perspective: "The reason for this is that most errors are not simply random intrusions of 'noise' into an otherwise clear and unambiguous flow of communication; they are almost always intimately connected with the speaker's intended message, and reveal something of it. Rather than blatantly standing out from the rest of the utterance, a typical error blends in smoothly with it". To put this is a more formal way, we are able to automatically correct errors because they arise from an environment similar to our own processing mechanism. Speakers and listeners share a sense of linguistic similarity in language usage. Therefore, if the erroneous phrase was close enough for the speaker to confuse it with the target utterance, it is probably close enough for the listener to transform it back to the original message. Thus, we have a metric we can use to find a non-literal but more appropriate meaning.

1.4.1 Disfluencies and their effect on language comprehension

According to estimates, in spontaneous discourse speakers encounter some problem or another in their production in every tenth word on average (Fox Tree, 1995; Bortfeld, Leon, Bloom, Schober, & Brennan, 2001). Such disfluencies include repairs (e.g., “The man, I mean the woman”), but also other markers of production difficulty like filled pauses (e.g., “uh”, “um”) and repetitions (e.g., “The—the man”). Namely, normal speech is regularly interrupted by faults of the speaker's production system. These frequent disfluencies may hinder processing on the comprehender's side. However, the overt marking of the production difficulty can also

help the listener in reconstructing the speaker's communicative intention and adapting a fitting interpretation strategy.

Indeed, the information which disfluencies carry about the producer's condition is not overlooked by listeners. Listeners use disfluencies as a cue for an interpretation shift and modify their predictions. Recent investigations of disfluency processing (Lowder & Ferreira, 2016; 2019) revealed for example that repair words like *I mean* initiate an immediate attention shift from the referent preceding the repair to other possible referents (e.g. when hearing ...*the cat, uh I mean...*, people turn to a plausible competitor like "dog"). Thus, listeners quickly exploit a repair cue to generate expectations regarding the speaker's intention and mentally correct for the apparent errors, even before the speaker spells out the correction.

Comprehenders may also use non repair disfluencies to predict the upcoming input, given that speakers tend to use filled pauses when uncertain about an upcoming word (Smith & Clark, 1993; Brennan & Williams, 1995). Thus, listeners might be more likely to predict a deviant utterance when the speaker becomes disfluent. Indeed, upon hearing a filled pause (like "uh" or "um") listeners expect the speaker to mention a new, unfamiliar, or difficult to label referent (e.g. Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Kam, & Tanenhaus, 2007). In addition, listeners seem to integrate an unpredictable word more easily when that word is preceded by a filled pause, relative to an unpredictable word in uninterrupted speech (Corley, MacGregor, & Donaldson, 2007). This provides additional evidence that comprehenders use their experience as producers to interpret the corrupted input they get.

1.4.2 Good Enough processing

How does comprehension work when errors are not marked by a disfluency? In written text we rarely observe explicit evidence for production difficulty before a deviant word or an immediate correcting utterance. Yet many times, even edited texts incorporate some typing errors, word substitutions, or other repetitions and omissions. When the linguistic utterance is unfaithful to the speaker or writer's communicative intention, we require a theory which generates imprecise or contextually biased representations.

One prominent hypothesis suggesting that comprehenders often develop inaccurate representations is called "Good Enough" processing (Ferreira, Ferraro, & Bailey, 2002; Ferreira, & Patson, 2007; Karimi & Ferreira, 2016). Under this approach, listeners and readers tend to generate shallow or superficial representations in order to spare the cognitive load associated with full analysis of the utterance. This means that we often generate an interpretation which is not consistent with the actual input, but is instead are biased by world knowledge of applicable events and general biases.

Various findings support the conjecture that our interpretations sometimes reflect more of our expectations and real-world probability, than of the actual linguistic input. For example, Ferreira (2003) found that readers often reverse the roles of the referents in passive sentences when their non-reversed interpretation is implausible. Thus, people sometimes take a sentence like "*the dog was bitten by the man*" to mean that the dog, rather than the man, is the agent of the biting action. These erroneous interpretations occur more than in plausible (e.g. "*the man was bitten by the dog*") or active sentences (e.g. "*the man bit the dog*").

Other findings which support hasty and inaccurate interpretations come from well-known riddles like those in (15) (which are commonly called the Moses Illusion, after the sentence in 15a). People often respond to the question in (15a) with "two", failing to notice that it was Noah rather than Moses who took animals on the ark on that biblical story (Erickson & Matteson, 1981). Similarly, people often fail to detect the anomaly in questions like (15b) (if a man has a widow, then that man is dead, and thus cannot marry anybody) and in various other statements (Barton & Sanford, 1993; Sanford & Sturt, 2002).

(15) a. How many of each type of animal did Moses take on the ark?

b. Can a man marry his widow's sister?

In the Good Enough processing approach, forming and transforming sentential representations takes time and effort. Since these resources might not be fully available during real life communication, Ferreira and her colleagues suggested over the years that the parser uses bounded rationality and heuristics. At its core, the idea behind this theory is that the linguistic representations we form during sentence processing are only good enough to tackle the task at hand (i.e. we use only the interpretation steps relevant to the current task). Instead of navigating full linguistic representations and creating a complete analysis, we utilize heuristics that allow quick-and-dirty interpretations. As a side effect, the inaccurate and biased interpretations may also allow accommodation of various speech errors (Ferreira, & Patson, 2007).

1.4.4 Rational noisy-channel processing

An alternative approach attributes "shallow" processing of the types mentioned above directly to rational inference regarding possible perception/production errors. This approach, called Noisy-Channel processing, suggests that readers maintain some uncertainty with regard to the input. Thus, readers can be pulled towards "near-neighbour", extrapolated interpretations, when these have higher probability than that of the perceived utterance.

In models of Rational (Bayesian) Noisy-Channel processing, the addressee weighs the probability of the different intended utterances against the probability that noise corrupted these utterances into the form of the current input (Levy, 2008b). More specifically, the addressee ranks the possible underlying representations (marked here as R) of the input (I), deducing the probability that the underlying structure of the input is R_i, by balancing two probability components: the prior probability of that underlying representation, $\Pr(R_i)$; and the likelihood of perceiving the current input, given that R_i is the underlying representation (and given a certain noise-model), $\Pr(I|R_i)$.

(16) Bayes Theorem: $\Pr(R_i|I) \propto \Pr(R_i) \cdot \Pr(I|R_i)$

As a result, the highest ranked structure may not be completely faithful to the input. The parser may overlook some properties of the input, or "edit" it in different ways, if this would produce a sentence with higher probability. The existence of such a process can account for comprehenders' ability to recover the intended meaning from noisy utterances. However, it also means that parser may treat sentences that have only one grammatical reading as ambiguous.

One this account, the Moses illusion may exhibit our automatic correction for speakers' errors in lexical retrieval. Speech errors may arise in faulty phonology but they might also

reflect erroneous lexical retrieval presenting a word semantically related to the target (Garrett, 1975; Harley, 1984). Since speakers sometimes utter a word similar in meaning, yet quite different from the one they meant to use, rational noisy-channel inference may transform one word into another, more context-appropriate one. For example, Moses in (15a) may be very likely as a speech error version of Noah. Similarly, in (15b) *widow* can be easily taken as an unfortunate mis-integration of the concepts "deceased" and "wife" (and indeed many interpret this as a question about the legitimacy of a widower marrying his sister in-law). Namely, in light of speech errors, such comprehension errors may not reflect shallow processing but an efficient rational autocorrect process.

Another naïve intuition which the Noisy Channel model might be able to target is the difficulty of proofreading. Natural reading misses out on various typing errors like repetitions or omission of function words (e.g. *while you are engaged in the the reading of the curent sentence you might miss some of errors*). This makes proofreading a conscious (sometimes effortful) task rather than an automatic by-product of conventional reading. One way to account for this would be appealing to routine application of noisy-channel inference. If in normal reading we unconsciously correct the errors we encounter, turning off this mechanism might require mindful effort. In line with this, Staub, Dodge, and Cohen (2018) found that people fail to notice function word repetitions, even when they do not skip either of the instances of the word. A repeated *the* is detected in 46% of all trials, and still only in 66% of the trials when readers fixate on both instances. The finding that readers frequently overlook function word errors even when their visual input provides all the relevant data might be attributed to automatic noisy inference based on the possibility of eye movement control errors (as argued by Staub, Dodge, & Cohen, 2018) or based on likely typing and editing errors.

Turning to empirical evidence for the usage of priors and likelihood computations, several studies have provided intriguing findings in the past decade. The first experimental evidence for noisy-channel inference came from Levy, Bicknell, Slattery, and Rayner (2009). Levy and his colleagues found, using eyetracking, that readers look back to reevaluate the spelling of an early word, when it has an orthographic neighbor which grammatically coheres with a (preferred) simpler structure downstream. They examined the reading of sentences like (17a). This sentence contains a passive reduced relative clause (bracketed), a marked and dispreferred structure. However, switching the preposition *at* with its near-neighbor *as* would license the analysis of *tossed* as an active main verb, a much more frequent structure. The rate of regressions to the preposition was significantly higher in (17a) than in (17b), where the verb *thrown* does not allow for an active main verb reading. This suggests that when there is an alternative structure with higher prior probability readers consider errors in their representation. Moreover, regression rate in (17a) was higher than in (17c-d), where the preposition (*toward*) had no relevant near-neighbor. This suggests that when the near-neighbor alternative is unavailable, the low frequency string is accepted as is.

(17) a. The coach smiled at the player [tossed the frisbee].
b. The coach smiled at the player [thrown the frisbee].
c. The coach smiled toward the player [tossed the frisbee].
d. The coach smiled toward the player [thrown the frisbee].

In further support of the Rational Noisy-Channel framework, Levy (2011) demonstrated garden path effects which result from compromising faithfulness to the input. The study examined readers' tendency to disregard a comma when a subsequent fronted prepositional phrase could be used as a complement of a preceding verb (in sentences like 18a), in order to avoid the low-frequency locative-inversion structure. In such sentences, the noisy-channel interpretation would suggest neglecting the comma, but should lead the parser to a dead-end which requires structural reanalysis at the verb *lurched*. In line with this, the results revealed an increase in reading times at the disambiguating verb in (18a) relative to (18b), indicating reanalysis costs which are required if an input-unfaithful analysis was constructed earlier.

(18) a. As the soldiers marched, toward the tank lurched an injured combatant.
b. As the soldiers marched into the bunker, toward the tank lurched an injured combatant.

Another type of support for the Noisy-Channel processing framework comes from the correlation between the likelihood of modification, or edit, to the input, and the rate of unfaithful parses. Manipulating the likelihood component of the rational inference model, $\Pr(I|R_i)$, can be achieved in two ways: (i) by increasing the overall noisiness of the signal (with the prediction that when a signal is conveyed over a noisier channel, interpretation will be systematically biased towards higher-frequency interpretations); (ii) by comparing different types of modifications (with the prediction that less non-literal analyses will be adopted when the required modification is more notable). Using comprehension questions, Gibson, Bergen, and Piantadosi (2013) probed the rate of unfaithful interpretations in different settings and structures. They found that an unfaithful, semantically more plausible, reading was chosen at a higher rate when this could be achieved with fewer edits of the input. They also showed that readers engaged in this "error correction" process at a higher rate when the experiment included a high proportion of sentences with syntactic errors. The rate of error correction thus increases with the perceived noise rate of the situation.

Overall, studies conducted under the framework of Noisy-Channel processing brought to light comprehension strategies which accommodate the various sources of noise in natural communication (speech, print and perception errors). Studies under this framework suggest that comprehenders are ambivalent with regard to the reliability of the input, and may opt for an interpretation which is not completely faithful to the input. Yet they exhibit such interpretations only when editing is minor and when this will make for a sentence with a more frequent meaning or structure.

1.5 Dissertation outline

This dissertation investigates how readers deal with linguistic errors in interpretation, memory and perception/production of sentences. I explore this topic in three separate studies which correspond to the three issues presented above: the difficulty of amending an initial erroneous interpretation and the cost-effectiveness of predictive processing; the vulnerability to interference (in memory retrieval or encoding) from intervening constituents; and mechanisms for compensation for possibly misleading or noisy input.

1.5.1 Study I: Predictions motivated by syntax and discourse, and their disconfirmation

After introducing the general methods used throughout this dissertation (Chapter 2), in Chapter 3 I turn to investigate how we balance the benefits of predictive processing and the possibility of interpretation and prediction errors. As explained above, when we take (calculated) guesses in every processing step, we are bound to fail sometimes. Many times, we place our bets on an interpretation which eventually turns out to be the correct one. But other times we will have to retrace our steps and modify our initial interpretation.

How do we overcome the difficulty of initial misinterpretation? How well do we manage to recover from incorrect expectations? I investigate the costs of failed predictions motivated by different considerations - syntactic licensing, rapid semantic integration, or alignment with pragmatic preferences. I tease apart the role of syntactic motivations, pragmatic considerations and contextual prediction in active dependency formation, based on comparison of Hebrew constructions as in (19). In both these cases there is a prediction for co-reference to the underlined antecedent. In (19a) this prediction is a demonstration of the classic Active Filler strategy, anticipating a syntactic dependency. However, in (19b) the motivation for the prediction is strictly discursive, based on the pragmatic function of the *regarding*-phrase.

(19) a. The doctor asked <u>which nurse</u> the patients persuaded…
b. The doctor asked regarding <u>the nurse</u> if the patients persuaded…

Using both production and comprehension measures, I observe prediction for the resolution of a long-distance dependency, even in the absence of syntactic licensing pressures (as in 19b). Disconfirmation of these predictions should be similar in terms of structural monotonicity, accessibility of the alternative resolution and inhibition of the wrong interpretation, as the sentences have similar structures and the predicted constituents (a gap vs. a pronoun) are minimally different. However, I find that when it is motivated by syntactic pressures, predictive dependency formation has earlier and higher costs at disconfirmation sites. Moreover, disconfirmation of syntactic predictions exhibits higher semantic persistence of the initial interpretation. I argue that the difference in reanalysis costs is attributed to degrees of commitment to the prediction. Specifically, I propose that active dependency formation is faster or more binding in syntactically-motivated predictions (i.e. filler-gap dependencies) relative to pragmatically-motivated ones. This derives higher prediction costs and failure to completely inhibit the initial interpretation.

1.5.2 Study II: Memory fallibility in subject-verb-reflexive dependencies

Chapter 4 is concerned with the errors in the encoding or retrieval of syntactic features. Memory fallibility to linguistic illusion was repeatedly observed in agreement relations. This study asks how agreement computations in different parts of the sentence interact. Errors in agreement computation may arise due to trusting a faulty copy in memory, failing to maintain an accurate representation over time, or mis-retrieving it. Thus, it might be beneficial for the parser to update the representation of agreement information in intermediate retrieval points, or to take some "shortcuts" which could minimize the role of accessing the target noun itself.

To investigate this, I tested the processing of reflexive pronouns. Such pronouns, realized in English using the *-self* form (*myself, himself, herself*), mostly refer back to the subject of the clause, while other pronouns refer to various antecedents within or outside their clause (20). They therefore require identifying a feature-matched subject, but they may also depend on association with another agreement-bearing constituent, namely the verb. Potentially, consulting the agreement features on the (linearly proximal) verb can provide a "shortcut" or updating point, as the verb agrees with the subject.

(20) a. Snow White$_1$ said that the jealous wife$_2$ of her father$_3$ loves herself$_{*1/2}$
b. Snow White$_1$ said that the jealous wife$_2$ of her father$_3$ hates her$_{1/*2}$
c. Snow White$_1$ said that the jealous wife$_2$ of her father$_3$ loves *himself$_3$ / him$_3$

These options were not investigated before since reflexive agreement attraction is usually manipulated though gender cues, while in most languages verbal agreement does not manifest gender features (at all, or most of the time). In Hebrew, however, verbal agreement mostly reveals the gender of its subject. Thus, Hebrew presents a unique environment in which we can test the interaction between verbal and reflexive agreement.

In this study I thus investigate how agreement marking on the verb affects the processing of a subsequent reflexive pronoun. I show that the processing of agreement features at the reflexive pronoun highly depends on the availability of agreement cues on the verb. I find that (i) when the verb mismatches the subject, a reflexive which matches the verb is preferred over one which matches the subject. (ii). Agreement attraction is less prominent when the verb manifests agreement cues. I interpret this as evidence that verbal agreement either alters the representation of the subject; or makes retrieval of the subject's features unnecessary (i.e. the reflexive does not reaccess the subject when verb agreement is available, but verifies verb-reflexive agreement). These strategies might be beneficial for processing as they allow the parser to reduce the reliance on the individual representation of the early noun phrase.

1.5.3 Study III: The Balance between possible errors and rare linguistic structures

Chapter 5 presents support for the application of noisy-channel inference in incremental processing. Specifically, I suggest that the prior probability of the syntactic structures which are faithful to the input modulates readers' choice of noisy interpretations. I exhibit that the noisy parse (which assumes an error occurred in the text) is given a high weight when the syntactic structure underlying the faithful analysis is extremely rare. Yet, when the parser has a syntactic alternative which is not as rare, it opts for this alternative, even if the structure underlying the noise-inflated parse is even more frequent.

I investigate this using Hebrew relative clauses in which the subject position is vacant, and the verb mismatches the filler in number (21). At the verb, the parser may construct a subject relative, ignoring the agreement mismatch, or choose a less frequent structure of an object relative (either predicting a post-verbal subject or assuming a null impersonal one).

(21) a. We met the student$_1$ who __$_1$ know the answer. Agreement error
b. We met the student$_1$ who know __$_1$ the teachers well. Post-verbal subject
'we met the student who the teachers know well'
c. We met that student$_1$ who *pro* know his$_1$ work. Impersonal null subject
'we met the student who people know his work.'

In four self-paced reading experiments and one production experiment, I show that the effect of agreement mismatch is modulated by the prior probability of the alternative analysis. I find evidence that a corrupted subject relative parse (as in 21a) is preferred over assuming the rare word order in (21b). This suggests that readers are willing to ignore the agreement information. Yet, I also find evidence that readers attempt to remain faithful to the input, when a grammatical analysis is not extremely rare. Thus, readers prefer constructing the structure of (21c), which is grammatically pristine and relatively common, over the more frequent yet corrupted subject relative reading in (21a). These findings suggest that readers rapidly apply elaborate probabilistic knowledge regarding the distribution of syntactic structures in their language. They are willing to construct ungrammatical subject-verb dependencies. Yet they do so not due to shallow processing, but as a rational strategy for avoiding highly improbable utterances.

1.5.4 Overarching significance

Over all, results from the three studies present a complex system of "checks and balances", which allow for efficient sentence processing given the prospect of various errors. Chapter 3 concludes that the parser modulates its commitment to predictions, and thus the costs of their reanalysis, based on the differences in the motivation for the prediction. Chapter 4 suggests that computation of agreement dependencies is aided by the flexibility of agreement representation/verification. This might allow the parser to overcome memory fallibility by enhancing or bypassing the retrieval of agreement of the non-local target. Finally, Chapter 5 presents evidence for application of rational noisy-channel processing during incremental processing.

Thus, the reported studies contribute to our understanding of the human comprehension mechanism and suggest several ways in which the system might compensate for different pitfalls of processing. that language comprehension encompasses various strategies for dealing with the uncertainty of its own representations. These strategies include modulating the commitment for some prediction types, utilizing local but indirect information as a cue, and rational probabilistic inference based on likely errors. Put in place, these strategies can allow readers and listeners to skip over obstacles like typing and speech errors, noisy perception, faulty memory remnants, and failures of eager parsing decisions.

Chapter 2

Methods overview

This thesis reports 13 experiments overall: Five experiments in Chapter 3, three in Chapter 4, and an additional five in Chapter 5. Of these experiments, eight are self-faced reading experiments, three are sentence completion experiments and the remaining two are acceptability judgement experiments. Since much of the experimental protocol is identical across the different studies, I report here the main points of the experimental procedures and data analysis methods which I use throughout the dissertation.

2.1 The experimental procedures

Participants were recruited through posting on university notice boards, mailing lists, and social media. Participants were mostly non-linguistics students and were naïve to the purpose of the experiment. They received monetary compensation or course credit. Monetary compensation was set to a rate of approximately 1NIS (or 30 cents) per minute. For example, 20-minute experiments were offered for an amount of 20NIS - approximately 6 US dollars. Participants gave informed consent before starting the experiment, and all experimental procedures were approved by Tel Aviv University's ethics committee.

2.2.1 The setup of self-paced reading experiments

For lab-based experiments, the experiment was conducted on a desktop PC using the Linger software package (Rhode, 2003). Web-based experiments were built using IbexFarm (Drummond, 2018), and participants took the experiment online, on their own computers. The self-paced reading experiments reported in this dissertation took approximately 20 minutes.

Sentences were presented in a moving window, non-cumulative self-paced reading display. Sentences were initially masked by dashes (with no spaces or punctuation, on lab-based experiments; with spaces and punctuation intact on web-based experiments). Participants pressed the space bar to reveal each word. Participants were instructed to read each sentence continuously, at their own pace, such that they understand it, but avoid lingering on words.

Before they began the experiment, participants underwent a practice block of four (on lab-based experiments) or seven items (on web-based experiments). Between 50%-100% of the trials, in both the practice and experimental blocks, were followed by a ‘yes/no’ comprehension

question (exact percentages are reported with each experiment). No feedback was provided for correct or incorrect answers. The order of presentation was randomized for each participant.

2.1.2 The setup of acceptability judgment experiments

Participants for these experiments were mostly volunteers (recruited through social media), who completed the experiment without monetary compensation or course credit. The questionnaires were built using Google forms. Participants filled the questionnaire online. They were instructed to rate the acceptability of the sentences on a 7-point scale, with 1 indicating that the sentence is not natural at all and 7 indicating a very natural sentence. Order of the items was randomly generated for each participant.

2.1.3 The setup of sentence completion experiments

The three sentence completion experiments reported in this dissertation are very different in nature. They were all run online, however their procedure ranges from a completely free completion task, with all sentences presented together and no time limit, to rapid serial visual presentation of the preamble and forced choice of the completion from two fixed alternatives. More about the procedure of each of these experiments appears in their individual methods subsection. The coding schemes for the free response experiments are also described the following chapters, along with the other methodological details of the experiment. I will note that participants of the production experiment of Chapter 3 where mostly volunteers while participants of the timed presentation experiments (on Chapters 4 and 5) received monetary compensation (of 20 and 15 NIS correspondingly).

2.2 Data analysis protocol

The data are analyzed in R (R Development Core Team, 2015) using mixed-effects linear models of the lme4 package (Bates et al., 2014). I conducted two types of analyses: a frequentist analysis (null hypothesis significance testing) and a Bayesian analysis using weakly informative priors. Contrast coding is derived from the experimental design and the predictions of each study. I therefore report the coding scheme separately for with each experiment.

The frequentist models were fitted using restricted maximum likelihood (REML) and *p*-values were extracted by Satterthwaite degrees of freedom approximation using the lmerTest package (Kuznetsova et al., 2017). I corrected *p*-values using the Bonferroni correction, by the number of pairwise comparisons which were conducted (number of comparisons used for the correction is reported along with every experiment).

The Bayesian hierarchical models were fitted using Stan (Carpenter et al., 2017), via the brms package (Bürkner, 2017). Results of the Bayesian models are derived from the posterior distributions of parameters of interest. For each model, I sampled from the posterior distribution by running four MCMC chains at 4000 iterations each. The first half of the samples was discarded as warm-up samples. Convergence was checked using the R-hat convergence diagnostic and by visual inspection of the chains (Gelman et al., 2014). I summarize the posterior distribution by presenting its mean along with the 95% credible interval (CrI). I consider an effect reliable if over 95% of the sampled posterior distribution is beyond zero.

Following Barr et al. (2013), a maximal random effect structure was constructed. Namely, in addition to random participant and item intercepts, I initially included correlated random slopes, by-subject and by-item, for each of the fixed effects as well as their interactions. However, in some cases, due to convergence failure, slopes accounting for the lowest variance were gradually removed from the frequentist analysis. The Bayesian models included a full random effects structure.

2.2.1 Analysis of self-paced reading data

Participants were excluded from the analysis if their performance on comprehension questions of experimental items was lower than 70%, or if their average RT in experimental items was more than 2.5 SDs above the group's average. For the remaining participants, RTs higher than 2.5 SDs above the individual's average RT were trimmed to that cutoff, and RTs shorter than 120ms were excluded. In each experiment I analyze RTs from the critical region, as well as from the spillover word (namely the post-critical word). In models which included two regions in one analysis (Experiments 8A-B and analysis of Experiment 3 on Appendix A), I included another by trial intercept to accommodate the fact that the two measurements (at the critical and post-critical word) were not independent.

Null hypothesis significance testing

To remove irrelevant effects, I used a two-step analysis. First, a basic model using log-transformed RTs was fitted to all the data, namely all the words in all the sentences, including both filler and experimental items (but excluding practice items). This model included word

length, position of the word in the sentence and trial number as fixed effect predictors, as well as a random intercept for participants.[2] Residuals of this model were then entered into a second model. In this second stage, only data from critical words (or from the spillover region) in the experimental items were considered. This model included as fixed effects the experimental factors and the log RT of the previous word.[3] I fitted the second stage model separately for the critical word and for the spillover word. Unless specified otherwise, a maximal random-effect structure was used in this model, with intercepts and slopes for all experimental factors and their interactions.

Bayesian analysis

In these models, I did not use residualization or non-experimental fixed-effects, to allow a straightforward transformation of the posterior means and credible intervals. Posterior means (and credible intervals) were back-transformed to the millisecond scale to facilitate interpretation of the results. I used regularizing, mildly informative priors (Gelman et al., 2014): a standard normal distribution, N (0, 1), was used as the prior distribution for the fixed effects and for the standard deviation parameters. For the intercept, I had a wider prior of N(0,10). Finally, for the correlation matrices of the random effects, priors were defined using the LKJ prior (Lewandowski, Kurowicka, & Joe, 2009).

2.2.2 Analysis of acceptability judgments

Within-subject z-transform was argued to eliminate variability that results from differences in scale biases across participants (Schütze & Sprouse, 2014). I therefore conducted my analyses z-transformed data, using linear mixed effects models. On the Bayesian analysis, I used regularizing, mildly informative priors: a standard normal distribution N (0, 1), was used as the prior distribution for the fixed effects and for the standard deviation parameters. For the intercept, I had a wider prior of N(0,3). Finally, for the correlation matrices of the random effects, priors were defined using the LKJ prior.

2.2.3 Analysis of sentence completion data

Coded completions present a binary data set (e.g. producing a reference to the critical entity or not, choosing a feminine or masculine form, completing the sentence as a subject or object relative). Therefore, analyses used a logistic mixed-effects model – the generalized linear mixed effect model of the lme4 package on the frequentist analysis, and a Bernoulli response distribution on the Bayesian analysis. I used the same regularizing, mildly informative priors as in the acceptability data analysis: a standard normal distribution N(0, 1) as the prior distribution for the fixed effects and for the standard deviation parameters, a wider prior of N(0,3) for the intercept, and the LKJ prior for the correlation matrices of the random effects.

---

[2] logRT ~ experimental + rcs(position) + log(trial) + length + (1|subject)
The restricted cubic spline (RCS) function is used here for word position in order to capture non-linear effects of this factor (e.g. wrap-up effects).

[3] ResidualRT ~ condition + previousWordRT + (1+ condition | subject) + (1+ condition | item)

Chapter 3

Predictions motivated by syntax and discourse, and their disconfirmation

3.1 Introduction

During sentence processing, comprehenders form different predictions regarding the unfolding of the sentence. Yet, every act of predictive processing puts the parser in greater risk of having to undertake reanalysis. Therefore, it might not always be cost-effective for the parser to engage in such strong predictive processes. For example, predictive structure building might come into play only when maintaining the structure in its incomplete form is too costly. The current study asks which types of pressures result in predictive dependency formation, and what are the costs of prediction failure in each case.

In this study, I investigate the contribution of syntactic pressures, pragmatic constraints, and contextual availability as motivations for prediction of long-distance dependencies. I test for differences in the parser's commitment to predictions based its motivation: Do costs of disconfirmation (i.e. reanalysis costs) differentiate syntactically-motivated and pragmatically motivated dependencies? Many previous studies exhibited prediction based on discourse preferences. For example, it was exhibited reference resolution, the interpretation of a definite referring expression within a visual array, is predicted rapidly based on various discourse variables like the interaction between on givenness and prosodic stress (Dahan, Tanenhaus, & Chambers, 2002), the partner's visual perspective (e.g. Keysar, Barr, Balin, & Brauner, 2000; Hanna, Tanenhaus, & Trueswell, 2003; Ryskin, Wang, & Brown-Schmidt, 2016), and the usage of scalar and non-scalar adjectives for contrast (Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Heller, Grodner, & Tanenhaus, 2008). Yet, to the best of my knowledge. there has not been much research focusing on the costs of disconfirming such predictions. This study uniquely provides a close comparison between pragmatic and syntactic effect, thus allowing insight into possible similarities and difference between the two. I will argue that the parser's commitment to the prediction, and thus the costs of its disconfirmation, are modulated by the source of the prediction.

We know that predictions of very different kinds are computed during incremental sentence processing. Various studies have revealed that readers and listeners construct interpretation and anticipate the upcoming input incrementally, rather than passively waiting for words to appear and indicate the correct interpretation. This is evident both in cases of local

ambiguity, were the parser rapidly favors one of the alternative analyses, and in the pre-activation of words which are expected to appear later in the sentence. However, disconfirmations of different predictions are not equally difficult. Specifically, lexical prediction and structural ambiguity resolution appear to be on the edges of the scale, with no consistent behavioral costs for the disconfirmation of lexical predictions on the one end, and robust processing disruption in garden path sentences on the other hand. To investigate intermediate levels on that scale I use prediction for resolution of filler-gap and binding dependencies. I would like to suggest that the parser can exhibit differential commitment to different predictions, depending on their linguistic necessity. To allow us to consider reanalysis costs as evidence for the parser's commitment to predictions, we shall first review different approaches to reanalysis difficulty.

3.1.1 What makes reanalysis difficult?

Reanalysis in cases of early misinterpretation is necessary for accurate language comprehension and is employed frequently, and often automatically, in human communication. Yet altering an interpretation you committed to may be challenging in terms of linguistic structure building, semantic interpretation, and employment of memory and inhibition mechanisms. As many authors have noted (e.g. Gorrell, 1995; Pritchett, 1992), reanalysis may present the parser with challenges varying in magnitude, even within the realm of structural ambiguity. Different factors may affect the difficulty of reanalysis processes. Thus in comparing syntactic and pragmatic effects we should consider whether the tested structures differ in with regards to these factors.

Modelling the commitment of the parser

On parallel processing models, disconfirmation of predictions requires re-ranking of the relevant alternatives (Levy, 2008a; MacDonald et al., 1994). Namely, to align with the bottom-up input, the parser needs to change the balance on the scales, increasing the weight of some interpretations and decreasing the weight of others. This reallocation of processing resources is costly, presumably as the parser is more committed to (or has invested more prediction resources in) interpretations of higher probability weight. The lower the weight of the current input was initially, the more disconfirmation costs will increase. Therefore, on parallel processing models, "reanalysis" costs, and commitment level, mostly reflect the plausibility and frequency of the final structure.

Another type of parallel processing framework is dynamical Self-Organizing Sentence Processing (Tabor & Hutchins, 2004). On this model, the alternative interpretations are not denoted by distinct representations but by different attachment choices, with different strength coefficients, on the same syntactic tree. The continuous coefficient value, which is a function of the fit between the two attached constituents, reflects the stability of the parsing choice. Thus, it determines how costly it would be to deconstruct it in a case of reanalysis. On this model the commitment level is based on the (syntactic and semantic) fit of the attachment, the power of the constraint, and the passing of time (a postulated attachment will get stronger if the incoming material is consistent with it, and thus, as long as the input is ambiguous, the “rich

get richer", in line with findings that reanalysis becomes harder when the ambiguous region is longer, Ferreira & Henderson, 1991; Tabor & Hutchins, 2004).

Differential commitment levels can arise also in serial processing. In serial processing models, only one interpretation is pursued at each moment. Therefore, on a single parsing event, there could not be multiple alternative interpretations or weights. An analysis choice is binary – an analysis is either pursued or not. However, commitment levels can still be implemented through the rate of choosing a specific analysis (across many trials), or the speed of constructing the analysis.

The Unrestricted Race model (van Gompel, Pickering, & Traxler, 2000) suggests that the parser's commitment level can be reflected in the consistency of its choices across trials. On this model, only one analysis is constructed at a time, but the probability of choosing it is a function of its likelihood (given the frequency of its structure and the plausibility of its meaning). Therefore, a strong bias (when one analysis is highly favored) would lead to consistent interpretation, a weak bias would lead to adopting one analysis most but not all of the time, and on a balanced ambiguity (where analyses are equally probable) the analysis will be chosen at random. Commitment levels which are implemented as the average rate of constructing a certain interpretation or prediction would also produce different means of reanalysis costs. When averaging across trials, as one usually does in psycholinguistic experiments, conditions where reanalysis is less frequent would appear as conditions with less processing difficulty than those where reanalysis had occurred on more trials.

Another proposal which may account for gradience in reanalysis costs is Construal (Frazier & Clifton, 1996). This theory distinguishes primary relations (between syntactically obligatory constituents) and non-primary ones (involving adjunct phrases). While primary relations are resolved immediately, non-primary ones are constructed relatively slowly. Namely, some relations may be left underspecified, or get attached in a way/timescale which allows later processes to refine them without notable processing costs. Thus, low commitment could also be implemented as a slow prediction. If a prediction does not arrive at a full representation or stable attachment by the time reanalysis cues arrive (or ever), we should observe low costs of disambiguation. This is relevant to the distinction between pragmatically and syntactically motivated dependencies (debated in the current study), as Construal presumably draws from the idea that adjuncts, possibly like non-syntactic predictions, have little consequences for other aspects of the sentence structure.

Consistency and transparency at the reanalysis site

The differential costs of reanalysis could also be attributed to the different computations that the parser has to execute in order to bring the initially dispreferred interpretation into the foreground. Different factors along these lines were suggested as accounts for more severe and less severe Garden Path effects. Contrasts in difficulty of reanalysis were sometimes attributed to the extent of convergence between the initial interpretation and the new one (representation-preserving models). This assumes that the costs of processing unpredicted material are associated with the need to modify some core aspects of the sentence representation that the parser has formed. Different authors postulated this based on varying types of "core aspects".

Pritchett (1988; 1992) suggested that reanalysis is costlier when it involves a change in the thematic domain of an argument. Thus, according to Pritchett, reanalysis is costly when it involves a major rearrangement of thematic relations. Later, many authors anchored reanalysis costs in monotonicity of syntactic structure, rather than thematic roles. Specifically, structural dominance, a transitive hierarchical relation associated with the position of one node relative to another in the syntactic analysis, was the heart of representation-preserving accounts. Weinberg (1993), Gorrell (1995) and Sturt and Crocker (1997; also Sturt, Pickering, & Crocker, 1999) suggest (in various implementations) that easy reanalysis occurs only when the required revision does not falsify any dominance relations asserted in a previous parse.[4]

In addition, semantic persistence was also suggested to partly modulate reanalysis costs. This term expresses the compatibility of the event depicted in the full sentence with the one envisioned based on the initial ambiguous region. Sturt (2007) suggested that the compatibility of the two readings affects the processing difficulty at the reanalysis region of sentences like (22a-b). Both these sentences require moving the *South Pole* from the direct object position of *found* to a position within this verb's sentential complement (amending the syntactic and thematic relations similarly in both cases). However, in (22a) the final interpretation is compatible with the notion that *the explorers found the South Pole* (as constructed in the initial parsing), but the continuation of (22b) explicitly falsifies it.

(22) a. The explorers found the South Pole was actually right at their feet.
b. The explorers found the South Pole was actually impossible to reach.

Apart from representation-preserving factors, accessibility of the alternative structure (Fodor & Inoue, 1994) can be invoked to account for variability in the difficulty of reanalysis. On this proposal, similar structures may require different reanalysis effort depending on the nature of the disambiguation site. Reanalysis is easier when the disambiguating material does not only indicate that there is something wrong with the sentence structure, but also highlights the wrong turn the parser had taken (i.e. which attachment was erroneous). For example, in the initial interpretation of (23a) the phrase *in the library* modifies the embedded verb *reading*. This reading is disproved when the sentence ends without a compatible location argument for the main verb *put*. This null evidence is difficult to decipher relative to overt evidence, under Fodor and Inoue's (1994) proposal. Thus, in sentence (23b), reanalysis should be relatively easy. The initial interpretation posits that the coffee table is the complement of the verb *put*, yet the following phrase cannot be attached unless the first is taken to be the original location of *the book* (i.e. a modifier of this noun phrase) rather than the target of the *putting* action. This overt symptom of the erroneous analysis facilitates reanalysis according to Fodor and Inoue.

(23) a. Susan put the book that she'd been reading all afternoon in the library.
b. Susan put the book on the living-room coffee table into her briefcase.

[4] Another representation-preserving account (Bader, 1998) suggests that the need to revise the original prosodic structure (i.e. the intonation in which the sentence was read) makes reanalysis difficult. Bader (1998) points out that readers create a prosodic representation of the sentence, along with its syntactic structure, even in silent reading. A syntactic misanalysis may require the reader to replace the original prosodic structure, and Bader provides experimental evidence that recovering from such a reanalysis is more difficult.

c. The fact that Susan put the book that she'd been reading all afternoon in the library bothered her parents.

Moreover, Fodor and Inoue (1994) note that overtness of the evidence is not the only factor determining the accessibility of error detection. Overt evidence may not help in diagnosing the error effectively, in cases like (23c). In contrast to (23a), the end of the relevant clause in (23c) is indicated by an overt symptom – the verb *bothered* – rather than by a null ending of the sentence. However, this verb and the higher clause have no direct repercussions for the phrase whose erroneous attachment has to be mended.

3.1.2 Ambiguity resolution and lexical prediction: Differences and similarities

Listeners and readers are immediately sensitive to the incompatibility between their predictions and the actual input. Upon encountering unexpected input, they rapidly revise their interpretation, whether this new input concerns the resolution of structural ambiguity (e.g. Frazier & Rayner, 1982; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) or of lexical prediction (e.g. Altmann & Kamide, 1999; Chow & Di Chen, 2020). However, the cognitive and behavioral costs of such revisions are not equally apparent in both cases.

How come the behavioral costs of disconfirmed expectations are so clear in the case of structural ambiguities but not in lexical predictions? Some lexical predictions have such high probability that in some respects they may be a "safer bet" than ambiguity resolution strategies are. Namely, in cases of highly constraining sentences, the probabilistic certainty regarding the next word might be greater than the that of structural ambiguity choices. Yet only for ambiguity resolution we observe clear reanalysis costs in reading times measures. Therefore, it does not seem likely that simple predictability or Surprisal considerations account for this contrast.

In addition, this puzzle probably cannot be resolved by simply appealing to the contrast between reanalysis of previous material (the ambiguous part of the sentence) and inhibition of material which was not in the input to begin with (the predicted word). Reanalysis costs in filler-gap dependencies (the filled gap effect) appear even when the resolution site is not expected to be null, in grammaticized resumption languages (Keshev & Meltzer-Asscher, 2017). Thus, we should search for a more complex solution to this puzzle.

Theories of reanalysis costs offer some directions as for why disconfirmed lexical predictions are easier to bounce back from than disconfirmed structural predictions. First, the former are associated with minimal phrase-structure change, aligning with structural monotonicity requirements (in line with Sturt & Crocker, 1996, among others). Second, disconfirmation of a lexical prediction directly indicates the correct interpretation and does not require complex diagnosis of the parsing error (in line with Fodor & Inoue, 1994), apart from possibly adjusting the probabilistic information which led to the prediction.

Lastly, there might be a different level of commitment to predictions which are motivated by structure building or grammatical licensing, relative to those related to meaning and world knowledge (which usually are the basis of lexical prediction). The intuition that syntactic information is more binding than context accompanied linguistic theory over the years. The cancelability of pragmatic inferences is part of what distinguishes pragmatics from syntax and

semantics. This has also found a way into the reanalysis literature as a side note. Fodor and Inoue (1994) suggest the intuitive judgment that semantic incongruency is not as good of a reanalysis cue as a syntactic one. In sentence (24a) readers may resist transforming the sentence structure, maintaining that the waiter met the story, however implausible it is for *story* to be the complement of *met*. In contrast, in (24b), the alternative structure immediately comes to mind as the verbal phrase *to go home* syntactically cannot be a complement of the verb *met*. This can be taken as indication that parsing gives higher weight to syntactic constraints relative to semantic-pragmatic ones.

(24) a. They told the customer that the waiter met the story.
b. They told the customer that the waiter met to go home.

Why should the parser exhibit different levels of commitment at all? Consider the advantages and disadvantages of anticipatory processes: On the one hand, prediction helps the cognitive system to prepare for the prospective processing operations. Such preparation may free up resources for later parts of the sentence (when the predicted material is encountered). On the other hand, wrong predictions may require extra processing costs for repairing the incorrect representation. Thus, prediction comes at a price of an endless chase after incremental interpretations and their amendments.

In addition, predictive processing entails extra computations, which add up to the ones required for deciphering the current input. Generating the prediction and maintaining it may increase the burden on the parser. For example, in the Syntactic Prediction Locality Theory (Gibson, 1998), predictions have sustained costs: each predicted constituent (which is required in order to syntactically license the sentence) gives rise to processing load until the prediction is satisfied by an input element. Given these possible disadvantages, it seems likely that our comprehension mechanism would not commit to every anticipation, and rather leave full predictive processing to cases in which the benefits overweigh the costs. Thus I propose a comparison between syntax based and discourse based predictions which controls the pragmatic content across critical cases. This close comparison would allow evaluation of active prediction in both cases.

3.1.3 Filler-gap and other dependencies

The current study aims to tease apart the role of syntactic, pragmatic and contextual considerations in processes of dependency prediction. To test this, I investigate processing of the four Hebrew sentence types exemplified in (25).

(25) a. *ha-menahel badak eyzo axot*$_1$ *ha-metupal šixnea* _$_1$ *lanuax*
The-manager checked which nurse the-patient persuaded _ to.rest

b. *ha-menahel badak legabey ha-axot*$_1$ *im ha-metupal šixnea ota*$_1$ *lanuax*
The-manager checked regarding the-nurse if the-patient persuaded her$_1$ to.rest

c. *ha-menahel badak im ha-axot*$_1$ *im ha-metupal šixnea ota*$_1$ *lanuax*
The manager checked with the nurse if the-patient persuaded her$_1$ to.rest

d. *ha-menahel badak im ha-metupal šixnea et ha-axot lanux*[5]
The manager checked if the-patient persuaded ACC the nurse to.rest

The embedded wh-question in (25a) is a filler-gap dependency, which should exhibit predictive processing and reanalysis costs. Filler-gap dependencies are a well-studied test case of active dependency formation. Readers usually predict the resolution of such dependencies before obtaining unequivocal evidence (known as the Active Filler strategy; Frazier, 1989). Moreover, they commit to the predicted analysis, to the degree that subsequent conflicting input disrupts processing. Namely, the parser is eager to resolve these dependencies, postulates a gap prematurely, and then may confront reanalysis if another argument occupies the supposed gap position. This reanalysis is exhibited in behavioural measures as elevated reading times on the word occupying the potential gap position (known as the Filled-Gap Effect; Stowe, 1986).

Going back to our example sentences, on the other end of the scale, the sentence fragment in (25d) could not elicit a parallel prediction of a dependency at the point of the embedded verb, since it does not contain a parallel referent which the comprehender might expect to receive the theme role of *persuaded*. This case is comparable to the baseline condition from Stowe's (1986) study and will thus provide an unambiguous baseline in the current study as well.

Sentences (25b-c) present cases of pronominal binding, with an antecedent at a position similar to that of the filler's and a pronoun at the same position as the gap (25a). Thus, sentences (25b-c) differ from (25d) in that their main clause includes an additional referent which can be used to predict a direct object for the verb *persuaded*. Upon reading the verb, comprehenders may predict a pronoun co-referring with this antecedent, to arrive at a rapid interpretation of the sentence. Therefore, these sentences may diverge from a sentence like (25d). However, these sentences clearly also diverge from (25a), as they do not involve a syntactic requirement for a subsequent co-reference site, as filler-gap dependencies do. The sentences in (25b-c) thus lack the syntactic motivation for active dependency formation.

There is also a critical distinction between sentences (25b) and (25c). These dependencies differ with regard to the information structure of the antecedent *the nurse*. Sentence (25c) merely presents it as another discourse referent which might or might not be relevant later. However, (25b) presents it as a topic, a discourse-prominent antecedent within a *regarding*-phrase. Intuitively, *regarding*-phrases as in (25b) require that the next clause will be a comment on the NP embedded in them. Since the clause is expected to be a statement about the antecedent in (25b), the element co-referring to the antecedent should be confined to prominent, matrix, positions (I test these assumptions explicitly in Experiment 1). This is similar to what has been claimed about "as-for" topics as well as fillers in filler-gap dependencies (Kuno, 1976). Discussing their discourse function, Kuno (1976, p. 420) postulated that "a relative clause must be a statement about its head noun". When the sentence introduces such a topic, the resolution site (co-referring to the topic antecedent) must be a potential focus domain. Namely, similarly to the gap site in (25a), the pronoun in (25b) should pragmatically be in the

[5] *im* 'with' and *im* 'if' are spelled differently in Hebrew (עם vs. אם, respectively).

part of the utterance which is asserted, rather than presupposed.[6] If sentences like (25b) give rise to predictive processing, this would be for their discourse function (and the pragmatic considerations that this function entails). Therefore, the case of *regarding*-antecedents allows us to construct an environment similar to that of filler-gap dependencies, but to eliminate the syntactic licensing pressure.[7]

Lastly, when we examine again the contrast between our examples in (25), we can observe that the *with*-phrase in (25c) allows re-usage of the antecedent embedded in it, but does not require it in any way. Readers may prefer the sentence to refer back to this contextually available antecedent over introducing a new discourse referent. However, there is no syntactic or pragmatic requirement here. Therefore, if prediction is observed in this case, we could suggest that predictive processing is hyper-active and occurs with any possible context. On the next subsection, I explain how comparing these types of dependencies can help us understand the parsers commitment to different predictions, and shed light on the contrast between reanalysis of lexical prediction and ambiguity resolution.

3.1.4 The current study

After establishing that *regarding*-phrases present a pragmatic requirement for coreference (Experiment 1), I test the viability of the prediction in the different sentence structure in (25), using production rates of co-referential elements (Experiment 2). I then use the filled-gap effect to examine reanalysis costs of *regarding*-dependencies and wh-questions (Experiments 3 and 4). This paradigm "fills" the predicted co-reference site (a gap in the case of wh-questions and a pronoun in regarding-dependencies) with a lexical NP as disconfirmation of that prediction.

The comparison of filler-gap dependencies with *regarding*-dependencies and non-topic antecedents allows us to closely examine the difference between predictions motivated by syntactic and non-syntactic factors, and to test whether disconfirmation costs are observable for cases which do not directly involve a structural ambiguity. In the context of the different factors laid out above, these dependencies allow us to control for differences in the monotonicity and transparency of the full correct structure. First, if the parser predictively forms *regarding*-dependencies (or pronominal dependencies following other antecedents as in the *with*-phrase case), their disconfirmation should involve structural modification similar to

---

[6] Kuno (1976) suggests that the prominent discursive status assigned to both aboutness topics (like *as-for* or *speaking-of*) and fillers results in a similar pattern of constraints. Thus, non-extraction functions seem to also be somewhat affected by "island" constraints, like relative clause islands (i), and other complex NPs (ii).

i. Speaking of violence, Snead is an Englishman who condones it. (Kuno, 1976)
ii. As for TextMangler, I heard the rumor that Bill Gates uses it. (Falk, 2009)

[7] It should be mentioned that some analyses suggest that the aboutness of a topic constituent can present a syntactic requirement for co-reference, rather than only pragmatic expectation. Aboutness in proleptic object constructions of various languages may require unique syntactic licensing through complex predication (Salzmann, 2017). However, the Hebrew counterpart *legabey* does not withstand the strict conditions of a syntactic requirement (Landau, 2011, mentions that the Hebrew counterpart prolepsis uses the preposition *al*), since they allow topichood without an explicit pronoun

i. *šamanu al Dan še-mašehu nora kara *(lo)*
we.heard about Dan that-something terrible happened *(to.him)
ii. *šamanu legabey Dan še-mašehu nora kara (lo)*
we.heard regarding Dan that-something terrible happened (to him) (Landau, p.c.)

that of filler-gap dependencies. On either of these we predict a co-referential element in the same embedded position, and need to replace it with another NP. Therefore, in terms of dominance, thematic or semantic monotonicity, reanalysis of such dependencies should incur similar costs. Second, on each case the disconfirmation does not indicate the correct resolution site of the dependencies (i.e. the gap site in wh-questions or the position of the pronoun co-referring with the antecedent in the *regarding*-phrase or the *with*-phrase dependencies).

As we maintain such aspects constant across conditions, we leave only the different motivation as an account for the differences we may trace. The motivation for predicting co-reference for a non-prominent antecedent, in the *with*-phrase case, could be maximizing the interpretation at any given point (Altman, 1999) and avoiding new discourse referents. In *regarding*-phrase sentences, prediction could be motivated by such maximal interpretation strategy, or by information structure considerations which require that the clause provides a comment about the *regarding*-antecedent. On wh-questions, prediction could draw from such context-based factors, but most dominantly also includes pressure to syntactically license the dependency through a gap site (and specifically a gap site in a non-island position). To preview my results, across production and reading data I find distinction between the three levels of prediction: Filler-gap dependencies exhibit the clearer predictive effects, *regarding*-phrases exhibit reliable yet slower and less committing predictive processing, and *with*-phrases reveal only high expectation for co-reference only in production.

3.2 Experiment 1: Requirement for co-reference in acceptability

To make sure that the NP embedded in Hebrew *regarding*-phrases indeed constitutes a pragmatically prominent antecedent, which requires the next clause to be about it, I conducted two acceptability judgments experiments. In Experiment 1A I test whether a backward referring expression is required following *regarding*- and *with*- phrases. I compare the two dependency types to control for a possible general preference for repeated use of previous discourse referents. This experiment, therefore, tests the acceptability of sentences containing a *regarding*-phrase (as in (25b) above) and a *with*-phrase (as in (25c) above) with and without a subsequent pronoun. If a backward referring expression is pragmatically required following a *regarding*-phrase, the contrast between the pronoun and no-pronoun conditions should be greater in this structure.

In experiment 1B I test whether *regarding*-phrases also require that the pronoun appears at an early/discourse-prominent position. This experiment, therefore, tested the acceptability of sentences containing a *regarding*-phrase and a *with*-phrase with early (matrix) pronouns and late (embedded) ones. If there is preference for an early pronoun following a *regarding*-phrase, we expect the effect of the position of the pronoun to be greater in this structure than in the *with*-phrase structure.

3.2.1 Methods

Participants

Sixty-four native Hebrew speakers (according to self-report) volunteered for the study; 32 in Experiment 1A (mean age: 23.67, range 19-30), and 32 in Experiment 1B (mean age: 25.7, range 20-36). Fourteen participants were bilingual speakers of Hebrew and either Russian, English or Spanish (distributed similarly between the two sub-experiments), and the rest were monolingual.

Materials

Each of the two experiments included 20 experimental items of four conditions. The materials in Experiment 1A manipulated the type of antecedent (*regarding*- vs. *with*-phrase) and the realization of a pronoun referring back to it. The materials in Experiment 1B manipulated the type of antecedent and the position of the pronoun referring back to it (matrix/embedded). See Table 3.2 for sample items. Within each experiment, materials were divided into four lists in a Latin square design. Forty filler items of various acceptability statuses were added to each list (comprising a 60-item questionnaire).

Condition	Sentence
Regarding, Matrix pronoun (Exp 1A&B)	*ha-šotrim badku legabey ha-soed im ha-xašuda* the-cops checked regarding the-customer whether the-suspect *mi-šxunat ha-mecuka šixne'a oto lidxof et* from.the-neighbourhood.of the-poverty convinced him to.push ACC *ha-melcarit ba-misada ha-yokratit*

	the-waitress in.the-restaurant the-luxurious 'The cops checked **regarding the customer** if the suspect from the poverty neighbourhood convinced **him** to push the waitress in the luxurious restaurant'.
Regarding, No pronoun (Exp 1A)	*ha-šotrim badku legabey ha-soed im ha-xašuda* the-cops checked regarding the-customer whether the-suspect *mi-šxunat ha-mecuka šixne'a et ha-tabax* from.the-neighbourhood.of the-poverty convinced ACC the-cook *lidxof et ha-melcarit ba-misada ha-yokratit* to.push ACC the-waitress in.the-restaurant the-luxurious 'The cops checked **regarding the customer** if the suspect from the poverty neighbourhood convinced the cook to push the waitress in the luxurious restaurant'.
Regarding, Embedded pronoun (Exp 1B)	*ha-šotrim badku legabey ha-soed im ha-xašuda* the-cops checked regarding the-customer whether the-suspect *mi-šxunat ha-mecuka šixne'a et ha-melcarit* from.the-neighbourhood.of the-poverty convinced ACC the-waitress *lidxof oto ba-misada ha-yokratit.* to.push him in.the-restaurant the-luxurious. 'The cops checked **regarding the customer** if the suspect from the poverty neighbourhood convinced the waitress to push **him** in the luxurious restaurant'.
With, Matrix pronoun (Exp 1A&B)	*ha-šotrim badku im ha-soed im ha-xašuda* the-cops checked with the-customer whether the-suspect *mi-šxunat ha-mecuka šixne'a oto lidxof et* from.the-neighbourhood.of the-poverty convinced him to.push ACC *ha-melcarit ba-misada ha-yokratit.* the-waitress in.the-restaurant the-luxurious. 'The cops checked **with the customer** if the suspect from the poverty neighbourhood convinced **him** to push the waitress in the luxurious restaurant'.
With, No pronoun (Exp 1A)	*ha-šotrim badku im ha-soed im ha-xašuda* the-cops checked with the-customer whether the-suspect *mi-šxunat ha-mecuka šixne'a et ha-tabax* from.the-neighbourhood.of the-poverty convinced ACC the-cook *lidxof et ha-melcarit ba-misada ha-yokratit* to.push ACC the-waitress in.the-restaurant the-luxurious

	'The cops checked **with the customer** if the suspect from the poverty neighbourhood convinced the cook to push the waitress in the luxurious restaurant'.
With, Embedded pronoun (Exp 1B)	*ha-šotrim badku im ha-soed im ha-xašuda* the-cops checked with the-customer whether the-suspect *mi-šxunat ha-mecuka šixne'a et ha-melcarit* from.the-neighbourhood.of the-poverty convinced ACC the-waitress *lidxof oto ba-misada ha-yokratit.* to.push him in.the-restaurant the-luxurious. 'The cops checked **with the customer** if the suspect from the poverty neighbourhood convinced the waitress to push **him** in the luxurious restaurant'.

Table 3.2. Example stimuli from Experiments 1A and 1B.

Procedure

The questionnaires were built using Google forms. Participants filled the questionnaire online. They were instructed to rate the acceptability of the sentences on a 7-point scale, with 1 indicating that the sentence is not natural at all and 7 indicating a very natural sentence. Order of the items was randomly generated for each participant.

Data Analysis

Statistical analyses were carried out on z-transformed data. A linear mixed-effects model was fitted separately for each experiment, with antecedent type, resolution type and their interaction as fixed effect predictors (factors were sum coded). In addition, two pairwise comparisons were conducted to test the effect of the resolution type, independently in for each antecedent type (thus for each model *p*-values on pairwise contrasts are Bonferroni corrected for two comparisons). Due to convergence failure on the frequentist analyses, one slope component was removed from the random structure (main effect of resolution on by-item effects) in Experiment 1A. For Experiment 1B convergence was more problematic and the final model included only the one slope (main effect of resolution on by-item effects).

3.2.2 Results

Experiment 1A

Z-transformed rating means by condition are presented in Table 3.3 and mean raw ratings are presented in Figure 3.1. Results of the frequentist and Bayesian analyses are summarized in Table 3.4.

	Matrix pronoun	No pronoun	Embedded pronoun
Experiment 1A:			
Regarding-phrase	0.36 (0.72)	-0.24 (0.79)	-

With-phrase	0.67 (0.55)	0.51 (0.73)	-
Experiment 1B:			
Regarding-phrase	0.23 (0.73)	-	-0.02 (0.82)
With-phrase	0.62 (0.55)	-	0.63 (0.65)

Table 3.3. Mean (SD) z-transformed acceptability ratings by condition, Experiments 1A-B.

Both the frequentist and the Bayesian analyses revealed a significant main effect of antecedent type, such that dependencies with *regarding*-phrase antecedents were judged as less natural than those with *with*-phrase antecedents. The analyses also detected a main effect of pronoun realization, such that sentences with a pronoun were judged as more natural than those without one. Crucially, the interaction of the two factors was reliable. Pairwise comparisons (examining the effect of pronoun realization for each antecedent type separately) revealed a decrease in acceptability for the no-pronoun condition in *regarding*-type structures. The corresponding comparison for *with*-type dependencies was not robust on either analysis.

	Bayesian		Frequentist			
	Estimate [CrI]	Pr(β>0)	Estimate	SE	*t*	*p*
Main effect of antecedent	-0.27 [-0.38, -0.18]	> 99%	-0.276	0.047	5.91	< .001
Main effect of resolution	0.20 [0.11, 0.29]	> 99%	0.196	0.035	5.56	< .001
Interaction	0.10 [0.03, 0.18]	> 99%	0.106	0.038	2.77	.011
Regarding: Pronoun - NP	0.60 [0.36, 0.84]	> 99%	0.604	0.116	5.19	< .001
With: Pronoun - NP	0.18 [-0.05, 0.42]	94%	0.178	0.090	1.98	.095

Table 3.4. Results of the frequentist and Bayesian analyses, Experiment 1A. For the frequentist analysis**,** the table shows the estimate, standard error, *t*-value and Bonferroni-corrected *p*-value of the tested contrasts. For the Bayesian analysis it shows the mean of the posterior distribution of each parameter of interest, together with a 95% credible interval (CrI).

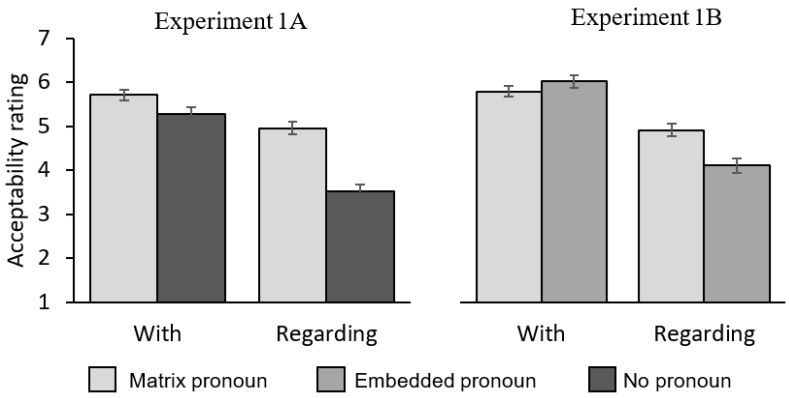

Figure 3.1. Mean acceptability ratings (on a scale of 1-7) by condition, Experiments 1A-B. Error bar represent +/-1SE.

Experiment 1B

Z-transformed rating means by condition are presented in Table 3.3 and mean raw ratings are presented in Figure 3.1. Results of the frequentist and Bayesian analyses are summarized in Table 3.5. Both the frequentist and the Bayesian analyses revealed a main effect of antecedent type, such that dependencies with *regarding*-phrase antecedents were judged as less natural than those with *with*-phrases. Crucially, the interaction of the two factors was reliable on both analyses. Pairwise comparisons (examining the effect of dependency length for each antecedent type separately) revealed a difference between the short and long versions of *regarding*-phrase dependencies, such that sentences with earlier resolutions were rated higher. A comparable contrast on the *with*-phrase conditions was not observed.

	Bayesian		Frequentist			
	Estimate [CrI]	Pr(β>0)	Estimate	SE	*t*	*p*
Main effect of antecedent	-0.26 [-0.36, -0.17]	> 99%	-0.264	0.025	10.5	< .001
Main effect of resolution	0.06 [-0.03, 0.16]	89%	0.060	0.043	1.38	.18
Interaction	0.07 [.007, 0.13]	98%	0.069	0.025	2.75	.006
Regarding: Matrix-Embedded	0.26 [0.03, 0.50]	99%	0.259	0.100	2.58	.02
With: Matrix-Embedded	-0.01 [-0.25, 0.22]	56%	-0.018	0.100	0.18	> .99

Table 3.5. Results of the frequentist and Bayesian analyses, Experiment 1B. For the frequentist analysis**,** the table show the estimate, standard error, *t*-value and Bonferroni-corrected *p*-value of the tested contrast. For the Bayesian analysis it shows the mean of the posterior distribution of each parameter of interest, together with a 95% credible interval (CrI).

3.2.3 Discussion

In Experiment 1, I tested the effect of the presence of a dependency and its length on acceptability ratings in sentences containing *regarding*- vs. *with*-phrase antecedents. The results of Experiment 1A suggest that *regarding*-phrases require a backward referring pronoun (in the subsequent clause), beyond the general preference for pronouns over new discourse referents. Moreover, the significant interaction in Experiment 1B indicates that a preference for early (matrix) co-reference (over late, embedded co-reference) also depends on antecedent type. I interpret these findings as evidence for a pressure to form an early co-referential dependency following *regarding*-phrases, over and above any such general preference in neutral contexts with no discourse prominent antecedents. In the following experiments I investigate the effects of this factor on predictions and their disconfirmations.

Incidentally, it might also be noted that, in both experiments, participants consistently rated the *regarding*-type dependencies lower than the *with*-phrase sentences. This might suggest that the usage of *regarding*-phrases is marked. This markedness could result from these structures' overall lower frequency in the language. In addition, there might be some additional discursive preference which applies to *regarding*-phrase sentences. It is possible that *regarding*-phrases are usually licensed by a fuller context than that available in one-sentence items. A corpus

survey indeed revealed that sentences of the form of our *with*-phrase interrogative constructions are much more common than the corresponding *regarding*-phrase ones. I tested the frequency of the structures in the HeTenTen corpus.[8] I found 16,591 instances of sentences containing a verb immediately followed by a *with*-phrase, which in turn is followed by an interrogative within the subsequent four words. The corresponding search for *regarding*-phrase constructions yielded only 377 instances. A Similar ratio was revealed when limiting the search to the verbs used in our materials (*with*-construction: 1,214 instances; *regarding*-construction: 56 instances). [9]

[8] HeTenTen is a web-corpus of about 10^9 Hebrew tokens (https://www.sketchengine.co.uk/hetenten-corpus/), which is part of the TenTen family of corpora of Sketch Engine (Jakubíček et al., 2013).

[9] The lower rating of *regarding*-type sentences relative to *with*-type sentences, could also reflect a contrast between the arguments and adjuncts. It could be that the referents in with-phrases get more precise and stable licensing within the matrix clause. Namely, since the role of regarding-phrases in the main clause is more general and pragmatic, such phrases might be harder to stabilize or integrate into a particular message. On this interpretation, the frequency contrast would reflect a grammaticalization process of this. I thank Whit Tabor for mentioning this interpretation to me.

3.3 Experiment 2: Prediction of co-reference in sentence completion

Although they provide a somewhat delayed measure of processing, sentence completion tasks are indicative of prediction as they allow manifestation of the predicted structure at the relevant point in the sentence. Experiment 2 thus aimed to examine how the predictability of different dependency types is reflected in production. The experiment included preambles of the four conditions exemplified in (25) above: wh-questions, assumed to present a syntactically active antecedent; *regarding*-phrase structures, assumed to present a discourse-prominent topic requiring a comment; *with*-phrase sentences, with an antecedent presented as a non-obligatory argument of the matrix clause; and bare if-questions, which eliminated the antecedent from the initial part of the sentence. I expected the production rate of co-referring elements to vary as a function of dependency prediction strength, such that antecedents which entail a dependency more strongly are expected to induce a higher rate of elements referring back to them.

3.3.1 Methods

Participants

Twenty-eight self-reported native Hebrew speakers, with a mean age of 30.31 (range: 20-53) volunteered for the study. Twenty-five of them were monolingual, and three were bilingual of Hebrew and either Arabic, Spanish or Italian.

Materials

Twenty sets of sentence-initial fragments in Hebrew were created for this experiment, each containing four conditions, manipulating the type of antecedent provided in the sentence fragment (see Table 3.6 for examples). Sentence fragments presented either a wh-question (using a *which-NP* phrase); an if-question preceded by a *regarding*-phrase; an if-question preceded by a *with*-phrase which was a non-obligatory argument of the matrix verb,[10] providing it a role of addressee to which the embedded question is directed; and bare if-questions, which eliminated the antecedent from the initial part of the sentence, leaving only one referent (the subject) in the matrix clause, and one in the embedded question. In all four conditions, the sentence was truncated following the verb of the embedded question.

Condition	Sentence
Which	*ha-mora badka eyze talmid ha-biryonim me-ha-kita* the-teacher checked which student the-bullies from-the-class *ha-makbila dirbenu* _____ the-parallel urged _____ 'The teacher checked which student the bullies from the other class urged _____'

[10] This was mostly an indirect object headed by the Hebrew counterpart of the preposition *with*, but in several sentences, it was realized as the direct object of the matrix verb.

Regarding	*ha-mora badka legabey ha-talmid im ha-biryonim* the-teacher checked regarding the-student whether the-bullies *me-ha-kita ha-makbila dirbenu* _____ from-the-class the-parallel urged _____ 'The teacher checked regarding the student if the bullies from the other class urged _____'
With	*ha-mora badka im ha-talmid im ha-biryonim me-ha-* the-teacher checked with the-student whether the-bullies from-the- *kita ha-makbila dirbenu* _____ class the-parallel urged _____ 'The teacher checked with the student whether the bullies from the other class urged _____'
If	*ha-mora badka im ha-biryonim me-ha-kita ha-makbila* the-teacher checked whether the-bullies from-the-class the-parallel *dirbenu* _____ urged _____ 'The teacher checked if the bullies from the other class urged _____'

Table 3.6. Example stimuli for Experiment 2.

Procedure

The questionnaire was built using Google forms, and participants filled it online. They were instructed to complete the sentence fragments as they wish, and to make sure they add at least one word.

Coding

Sentence completions were excluded from the analysis if they gave rise to an ungrammatical sentence, suggesting that the participant misread some part of the sentence fragment (removing four productions, comprising 0.71% of the data). The rest of the sentences were categorized based on the referent used as the first object of the embedded verb. I examined the rate at which participants produced, at that position, a backward referring element (BRE), either a gap or a pronoun, matching the critical antecedent.

A pronoun completion was considered a BRE if and only if the pronoun was the first argument produced for the provided verb and agreed in number and gender with the critical antecedent (26a). Pronouns which did not agree with the antecedent (*her* in 26b) or were not an argument of the critical verb (i.e. possessive pronouns modifying the verb's object, like *his* in 26c) were coded as "other pronouns".[11] Productions which provided a lexical NP as the first

[11] This category was defined to allow comparison with the if-clause baseline, where the critical antecedent was missing. We included the NPs with possessive pronouns in this category rather than in the NP category described below in order to allow a conservative estimate of the contrast with the baseline condition (i.e. to assume the highest possible rate of such pronouns in that condition).

argument of the verb were coded as NP, even if a pronoun or a gap occurred as an additional argument later in the production (26d).

(26) The kindergarten teacher forgot regarding the boy whether the bullies asked
  a. … him to push the girl.
  b. … her to punish him.
  c. … his sister to push him.
  d. … the girls to ignore him.

Productions were considered as containing possible gaps only in the wh-question condition and were coded as BRE if and only if there was a missing argument position for the provided verb adjacent to it, and no other possibly missing argument position in the production (27a). Namely, if the production included another verb which could be transitive, the gap was assumed to be an object of that verb (27b). In this kind of case, the provided verb does not have an argument which can be coded as any of the previously presented categories. Such completions, as well as cases of missing arguments in conditions other than the wh-question condition, were coded as intransitive uses of the provided verb (I use here the term intransitive to indicate that the verb had only a sentential complement with no direct object).

(27) The kindergarten teacher checked which boy the bullies asked
  a. … to climb the fence.
  b. … to help.

Note that in the *if*-question condition, production of a BRE is impossible since no relevant antecedent is available in the input. Thus, I consider the production rate of a pronoun of any kind in this condition as a baseline to which to compare the production rate of BREs in the other conditions.

Data analysis

A logistic mixed-effects model was fitted in R to the production data based on BRE completion (or pronoun completion in the if-question condition). I used a sliding contrast scheme to detect an increase in BRE completion rate for every condition relative to the previous level. I therefore report three pairwise comparisons (using the Bonferroni correction for three comparisons): *with*-phrase condition vs. the *if*-clause baseline, *regarding*-phrase condition vs. *with*-phrase condition, and *wh-question* vs. *regarding*-phrase condition. Due to repeated convergence failures on frequentist analysis, the final model included only one slope in the random structure (the *with-if* contrast on by-item effects).

3.3.2 Results

Distribution of BRE productions by condition is presented in Table 3.7 and Figure 3.2. I report pairwise comparisons between every two consecutive conditions, in order to test the contribution of each type of dependency to the rate of BRE completions. Results of the frequentist and Bayesian modelling are summarized in Table 3.8.

	BRE	Other pronouns	NP	Intransitive	

Wh-question	130	0	3	4	137
Regarding-phrase	99	13	20	7	139
With-phrase	99	8	29	4	140
If-question	-	33	104	3	140

Table 3.7. Production frequency of different completions in the four conditions, Experiment 2.

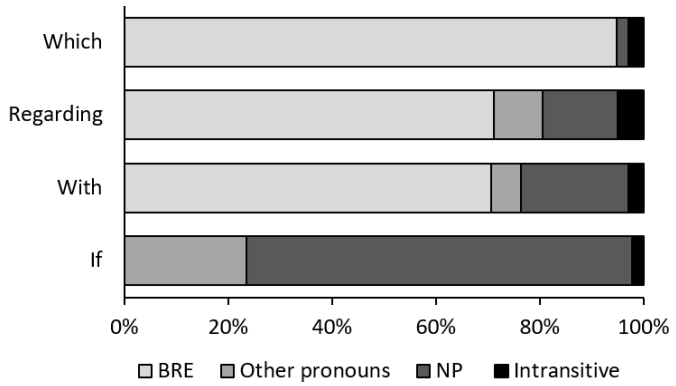

Figure 3.2. Production rates of different completions in the four conditions in Experiment 2.

On both analyses, BREs were produced in the *with*-phrase condition at a reliably higher rate than that of pronoun completions in the *if*-question condition. The analyses failed to detect a significant difference between the rate of BRE productions in *regarding*-phrase and *with*-phrase conditions. However, in wh-questions, BRE completions were significantly more common than in *regarding*-phrase sentences.

	Bayesian		Frequentist			
	Estimate [CrI]	Pr(β>0)	Estimate	SE	*t*	*p*
With - Baseline	44% [26, 59]	> 99%	2.96	0.58	5.13	< .001
Regarding - With	5% [-9, 19]	75%	0.05	0.31	0.18	> .99
Which - Regarding	34% [17, 50]	>99%	2.13	0.44	4.80	< .001

Table 3.8. Results of the frequentist and Bayesian analyses, Experiment 2. For the frequentist analysis, the table shows the estimate, standard-error, *t*-value and Bonferroni-corrected *p*-value of the tested contrast. For the Bayesian analysis it shows the mean of the posterior distribution of each parameter of interest, together with a 95% credible interval (CrI).

3.3.3 Discussion

Experiment 2 investigated to what extent speakers anticipate a dependency, namely a backward referring element, for different types of antecedents, via a production task. Results revealed a high proportion (over 70%) of backward referring elements in all cases where a contextually relevant referent was available in the sentence fragment, namely in wh-questions, *with*-phrase sentences, and *regarding*-phrase sentences. This completion type was far more frequent than

completion of other pronouns or lexical NPs. In addition, the baseline production rate of pronouns in the *if*-clause condition was significantly lower than the BRE production rate in the other three conditions.

Despite the high proportion of reference to the relevant antecedent in all the three non-baseline conditions, wh-questions were significantly the most constraining. Readers predicted a continuation with an immediate gap in almost all trials. I interpret this to reflect the effect of syntactic processes, which underlie the strong gap prediction. In contrast, the pragmatic constraint in the *regarding*-phrase condition did not seem to affect pronoun production rates significantly more than the general contextual availability of an antecedent as in the *with*-phrase condition. I return to discuss this in the General Discussion of this chapter (section 3.5).

Since wh-questions revealed a higher proportion of early reference completions, the absence of an effect for the pragmatic constraint in the *regarding*-phrase condition cannot be due to a ceiling effect. Therefore, it seems that the discourse-prominence of the antecedent does not assert a strong influence in this task. One possible (if unlikely) explanation for this is that, for some reason, information structure is not considered in prediction. Another possibility is that in this task, the parser tries to utilize every possible clue for generating the completion. If this is the case, a general prediction of a backward referring elements due to the availability of a referent (as in the *with*-phrase condition) is strengthened to match that of a discourse-prominent antecedent (in a *regarding*-phrase) in the current task. The next experiment examines whether reading time measures of active dependency formation can differentiate dependencies required by information structure considerations and those which are (not required but only) enabled by the context.

3.4 Experiment 3: Reanalysis costs

How does the processing difficulty associated with a failed gap prediction compare with other unrealized expectations? The third experiment examines the costs of disconfirmed expectations in online processing. For this purpose, I use the filled-gap design, in which a potential gap/BRE position is filled with a lexical NP. I expected the size of the "filled-gap effect", reflected in increased reading times on this NP, to vary as a function of dependency prediction strength, such that antecedents which entail a dependency more strongly would give rise to higher reanalysis costs.

3.4.1 Methods

Participants

Participants were 92 native speakers of Hebrew (according to self-report) (mean age: 25.37, range: 19-35). Seven participants were bilingual of Hebrew and either Russian or French, and the rest were monolingual.

Materials

Materials were based on sentences from the previous experiments, and included 24 sets, with four conditions each (see example sentences in Table 3.9). Experimental items followed the filled-gap design, namely a lexical NP appeared in the direct object position of the first embedded verb, where a gap/co-referential argument would be predicted. The gap/pronoun was eventually realized within the clausal complement of that verb. Conditions included the three antecedent types (*wh-question*, *regarding*-phrase, and *with*-phrase) and an *if*-question baseline in which this additional referent did not appear, eliminating the possible prediction of a backward referring element.

Condition	Sentence
Wh-question	*ha-ovdim ša'alu eyze kupai ha-menahelet ha-gvoha hixrixa* ***et*** the-workers asked which cashier the-manager the-tall forced **ACC** ***ha-šomeret*** *ha-xadaša lesalek ba-šavua* še-avar **the-guard** the-new to.banish in.the-week previous 'The staff asked which cashier the tall manager forced **the new security-guard** to throw out last week.'
Regarding -phrase	*ha-ovdim ša'alu legabey ha-kupai im ha-menahelet ha-gvoha* the-workers asked regarding the-cashier whether the-manager the-tall *hixrixa* ***et ha-šomeret*** *ha-xadaša lesalek oto ba-šavua* še-avar forced **ACC** **the-guard** the-new to.banish him in.the-week previous 'The staff asked regarding the cashier if the tall manager forced **the new security-guard** to throw him out last week.'

With-phrase	*ha-ovdim ša'alu et ha-kupai im ha-menahelet ha- gvoha* the-workers asked ACC the-cashier whether the-manager the-tall *hixrixa* ***et ha-šomeret*** *ha-xadaša lesalek oto ba-šavua* še-avar forced **ACC the-guard** the-new to.banish him in.the-week previous 'The staff asked the cashier if the tall manager forced **the new security-guard** to throw him out last week.'
If-question	*ha-ovdim ša'alu im ha-menahelet ha-gvoha hixrixa* ***et ha-šomeret*** the-workers asked whether the-manager the-tall forced **ACC the-guard** *ha-xadaša lesalek et ha-kupai ba-šavua* šeavar the-new to.banish acc. the-cashier in.the-week previous 'The staff asked if the tall manager forced **the new security-guard** to throw out the cashier last week.'

Table 3.9. Example set from Experiment 3.
The critical NP is marked in bold; ACC = accusative case marker.

The 24 experimental items were distributed in a Latin square design across four lists. Within each list, the target sentences were combined with 40 grammatical filler sentences, for a total of 60 sentences. Since the experimental items all presented a dependency spanning a long distance, I included several types of dependencies among the filler sentences, to prevent adaptation to the materials of the experiment. Twenty filler sentences contained a short dependency, namely an embedded question with a gap or pronoun following the first verb (five sentences for each of the four structures in the experimental design). Fifteen other filler sentences contained neither a filler-gap dependency nor pronominal co-reference. These too mirrored the structures of the experimental items,[12] with the exception of the *wh*-questions (which would be ungrammatical). The remaining five items were adjunct wh-questions.

Procedure

The experiment was a web-based self-paced reading experiment. Seventy-five percent of the sentences were followed by a 'yes/no' comprehension question.

Data Analysis

Participants were excluded from analysis if their performance on comprehension questions of experimental items was not significantly different from chance level (resulting in the removal of eleven participants), or if their average RT in experimental items was larger than 2.5 SDs above the group's average (resulting in the removal of three participants). For the remaining 78 participants, RTs higher than 2.5 SDs above the individual's average RT were trimmed to that cut-off (affecting 2.39% of the data). RTs shorter than 120ms were excluded (affecting 0.079% of the data).

[12] Note that in Experiment 1A, sentences containing a *regarding*-phrase with no subsequent pronoun received low acceptability ratings. For the current experiment, we constructed filler sentences of this type that would nonetheless be sensible and acceptable, by keeping the clause closely related to the *regarding*-phrase using strongly associated concepts like *apartment* and *lease* (e.g. "The real estate broker checked regarding the apartment whether the lease restricts animal housing").

I analysed RTs from the critical (filled-gap) region, as well as from the spillover word. I used a sliding contrast scheme of pairwise comparisons to detect an increase in RT for every condition relative to the previous level. Namely, for every region, three comparisons were made: *with*-phrase condition vs. *if*-baseline, *regarding*-phrase condition vs. *with*-phrase condition, and *wh-question* vs. *regarding*-phrase condition. Thus, on the frequentist analysis I corrected p-values using the Bonferroni correction for three comparisons. Due to convergence failures on the frequentist analysis, the final model of the spillover region included only intercepts and no slopes in the random effects structure. At the critical filled-gap region, the full random effects structure was retained.

3.4.2 Results

Word-by-word reading times for the different experimental conditions, and by-condition means on the tested regions, appear in Figure 3.3. I report pairwise comparisons between every two consecutive conditions, in order to test the contribution of each type of dependency to the filled-gap effect. The results of the frequentist and the Bayesian analyses are summarized in Tables 3.11 and 3.12, correspondingly.

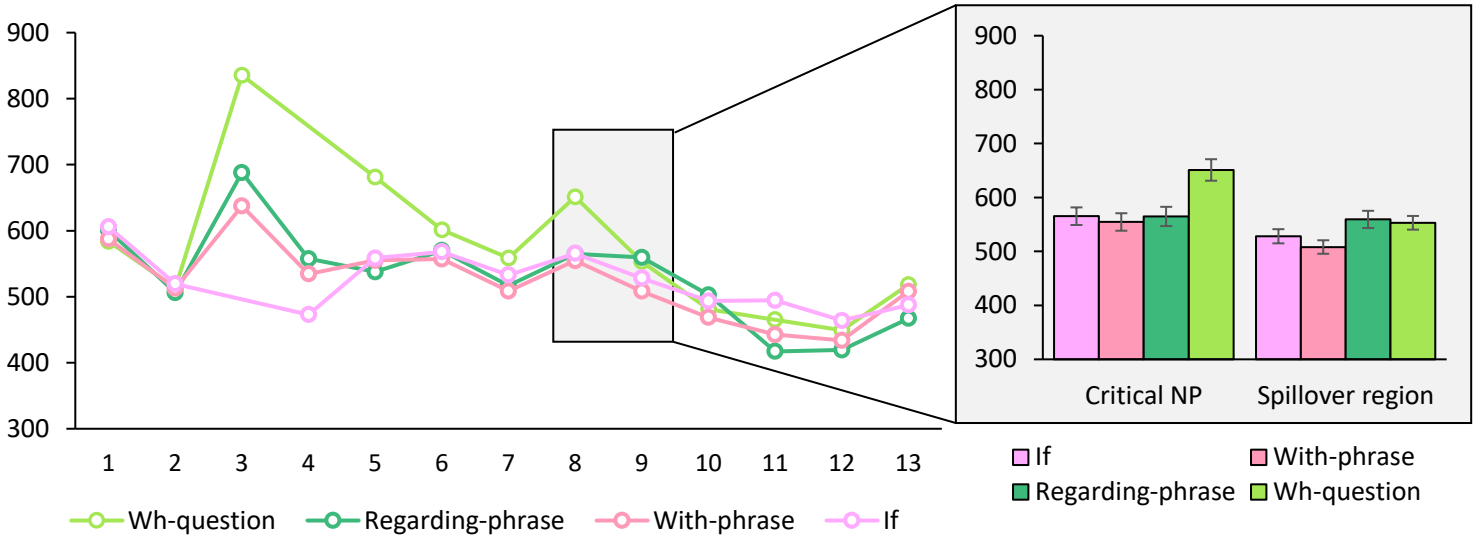

Figure 3.3. Word-by-word mean RTs (ms) by condition, Experiment 3. Error bars represent +/- 1SE.

|The-workers|$_1$ |asked|$_2$ |regarding the-cashier|$_3$ |whether|$_4$ |the-manager|$_5$ |the-tall|$_6$ |forced|$_7$ |ACC the-guard|$_8$ |the-new|$_9$ |to.banish|$_{10}$ |him|$_{11}$ |in.the-week|$_{12}$ |previous|$_{13}$

On the critical noun, both the Bayesian and frequentist analyses failed to detect a difference between the *with*-phrase condition and the baseline, or between the *regarding*- and *with*-phrase structures. However, a reliable increase in RT was observed (in both analyses) on the critical noun in wh-questions relative to *regarding*-phrase sentences. At the spillover region, there was still no evidence for increased RTs in the *with*-condition relative to baseline. However, the analyses did detect an increase in RTs in the *regarding*-phrase condition relative to the *with*-phrase condition. In that region, RTs in the *wh*-question condition did not differ from those in *regarding*-condition, on either analysis.

	Estimate	SE	*t*	*p*
Critical region:				
With-phrase vs. if-baseline	0.016	0.026	0.61	> .99
Regarding-phrase vs. with-phrase	0.004	0.027	0.15	> .99
Wh-question vs. regarding-phrase	0.077	0.026	2.92	.02
Spillover region:				
With-phrase vs. if-baseline	0.001	0.021	0.035	> .99
Regarding-phrase vs. with-phrase	0.063	0.025	2.52	.047
Wh-question vs. regarding-phrase	0.020	0.027	0.9	> .99

Table 3.11. Results of the frequentist analysis of Experiment 3: Estimate, standard-error, t-value and (Bonferroni-corrected) p-value of the tested contrasts.

	Posterior mean and CrI (ms)	Probability of the posterior beyond zero
Critical region:		
With-phrase vs. if-baseline	-12 [-41,16]	78%
Regarding-phrase vs. with-phrase	2 [-26, 29]	56%
Wh-question vs. regarding-phrase	60 [29, 92]	> 99%
Spillover region:		
With-phrase vs. if-baseline	-14 [-37, 8]	90%
Regarding-phrase vs. with-phrase	31 [6, 57]	99%
Wh-question vs. regarding-phrase	9 [-19, 37]	73%

Table 3.12. Results of the Bayesian analysis of Experiment 3: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

Results also seem to suggest differences in the effect size of the detected contrasts. Let us inspect the effect size estimates (derived from the Bayesian analysis) for the two statistically robust effects: the filled-gap effect detected for wh-questions (the contrast between wh-questions and *regarding*-phrases, at the critical word), and the one observed in the *regarding*-phrase condition (the contrast between *regarding*- and *with*-conditions, at the spillover region). There seem to be equivocal evidence that the magnitude of the effect in the wh-questions was larger than in regarding-phrases. The effect in wh-questions produced a numerically bigger estimate than that in *regarding*-phrases (with a posterior mean of 60 vs. 31). However, the CrI of those contrasts overlap on one side (see Figure 3.4).

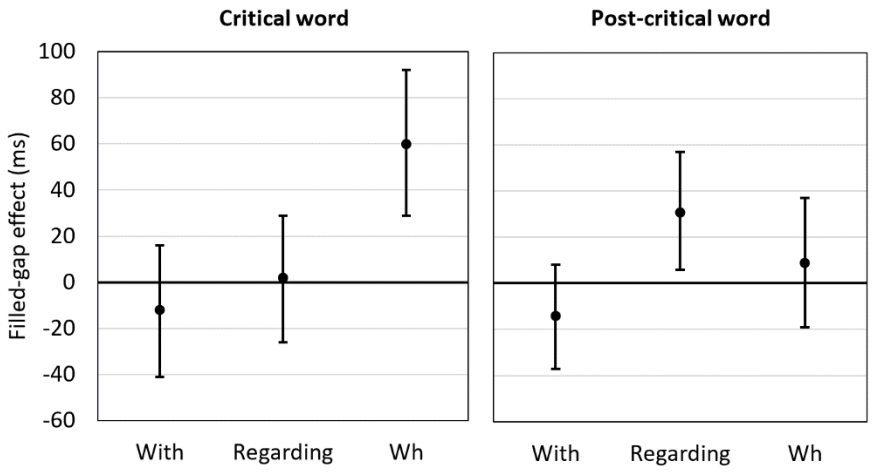

Figure 3.4. Posterior means and credible intervals (CrI) by condition, Experiment 3.

3.4.3 Discussion

In Experiment 3, I tested the processing costs ensued by a failed prediction for a gap/pronoun. I assume that, in the attested dependencies, reanalysis costs are an index for strength of the prediction. I observed active dependency formation in *wh*-questions, indicated by a filled-gap effect at NP following the first sub-categorizing verb. Importantly, this effect was observed above and beyond the costs in the corresponding region of *regarding*-dependencies. The *regarding*-phrase condition seemed to exhibit a belated "filled-gap" effect, with increased RTs compared to the *with*-phrase condition only at the spillover region. At this region, reanalysis costs were comparable in *regarding*-dependencies and wh-questions.

I suggest that the different patterns of the reanalysis costs reflect different degrees of prediction associated with pragmatically and syntactically motivated dependencies. Information structure biases (in the *regarding*-phrase condition), create a preference for an early argument referring to that antecedent. This incites some prediction for a dependency, which is then reanalysed following the filled-gap NP. Yet, syntactic licensing pressures (in filler-gap dependencies) result in a faster and/or stronger prediction for a gap, as expressed in earlier (and possibly larger) reanalysis costs.

It should also be mentioned that in this experiment, I did not find support for predictive dependency formation following non-prominent antecedents. The experiment did not detect increase in processing costs for the *with*-phrase condition relatively to the baseline case, where no initial antecedent is available. Namely, we failed to observe reanalysis costs when the antecedent is neither syntactically nor pragmatically prominent.

Differences in reanalysis effects reflect degrees of commitment

Processing difficulty was observed immediately (on the filled-gap noun) in the wh-condition, but in the spillover position in the *regarding*-phrase condition. How can we make sense of the divergence in the position of reanalysis? The nature of the self-paced reading task makes it hard to establish the fine-grained time-course of reanalysis. Thus, the belated effect could reflect a delay in different phases of the reanalysis process. One possibility is that in

pragmatically-motivated dependencies (relative to syntactically-motivated ones), the parser is late to detect the required reanalysis. This might mean that readers allocate less attention to confirming the well-formedness of such dependencies (relative to filler-gap dependencies).

It is also possible that the late reanalysis effect in *regarding*-phrase sentences does not reflect a delay in detecting the need for reanalysis, but rather a delay in more advanced phases of the reanalysis. Namely, the delayed effect could indicate that readers more easily postpone full reanalysis in processing these pragmatically-motivated dependencies. This could suggest that reanalysis costs in this case are not too severe, so that readers still engage in parsing of the input while reanalysing. In turn, this simultaneous forward and backwards processing makes the effect "spread out" to the following region.

Overall, I suggest that the motivation for structure building affects the parser's commitment to it and thus the costs of its reanalysis. Based on these results, it seems that lower commitment, in pragmatically-motivated dependencies, results in reallocation of some reanalysis costs to a later stage of processing. This could also be in line with the trend of effect size difference.

Parallel and serial processing of committing predictions

Fully parallel processing models should suggest that the effect we observe does not reflect commitment to the gap/pronoun prediction, but rather lack of prediction of the lexical NP which occurred in the input. If we keep all continuations active (proportional to their gradient probability), the only probability that should matter is that of the actual input. Namely, if the parser does not commit to one specific prediction, processing difficulty should be a not function of a specific unrealized prediction (in the terms of lexical prediction: it should reflect cloze probability effect, but not sentential constraint). Surprisal theory (Hale, 2001; Levy, 2008a) posits that the cost of processing a word in the input depends on that word's conditional probability in the sentential contest (i.e. the probability of that word given the previous input). Referring back to the critical antecedent and producing a full NP are not in complementary distribution since there is also the possibility of using another pronoun, referring to a different entity. Thus, it could be that the constructions we investigate are also distinguished by the predictability of the filled-gap element, and this might account for the results of Experiment 3. Indeed, we observed in Experiment 2 different rates of NP production across the different conditions (rate of NP production was 74% in the if-condition; 21% in the *with*-phrase condition, 14% in the *regarding*-phrase condition, and 2% in the *wh*-question condition).

To test the hypothesis that the results of Experiment 3 are driven by the probability of an NP in the filled-gap position, I conducted an additional analysis of the experiment's results, using a surprisal measure derived from the production results. As 20 out of the 24 sets of Experiment 3 were based on materials from Experiment 2, we can derive by-item surprisal measures for most items in the four conditions. If surprisal drives the contrast between the conditions, we might expect that this measure would account for the data better than our experimental contrast. However, the results failed to exhibit a significant effect of NP-completion rate at the critical word (Estimate = -0.021; SE = 0.042; t = 0.51; p = .61; Posterior mean and CrI: -21[-69, 28]; Probability of the posterior beyond zero: 80%) and at the spillover region (Estimate = 0.029; SE = 0.038; t = 0.77; p = .44; Posterior mean and CrI: 9[-32, 48];

Probability of the posterior beyond zero: 67%). Moreover, even after adding the NP-production rate as a regressor, the effects of the categorical contrasts were reliable and exhibited exactly the same pattern as in the original analysis. This suggests that there is indeed some commitment to the pronoun/gap prediction above and beyond the plausibility of a lexical NP in this position.

It is also possible to consider a partially parallel model in this context, where low probability predictions are abandoned. Namely, the weighed set of predictions includes only a subset of the (higher ranking) predictions. On this interpretation, it is possible that the NP prediction is lost in some of the conditions (due to its low probability). In these cases, the filled-gap NP should create reanalysis costs beyond its raw predictability. However, I find it unclear why such a process would distinguish wh-questions and *regarding*-dependencies from the *with*-phrase condition and the *if*-baseline, given the production probabilities mentioned above. More natural cut-off points seem to distinguish wh-questions from the rest of the conditions.

Another possible interpretation of the commitment level could be one where the lower commitment to a pragmatic prediction is reflected by lower probability, across trials, of actively forming the relevant structure/interpretation. This hypothesis would be in line with a stochastic serial parser like the one suggested by the Unrestricted Race model (van Gompel, Pickering, & Traxler, 2000). This scenario would predict the trend observed in the effect size, whereby the detected reanalysis costs on the critical region of wh-questions seem slightly larger than the cost on the regarding-phrase condition (at the spillover region). The Unrestricted Race model would suggest that on average, there are less reanalysis trials in the *regarding*-phrase condition, which may drive differences in average reanalysis estimates. However, this proposal is not in line with the delay of the reanalysis costs. If on each trial reanalysis is binary (it either occurs or not), there is no reason for slower computation or spreading of the effect. Thus, if this finding of timing differences is reliable, it might favour the last remaining interpretation of commitment level, namely slow or partial predictive processes which are in turn reverted relatively easily.

Thus, it seems that the interpretation which fits with the results most naturally is the one where a more committing predictive process provides a fuller representation. This could be implemented by differentially limiting the extent of the predictive processing. For example, the parser could limit prediction to a general conceptual association rather than a full syntactic structure, or vice versa, predict a pronoun without establishing full thematic/semantic relations. In addition, a partial representation could also reflect simply a slower prediction process. If the timescale for predictive processing is slower on pragmatically motivated dependencies, it would arrive at a full representation by the time the disconfirmation arrives. However, these conclusions are somewhat tentative and require additional investigation. Thus, Experiment 4, presents additional support for this interpretation and consider again the stochastic account of prediction levels (in order to refute is eventually).

3.6 Experiment 4: Semantic persistence

In the previous section I tested whether predictions motivated by pragmatics and syntax generate similar reanalysis costs, as a measure of the parser's commitment to such predictions. I observed significant reanalysis costs at the disconfirmation of pragmatically predicted dependencies. Yet an additional, earlier cost was observed for syntactically motivated predictions. I suggested that the parser exhibits lower commitment to predictions when they are motivated by discourse preferences rather than by a need for grammatical licensing. This gives rise to the following question: Does the contrast between reanalysis costs reflect a qualitative difference in predictive processing, or a quantitative difference in prediction or in reanalysis costs?

There are several ways in which the parser could establish gradient commitment levels. In this section I focus on the two serial processing options: slow/limited predictive processing, and low rate of predictions. The low rate interpretation suggests that variability across trials would manifest as relative ease in reanalysis of pragmatic dependencies. On this approach the serial parser is stochastic and the rate of choosing a certain analysis or executing predictive processing depends on the parser's certainty. Thus, the parser can decrease the overall rate of acting upon some predictions if it has lower commitment for them (i.e. in pragmatically motivated predictions). Alternatively, we could assume that the probability of executing predictive processing (and thus the probability of reanalysis) is the same on syntactically and pragmatically motivated dependencies. Thus, the difference between these reflects a qualitatively difference type of reanalysis. If for example the prediction does not go through all the stages of processing before the parser gets to the cue for reanalysis, this should make that reanalysis easier.

I try to distinguish these interpretations by testing whether disconfirmation of pragmatic prediction is modulated by factors which affect the difficulty of classic syntactic reanalysis. The core idea is that if syntactically- and pragmatically-motivated prediction and reanalysis involve exactly the same processes, and are distinguished only by frequency of acting upon the prediction, then the same effects should occur on the two types of reanalysis. Reanalysis of the two dependency types should be modulated by the same factors, and to a similar degree. However, if predictive processing is somewhat shallower on pragmatically motivated dependencies, our ability to experimentally manipulate the reanalysis costs should be limited. Namely, we should observe an interaction such that the *regarding*-phrase conditions exhibit not only easier reanalysis but also lower sensitivity to factors which are known to affect reanalysis costs.

One such factor, suggested to contribute to the difficulty of reanalysis, is the semantic compatibility of the initial reading and the globally correct one. The initial interpretation persists even after reanalysis, as reflected in the offline (erroneous) comprehension of garden path sentences (Christiansen, Hollingworth, Halliwell, & Ferreira, 2001; Patson, Darowski, Moon, & Ferreira, 2009; Huang & Ferreira, 2020). The need for inhibition of the initial reading could be manipulated experimentally by changing the compatibility of the two semantic representations. For example, Christiansen and colleagues (2002) did this by using sentences like (28). In sentence (28a), even after constructing the globally correct sentence structure, it is

still conceivable that Bill tried to hunt the deer (congruent interpretations), but in (28b) it seems implausible (incongruent interpretations). Indeed, this manipulation affected the persistence of the initial meaning in offline comprehension, with lower rates of persistence in the incongruent interpretations condition. Later, Sturt (2007) found in an eyetracking study that (in)congruency of the two readings affects the reanalysis of other structural ambiguities too (using the NP/S ambiguity, see example 22 and details on p. 39), and has a mark on reading times as well.

(28) a. While Bill hunted the deer ran into the woods
b. While Bill hunted the deer paced in the zoo

I manipulate the congruency between a reading with a gap/pronoun at the first object position and at the embedded object position using pairs like (29). In this way, I aim to test how semantic persistence affects the reanalysis of syntactically- and pragmatically-motivated dependencies.

(29) Sarah asked…
a. {Which boy | Regarding the boy if} the teacher liked praising {_/him}
b. {Which boy | Regarding the boy if} the teacher liked punishing {_/him}

I conducted two pre-tests in order to make sure that the sentences in each set are distinguished by the (in)congruency of the initial and final interpretation, but not in the overall plausibility of the final reading. I initially constructed 38 sets and then used the pre-test to choose the most fitting 24 out of these. On the first pre-test, I presented sentence pairs which represented the initial interpretation and the final one without a long-distance dependency (30). Participants were asked to rate on a seven-point scale the extent to which the meanings of the two sentences align. They were instructed to give a rating of seven if the sentences have an identical meaning, to choose one if the sentences contradict each other, and to use the intermediate scores for partial match/mismatch. The sets were divided to two lists such that participants saw congruent and incongruent pairs but never saw both versions form the same set. On the second pre-test, I used only the versions of the final readings and asked participants to rate the plausibility of each sentence. Participants saw only one item from each set and were instructed to use a seven-point scale, with one marking completely implausible sentences and seven marking very plausible sentences. The properties of the 24 chosen sets are provided on Table 3.14.

(30) a. The teacher liked the boy ---------- The teacher liked praising the boy
b. The teacher liked the boy ---------- The teacher liked punishing the boy

	Congruent	Incongruent	Comparison:
Compatibility with the initial meaning (of an early object gap/pronoun)	6.45 (0.37)	3.15 (0.95)	$t = 14.62$, $p < .001$
Plausibility of the final interpretation	5.10 (1.14)	5.08 (0.99)	$t = 0.002$, $p > .99$

Table 3.13. Pretest results for the experimental sets of Experiment 4: Mean (SD) ratings for the chosen sets and statistics of their comparison.

3.6.1 Methods

Participants

Participants were 52 native speakers of Hebrew (according to self-report) (mean age: 24.85, range: 19-35). Six participants were bilingual of Hebrew and either Spanish, Russian or English, and the rest were monolingual.

Materials

The experiment included 24 item sets of four conditions (see example set in Table 3.14), manipulating the type of the dependency (wh-question vs. *regarding*-phrase) and the semantic compatibility of the final reading with the initial one (congruent vs. incongruent). Note that the semantic (in)congruency manipulated here is between the initial interpretation of the filler as the object of the first verb (*engineer planned the house*) and the final interpretation of it as the object of the next verb (*engineer planned to ruin/build the house*).

In contrast to other filled-gap cases, the word initiating the reanalysis was not an NP, but rather a non-finite verb. The verb's meaning was used for the congruency manipulation. I could therefore provide the cue for reanalysis and the modulation of the interpretation in the same sentential region, i.e. the verb. A one-word adverb was inserted as a buffer between the first verb and the critical verb in order to allow time for constructing the prediction and committing to it.

To create the ambiguity, the first verb had to accommodate an object gap position on one interpretation, and a non-finite complement clause on the other, without allowing both complement types concurrently. This highly restricted the inventory of possible verbs. To avoid repetition, I included also verbs which could accommodate a gap in an indirect object position. In these sets (7 out of the 24) the second verb had the same type of prepositional complement (e.g. *with which boy did the teacher agree to play*).

Condition	Sentence
Congruent, wh-question	*Neta šiara eyze bayit ha-mehandes ha-baxir tixnen* Neta speculated which house the-engineer the-senior planed *kvar **livnot** ba-šxuna ha-mitxadešet* already **to.build** in.the-neighborhood the-renewing 'Neta guessed which house the senior engineer planned to build on the neighborhood renovation'.
Incongruent, wh-question	*Neta šiara eyze bayit ha-mehandes ha-baxir tixnen* Neta speculated which house the-engineer the-senior planed *kvar **laharos** ba-šxuna ha-mitxadešet* already **to.destroy** in.the-neighborhood the-renewing 'Neta guessed which house the senior engineer planned to ruin on the neighborhood renovation'.

Congruent, *regarding*-dependency	*Neta šiara legabey ha-bayit im ha-mehandes ha-baxir* Neta speculated regarding the-house if the-engineer the-senior *tixnen kvar* ***livnot*** *oto ba-šxuna ha-mitxadešet* planed already **to.build** it in.the-neighborhood the-renewing 'Neta guessed regarding the house whether the senior engineer planned to build it on the neighborhood renovation'.
Incongruent, *regarding*-dependency	*Neta šiara legabey ha-bayit im ha-mehandes ha-baxir* Neta speculated regarding the-house if the-engineer the-senior *tixnen kvar* ***laharos*** *oto ba-šxuna ha-mitxadešet* planed already **to.destroy** it in.the-neighborhood the-renewing 'Neta guessed regarding the house whether the senior engineer planned to ruin it on the neighborhood renovation'.

Table 3.14. Example set from Experiment 4.
The critical verbs are marked in bold.

The experimental items were distributed in a Latin square design across four lists. Within each list, the target sentences were combined with 40 grammatical filler sentences, for a total of 64 sentences. The filler items included 10 wh-questions with an early object gap position, 10 if-clauses of a simple verbal structure, and 20 sentences with embedded declaratives which included non-finite verbs. The verbs on the wh-question fillers were of similar properties to those used in the experimental items (generated from the sets which were excluded on the pre-test phase). This was done to avoid adaptation to the experimental sentence structure.

Procedure

The experiment was a lab-based self-paced reading experiment. Fifty percent of the sentences were followed by a 'yes/no' comprehension question.

Data analysis

I analyzed the reading times of the critical (embedded) verb and of the word following it (as the spillover region), in two separate models. The experimental fixed effects in this analysis, dependency type and congruency, were sum coded. In addition, within each model, I conducted pairwise comparisons contrasting the congruent and incongruent levels of each dependency. Thus, *p*-values on the frequentist analysis were corrected using Bonferroni correction for two comparisons. Due to convergence failure, the final model of the spillover region included only two slopes on the critical region analysis (the main effect of dependency on both by-item and by subject effects) and only one slope on the spillover region (the main effect of congruency on by-subject effects). On the verb region, the final model included only the slope corresponding to the dependency type factor, on both by-subject and by-item random effects.

After this analysis was conducted, I was concerned that the effect might derive from a possible misalignment in the timing of the effect in the two dependency types. Since in Experiment 3 the reanalysis appeared one region earlier for wh-questions, I conducted an

additional analysis where the filled-gap verb was considered the critical region of wh-question conditions, but for *regarding*-dependencies the critical region was the word following that verb. Spillover results were adjusted accordingly, using the first word after the verb for wh-questions, and the second word after the verb for regarding-dependencies. The effects on these (non-matching) regions were analysed in two additional models and produced the same results, see Appendix B.

Exclusion criteria resulted in removal of five participants (four for low accuracy in comprehension questions and one for abnormal reading times). For the remaining 48 participants, trimming of high and low RTs affected 2.69% and 0.02% of the data, correspondingly.

3.6.2 Results

Word-by-word reading times for the different experimental conditions, and by-condition means on the spillover region, appear in Figure 3.5. The results of the frequentist and the Bayesian analyses are summarized in Tables 3.15 and 3.16, correspondingly.

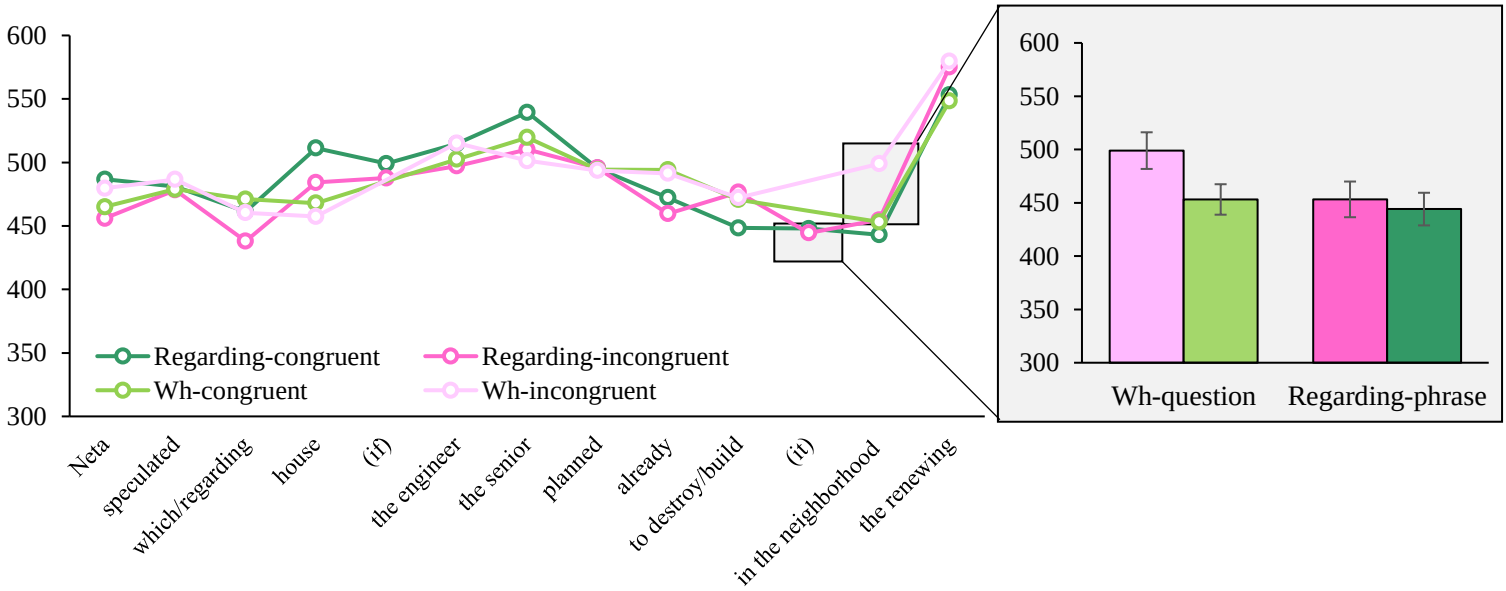

Figure B1. Word-by-word RT means (ms) by condition, Experiment 4. Error bars represent +/-SE.

The analysis of the critical region did not reveal any effects on either of the analyses. At the spillover region, both the frequentist and Bayesians analyses detected a main effect of dependency type, such that reading times were faster on regarding-phrase conditions relative to wh-questions. I also observed a reliable main effect for congruency. When the embedded verb casted a meaning incongruent with that of initial analysis, reading times were slower. The crucial interaction was significant only on the frequentist analysis, and not on the Bayesian one. However, both analyses converged on the pairwise comparisons again, exhibiting a congruency effect for wh-questions, but failing to detect the parallel contrast in the *regarding*-phrase dependencies.

	Estimate	SE	*t*	*p*
Main effect of dependency	-0.022	0.008	2.23	0.006
Main effect of congruency	-0.021	0.010	2.77	.033
Interaction	0.019	0.008	2.43	.015
Pairwise semantic persistence comparisons:				
Wh-question	-0.082	0.025	3.23	.003
Regarding-phrase	-0.005	0.025	0.19	> .99

Table 3.15. Results of the frequentist analysis of Experiment 4, at the spillover region: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-value of the tested contrasts.

	Posterior mean and CrI (ms)	Probability of the posterior beyond zero
Main effect of dependency	-13 [-22, -5]	> 99%
Main effect of congruency	-10 [-20, 0]	97%
Interaction	7 [-2, 17]	94%
Pairwise semantic persistence comparisons:		
Wh-question	42 [16, 67]	> 99%
Regarding-phrase	12 [-13, 36]	82%

Table 3.16. Results of the Bayesian analysis of Experiment 3: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

3.6.3 Discussion

The results of Experiment 4 suggest that semantic persistence differentially affects syntactically- and pragmatically-motivated dependencies. In wh-questions, a meaning shift increases the costs of a filled-gap reanalysis. Namely, reanalysis costs were more severe when the disambiguation entailed a semantic interpretation inconsistent with the initial one. However, I did not observe evidence for a corresponding effect in *regarding*-dependencies. In addition, the processing of these dependencies was overall easier than that of wh-questions. This suggests that the parser does not require extra processing effort for inhibition in *regarding*-dependencies.

The lack of semantic persistence (or inhibition costs) in pragmatically-motivated dependencies signifies that there is something qualitatively different in the prediction and/or reanalysis process in this dependency type. Namely, pragmatically-motivation consistently generate predictions in a different way than syntactic one – either more restricted or slower representation building. This is in line with the notion, promoted in this chapter, that the parser is less committed to pragmatically-motivated predictions. It suggests that this difference is not likely reducible to differences in the rate of prediction. Under the logic of a stochastic serial parser, reanalysis estimates could seem smaller when averaging across trials with a lower rate

of prediction/reanalysis. However, on such an account, since each act of predictive dependency formation is completely committing, the congruency manipulation should affect the ease of reanalysis, whenever reanalysis is needed. Therefore, effects of semantic persistence/inhibition should be similar in high rate (syntactic) predictions and in low rate (pragmatic) ones, contrary to my findings. Thus, the different commitment level should probably be implemented as a qualitative difference in the prediction, where only partial representations are predictively formed by the reanalysis point.

The current findings may also suggest a possible locus of the difference between the predictive processes. Since reanalysis of pragmatically-motivated dependencies seems to depend less on interpretative (dis)similarity, we could postulate that syntactically-motivated predictions involve a richer, or more precise, semantic representation. If pragmatically-motivated predictions do not trigger a full semantic interpretation, the inhibition part of the reanalysis should be less costly (as the meaning is less activated to begin with). This should also result in overall easier reanalysis. However, at the moment this hypothesis is merely speculative. The low commitment to pragmatic predictions could manifest in different ways, and shallow or slow processing on other levels of representation could also result in a similar pattern.

It could be interesting to test whether disconfirmation of pragmatic predictions reveals other effects which reanalysis canonically manifests. For example, several studies have shown that increasing the distance between the onset of the local ambiguity and the disambiguation site can make Garden Path reanalysis harder (Ferreira & Henderson, 1991; Tabor & Hutchins, 2004). Does the commitment to the initial parse increase with time in pragmatically-motivated predictions, as it does for Garden Path sentences? It is also well-known that the costs of prediction and reanalysis reduce acceptability ratings of syntactically-motivated dependencies (e.g. for Garden Path sentences - Warner & Glass, 1987; Ferreira & Henderson, 1991; and for filler-gap dependencies - Sprouse, 2008; Keshev & Meltzer Asscher, 2019). Is such acceptability penalty mirrored with *regarding*-dependencies?

Overall, Experiment 4 presents additional evidence for the parser's differential commitment level to pragmatically- and syntactically-motivated dependencies. I suggest that the reanalysis of pragmatically-motivated dependencies is less costly, and therefore it inflates reading times by a smaller margin and produces less semantic consequences. Yet it should be noted that these conclusions are somewhat hindered by the non-reliable interaction on the Bayesian analysis (yet see Appendix B for the alternative spillover analysis).

3.7 General discussion

The current study aimed to clarify the role of the motivation for prediction in reanalysis costs. Different models of human language processing converge in assuming that the parser is often reluctant to modify an expected structure. However, it is well-known that some reanalyses are harder than others. Moreover, previous studies failed to detect behavioural costs for disconfirmed lexical predictions, namely predictions which are not based on the need to construct a syntactic structure. This contrasts with the robust reanalysis effects in Garden Path sentences and in long-distance filler-gap dependencies. One possibility is that this contrast could be traced back to the different motivations for prediction, where licensing syntactic structure might be more committing than establishing a specific context-appropriate meaning. However, it could also reflect differences in monotonicity and transparency of the reanalysis site. To investigate this topic, I used more closely matched cases than previous studies. My findings suggest that discourse-based predictions can be committing enough to result in significant reanalysis costs. However, reanalysis of pragmatically-motivated dependencies involves different processes relative to that of syntactically-motivated ones, and it incurs lower costs.

3.6.1 Degrees of prediction and reanalysis

In this study I examined the prediction of long-distance dependencies with three different antecedent types: (i) fillers, which syntactically require co-reference; (ii) antecedents in *regarding*-phrases, which initiate a pragmatically-based prediction for a co-referring element; and (iii) antecedents which are contextually available but do not prompt any syntactic or discursive incentive for dependency formation (in *with*-phrases).

I observed evidence that the parser predicts an early resolution for filler-gap dependencies, in both production (Experiment 2) and comprehension (Experiment 3). In the case of discourse-prominent antecedents (in *regarding*-phrases), I also found a substantial (though lower) rate of dependency prediction in production. These structures also exhibited a belated and smaller filled-gap effect. Taken together, the results suggest that predictive dependency formation occurs with both pragmatic and syntactic motivation, but with different degrees of commitment, which affect the timescale and costs of the reanalysis. In addition, the results of Experiment 4 revealed that filled-gap effects in wh-questions, but not reanalyses of *regarding*-dependencies, were sensitive to the semantic compatibility of the initial and final readings. This suggests that inhibition of the initial interpretation is easier in *regarding*-dependencies, and thus that this interpretation is less committing.

Furthermore, general contextual prediction (represented in the *with*-phrase condition of Experiments 2 and 3) exhibits yet another level in the gradient level of dependency prediction. First, note that the measure of reanalysis cost (in Experiment 3) in *regarding*-dependencies was taken with reading times in the *with*-phrase condition as a baseline. Therefore, the effect in *regarding*-dependencies presents costs above and beyond any reanalysis which may have occurred in the case of predictions purely based on context availability. In addition, while the production experiment revealed a strong preference for using co-referential arguments even in

the *with*-phrase condition (in line with the level of prediction in *regarding*-dependencies), I failed to detect a cost for disconfirmed predictions for this sentence type in Experiment 3. I believe there are two ways in which this seeming contradictory pattern of results can be settled: it could result from an inflated estimate of prediction processes in production tasks, or from lack of reanalysis costs in such contextual predictions.[13]

Considering the first possibility, production tasks may bias participants to use any possible contextual cue, because using given information can facilitate production, removing the need to come up with an additional discourse referent and retrieve a new noun from the lexicon with little cues. In comprehension, on the other hand, participants are not required to think up a referent by themselves, but merely need to recognize it in the text. Therefore, in production-based tasks, the strong preference for re-usage of referents might give rise to effects similar to those generated by a pragmatically-motivated prediction for co-reference.

Alternatively, it could be that the *with*-phrase condition does initiate prediction processes, which were nevertheless not detected in the filled-gap experiment. It could be that the prediction is not generated fast enough, and thus affects only late measures like production rates (i.e. if the prediction is not made by the time the parser encounters the "filled-gap" NP, no reanalysis is required). In addition, the context-based prediction in the *with*-phrase condition might generate a qualitatively different process of low commitment prediction. Low commitment of the parser to the prediction can mean, for example, that a conflict between this prediction and the actual input is not costly at all, in line with different cases of lexical prediction, which failed to exhibit robust behavioural costs upon disconfirmation.

In the lexical prediction literature, an important distinction has been made between ease of integration due to partial overlapping activation, and predictive pre-activation of a constituent (Van Petten & Luka, 2012). Ease of integration is exhibited in comparing completions which are congruent or incongruent with a prediction (e.g. using the N400 effect). This is believed to indicate low level prediction (possibly via passive spreading activation from previous words in the context). However, if a lexical prediction is strong and specific enough, the processing of a different completion may exhibit, in ERP measures, some disruption relative to cases where no specific prediction was made. Thus, prediction in *with*-phrase cases could either be passive, in the sense that it elicits only ease of integration effects, or comparable to specific lexical predictions, which exhibit disconfirmation effects only in ERP measures.

3.6.2 What does "commitment level" mean?

I considered several possible interpretations for the parser's "commitment level": scaled weights of multiple predicted structures in a parallel processing model; rate of implementing

[13] Note that other factors could also constitute a possible source for the contrast between Experiments 2 and 3, so additional caution in interpreting this contrast is required. The factors could be:

(i) Demographics: the mean age of participants was higher in Experiment 2. I do not know of any reason to suspect that older readers over-predict only in the general context conditions (*with*-phrase sentences). However, one could speculate that this contributes to the contrast.

(ii) Statistical power: It is possible that the filled-gap effect in the contextual prediction is smaller, and that Experiment 3 did not have enough statistical power to detect this subtle difference.

the predictive structure building in a serial processing model; and pace/completeness of the predictive processing (again within serial processing). Each of these approaches could result in higher average measure of reanalysis costs in syntactically-motivated dependencies relatively to pragmatically-motivated ones.

Results of Experiment 3, however, seem to point against the parallel model. In parallel processing, all possible completions should be weighed. Therefore, the processing difficulty at the filled-gap NP should not be a direct function of co-reference predictability. Instead it should align with the predictability of the filled-gap NP. However, this factor did not seem to correlate with the reading times in Experiment 3, and did not reduce from the effect of the categorical distinction between the dependency types.

When considering the two serial processing options, the results of Experiment 4 favor the model in which the commitment level is implemented as the depth or speed of predictive processing. The interpretation of commitment level as likelihood of generating the structural prediction is in line with models such as the Unrestricted Race model (van Gompel, Pickering, & Traxler, 2000). On such an approach, reanalysis should occur on a smaller portion of the trials on *regarding*-phrase conditions relative to wh-questions. However, reanalysis trials are expected to be sensitive to similar manipulations since the process of revising the initial prediction proceeds in a similar manner in both cases.

On the other hand, we could describe the parser commitment as the extent to which the predicted dependency rapidly gets through various stages of analysis (syntactic structure building, construction of semantic meaning etc.). This could be similar to the Construal theory (Frazier & Clifton, 1996) in which some relations are not resolved immediately and thus are more malleable, or to Prediction by Production (Pickering & Gambi, 2018), which suggests that predictions can arrive at different levels of completeness as we move through the different stages of production. This is in line with the current hypothesis that the parser may utilize only partial representations or generate these representations relatively slowly. Overall, such a hypothesis would suggest that prediction (and thus reanalysis) occurs in similar rates for syntactically- and pragmatically-motivated predictions, but the malleable state of the pragmatic prediction makes its revision easier.

We observed evidence in favor of the latter approach suggesting a qualitative difference in the prediction and reanalysis process, over the first (quantitative difference in prediction and reanalysis rate). In Experiment 4, we observed differential patterns of semantic persistence. Semantic incongruency of the two meanings was costly only for syntactically-motivated dependencies. This cannot be explained by reanalysis rate, and thus favors the interpretation of commitment level as the stage of predictive processing accomplished by the reanalysis site.

It should also be mentioned that the approach of a qualitative difference in the prediction and reanalysis process could also be accommodated within the model of Self-Organizing Sentence Processing. In this model, the process of building structure involves gradual bonding of the elements. Expectations of the sort discussed here can be made only when the structure is relatively stable. One could assume that syntactic constraints drive faster bond formation and/or stabilization. Thus, the syntactic expectation for a gap could emerge earlier relative to the pragmatic prediction of a pronoun (producing the earlier reanalysis). In addition, this should

mean that the syntactic expectation would be more difficult to undo, resulting in higher costs of reanalysis and/or of semantic inhibition. Under a dynamical system like that of Self-Organizing Sentence Processing, qualitative contrasts can arise due to a continuous parameter change, in this case the attachment's stability, even though the underlying processes are the same ("phase transitions").

3.6.3 Implications for the interpretation of the Active Filler strategy

Most studies of the Active Filler strategy consider it (implicitly or explicitly) a structurally-motivated principle. Some assume that this parsing strategy reflects a primary structural preference, which does not reduce to anything else (e.g. Frazier & Clifton's 1989 original formulation); others propose that it derives from a principle of minimizing structure (De Vincenzi, 1991); yet others argue that it reflects the urgent need of the parser to locate the canonical position of the filler and assign it a thematic role (Pritchett, 1992; Aoshima, et al., 2004), or results from a need to confirm that the dependency is well-formed (i.e. does not cross an island boundary, Wagers & Phillips, 2009).

However, non-syntactic accounts of the Active Filler strategy are also found in the literature. Altmann (1999) suggested that the strategy results from the need to interpret (rather than to license) the sentence rapidly. Under this view, active gap-filling is a manifestation of general anticipatory processes. Lastly, another possible account for the Active Filler strategy can be formulated based on information structure premises. The pragmatic discourse function of filler-gap dependencies requires the following clause to be about the filler (Kuno, 1976). Thus, the preference to resolve the dependency in a discourse-prominent slot may lead the parser to predictively posit a gap at early matrix positions, which are also more prominent.

In previous studies, the structures used as baseline for testing the Active Filler strategy in filler-gap dependencies did not permit building a dependency at all. A better baseline should eliminate only the syntactic motivation for creating a dependency, while still leaving semantic/pragmatic motivations intact. Therefore, previous evidence for active dependency formation could not tease apart the contribution of these different processes.[14]

The results obtained in the current set of experiments provide evidence that the Active Filler strategy may be partly motivated by discourse considerations, but is not fully reducible to them. In line with the syntactic licensing perspective, I observed evidence for active dependency formation following filler-antecedents, above and beyond the effects observed for non-filler-antecedents (in *regarding* and *with*-dependencies). However, I also detected active dependency formation with non-filler-antecedents, when they presented discourse prominent

[14] Sussman and Sedivy (2003) targeted the question of contextual prediction in a visual-world eye tracking study. They monitored eye movements of participants scanning several object illustrations related to short background stories. They found predictive fixations on the picture corresponding to the direct object, at the offset of the verb in wh-questions (*What did Jody squash the spider with?*) relatively to yes/no questions (*Did Jody squash the spider with her shoe?*). They suggest that this reflects active gap filing above and beyond contextual prediction, since participants are able to deduce the complement of the verb (from the background story) even in the yes/no baseline. However, fixations on this object in the wh-condition may also reflect utterance planning processes (Griffin & Bock, 2000) since in the wh-condition the object had to be named. Thus, these results are not conclusive by themselves.

topics. This suggests that pragmatic motivations and discourse prominence partly contribute to strategies of predictive dependency formation.

Teasing apart these different prediction motivations also has some methodological implications. From a methodological point of view, the current findings suggest that early filled-gap effects are better indicators of syntactically-motivated dependency formation, while spillover effects may be more effected by discourse prominence motivations for prediction.

3.6.4 Directions for future research

The effects observed in this study set the stage for several additional avenues of investigation. A possible line of future work could test whether the prediction of *regarding*-dependencies aligns with island constraints. Kuno (1976) suggests that the discursive status of both aboutness topics and fillers give rise to similar preference for matrix co-reference, and avoidance of resolution in backgrounded constituents. Thus, over the years, several islands were suggested to be an instance of such pragmatic violations, exhibiting extraction from presupposed or other non-focused domains. These include the coordinate structure constraint (Kuno, 1976; Deane, 1991), relative clause islands and other Complex NPs (Kuno, 1976; Van Valin, 1996; Goldberg, 2006), subject islands (Erteschik-Shir 1973; Goldberg, 2006), and manner-of-speaking verbs (Goldberg, 2006). It is well-established that during active formation of filler-gap dependencies, the parser avoids postulating a gap site within island structures. If *regarding*-dependencies are subject to similar information structure constraints and initiate predictive processing like filler-gap dependencies, the following question arises: Are predictions of *regarding*-dependencies similarly sensitive to island constraints in incremental processing? Or is the parser willing to predict pronoun positions in presupposed islands?

In addition, the current set of experiments treats filler-gap dependencies as mostly syntactic ones. However, different sub-types of such dependencies may involve different levels of discourse prominence and different information structure considerations. For example, topicalization may induce higher pragmatic pressures relative to wh-questions, used in the current study. In addition, Abeillé, Hemforth, Winckel, and Gibson (2020) recently suggested an additional complication to the information structure constraints on filler-gap dependencies. They claim that in contrast to wh-elements, which are given focus position and seek for new information, relative clauses add a property to a given entity, and thus are less restrictive in their preference for non-backgrounded information. Future research could test if syntactic and pragmatic influences can conspire to produce different levels of predictive processing even within the family of filler-gap dependencies.

Similarly, such a research program could also benefit from more fine-grained frequency and pragmatic prominence measures for non-filler-gap dependencies. There are many different "aboutness" antecedents within and across languages (e.g. in English: *as for*, *speaking of*, *about*). As mentioned in footnote 8, some antecedents like English *about* may also pose syntactic constraints. Therefore, some antecedent types may present higher/lower requirement for subsequent coreference, or different sets of restrictions on the focus position of their co-referential pronoun. Thus, it could be interesting to test whether these generate different prediction and reanalysis patterns. While I argued that the reanalysis patterns in this study are

not reducible to simple frequency measures, additional fine-grained differences could provide a fuller picture of the interaction between the frequency of the dependency, its linguistic source and the prediction strength.

3.6.5 Conclusions

To conclude the first study, I propose that the parser establishes predictions with different levels of commitment. My findings exhibit three different degrees of prediction: Contextually available antecedents may incite a preference for co-reference, yet the prediction associated with them is weak, and does not involve costly reanalysis; Pragmatic motivations (e.g. information structure considerations) can provoke predictive dependency formation whose disconfirmation is costly; Lastly, syntactic licensing pressures enhance the predictive process such that dependency formation is observable in earlier measures, with even higher disconfirmation costs, and fuller semantic consequences.

In the overall context of this dissertation, these findings indicate a way in which the parser can modulate the potential costs of prediction, and the severity of prediction errors. Since our interpretation of the language input has to flexibly adapt to evidence from different sources, these different degrees of prediction may provide an important mechanism in our communicative ability. Readers and listeners may utilize this mechanism to balance the benefits and costs of eager parsing. Given that some initial misinterpretations persist even after reanalysis, modulating our commitments, and thus controlling the difficulty of dependency detachment, can facilitate accurate interpretation.

Chapter 4

Memory fallibility in subject-verb-reflexive dependencies

4.1 Introduction

Linguistic dependencies are subject to many possible sources of noise due to multiple levels of information, complex processing strategies and our generally limited cognitive capacity. It is fascinating, not that errors in fact occur, but that they seem to be (at least partially) systematic glitches of the mechanism rather than random background noise. Indeed, every step of normal processing might lead the parser into a path of misanalysis.

Do such memory faults last and cascade? Does the parser have some safety net to allow later recovery? I look into this question focusing on agreement relations, dependencies which are notorious for their systematically error-prone processing. I ask whether errors in one agreement dependency (between the subject and the verb) may propagate to another dependency (between the subject and a reflexive pronoun).

I utilize unique characteristics of Hebrew agreement in this investigation. In the rich agreement system of Hebrew, gender agreement is mandatory in almost every sentence – barging into the association between nouns and their corresponding adjectives, pronouns, and verbs. As the Israeli poet Yona Wallach described it "Hebrew peeks at you through the keyhole \\ the language sees you naked". Since both reflexives and verbs manifest the gender of the subject noun, Hebrew presents unique grounds for testing the interaction between verbal and anaphoric agreement.

4.1.1 Agreement attraction

Grammatical agreement links various constituents in the sentence through morphosyntactic marking of number, gender and person. Research has consistently observed an intriguing pattern of errors in this process, exhibited similarly in comprehension and production. Specifically, a structurally irrelevant antecedent may prompt speakers and readers to establish agreement with it, instead of with the target noun.

For example, in (31), taken from Dillon and colleagues' (2013) paper, the underlined verb should agree with *executive* but instead gets a plural marking matching the distractor *managers*. Such errors are commonly observed in production studies and missed by readers in

comprehension experiments (i.e. induce an illusion of grammaticality). Agreement attraction has been robustly observed in subject-verb dependencies, and to some extent also in various pronoun types (reflexives - Bock, Nicol, & Cutting, 1999; Jäger, Mertzer, Van Dyke, & Vasishth, 2020; tag pronouns - Bock, Eberhard, & Cutting, 2004; object clitics - Paspali & Marinis, 2020; demonstrative and relative pronouns - Meyer & Bock, 1999; ellipsis - Martin, Nieuwland, & Carreiras, 2012; 2014). Attraction in reflexive pronouns (32) is perhaps even more surprising than verbal agreement attraction errors. The subject, with which the reflexive has to agree, should have been reactivated just before arriving at the reflexive, namely at the verb, for thematic integration and possibly for verification of subject-verb agreement. This could in principle make the association of the subject with the reflexive pronoun quite transparent. Yet, verification of reflexive-subject matching can be susceptible to interference (Parker & Phillips, 2017; Sloggett, 2017; Jäger, Mertzer, Van Dyke, & Vasishth, 2020).

(31) The new executive who oversaw the middle managers apparently were dishonest about the company's profits

(32) The new executive who oversaw the middle managers apparently doubted themselves on most major decisions

However, most agreement attraction investigations have targeted reflexive agreement when the verb does not provide any relevant agreement cues. This is achieved either by manipulating reflexive number and using verbs which do not carry number information (e.g. past tense verbs in 32, taken from Dillon et al., 2013), or, more commonly, by testing the reflexive's gender agreement[15] (Sturt, 2003; Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Parker & Phillips, 2017; Sloggett, 2017), as most languages do not manifest gender in verbal agreement. This creates a gap in the literature as to how accessing the subject at the verb, with or without targeting its agreement, might affect the processing of the subsequent agreement relations at the reflexive pronoun.

4.1.2 Gender agreement attraction on predicates

While number agreement is relatively well-studied, considerably less work has been conducted addressing gender agreement, perhaps because English, the most accessible and well-represented language in the literature, lacks grammatical gender specifications. Even in languages with a grammatical gender system, verbal agreement rarely carries gender features.[16] The research available so far on gender agreement presents patterns which are somewhat different from those of the number agreement and suggest that gender attraction is (at least partially) independent from number attraction. Therefore, before turning to the current study we should first review the findings regarding verbal gender agreement.

[15] Using gender agreement attraction in the case of pronouns has the advantage of eliminating the possibility of a collective reading where the pronoun refers to both the target (the *executive*) and the distractor (the *managers*).

[16] Grammatical gender systems appear in many languages, including in Romance (French, Italian, Spanish), Germanic (German, Dutch, Norwegian), and Slavic languages (Russian, Polish, Czech). However, the presence of grammatical gender marking on nouns does not necessarily require agreement in subject-verb dependencies. In Romance languages, for example, only adjectives, and not verbs, manifest gender agreement. In Slavic languages gender agreement is established with verbs, but only in their past tense form.

Gender agreement was first studied in the production of predicative adjectives in Romance languages like Italian, French and Spanish (Vigliocco & Franck 1999; Vigliocco & Franck, 2001; Antón-Méndez, Nicol, & Garrett, 2002; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008). In these studies, participants produced predicative adjectives with erroneous gender marking, when a local distractor noun was embedded within the subject phrase (see example in 33, from the experiments on Italian in Vigliocco & Franck, 1999). Attraction was more prominent when the gender of the target noun was determined grammatically (33b) rather than conceptually (33a). Later, gender attraction was also observed in production of verbs in Hebrew (Deutsch & Dank 2009, 2011; Dank & Deutsch, 2010), and in Slavic languages (Slovak - Badecker & Kuminiak, 2007; Russian - Malko & Slioussar, 2016).

(33) a. *L'inquilino della casa*
The tenant-M of the house-F
b. *Lo sgabuzzino della casa*
The closet-M of the house-F

Although most studies on gender attraction have focused on production, there are a few relatively new studies of gender attraction in comprehension. Dank, Deutsch, and Bock (2015) exhibited gender attraction in an eyetracking study of Hebrew. They found that the conceptual vs. grammatical source of the target noun's gender affects the robustness of the attraction in comprehension too (similarly to production results of Vigliocco & Franck, 1999). Additional evidence for gender attraction in comprehension was found in Russian verbs (Slioussar & Malko, 2016), and recently in Greek predicative adjectives (Paspali & Marinis, 2020). The most comprehensive study of gender attraction in comprehension was conducted in a language closely related to Hebrew, Modern Standard Arabic (Tucker, Idrissi, & Almeida, 2016).

Overall, studies on both production and comprehension paint a picture of independent paths for number and gender agreement. First, gender errors can occur without number errors and vice versa when elicited by the same distractor (i.e. one noun, distinct from the target subject in both number and gender). In a production study, Antón-Méndez and colleagues (2002) observed predicative adjectives which matched the distractor in one agreement aspect but retained the correct feature of the target noun in the other (see 34 for an example). They suggest that the parser computes gender and number agreement separately.

(34) *la vista de los puertos*
the view.F.SG of the beaches.M.PL
{ a. *son bonitas*
are pretty.F.PL
b. *es bonito*
is pretty.M.SG

In comprehension, Tucker and colleagues (2016) found several prominent differences between gender attraction and number attraction. First, they observed evidence for different effect sizes on number and gender attraction, suggesting that gender features yield attraction effects of a larger magnitude. Thus, Tucker and colleagues suggest that a gender mismatch is more susceptible to grammaticality illusions. This could be in line with results from production of Spanish predicative adjectives (Antón-Méndez et al., 2002), where gender errors were more frequent than number errors. In addition, Tucker and colleagues (2016) exhibited that gender

attraction may affect reading times on a slightly different time scale. Specifically, in their self-paced reading data, number attraction emerged systematically earlier than gender effects (at the verb vs. one word after it; for a detailed distributional analysis see Almeida & Tucker, 2017).

Lastly, there is contradicting evidence as to the markedness asymmetry in the context of gender attraction. A hallmark of number attraction is an asymmetric pattern where a distractor carrying the marked (plural) feature induces more attraction than a distractor of the unmarked (singular) form. In gender attraction, a similar pattern could be expected, by which feminine distractors would interfere with subject-verb agreement to a higher degree relative to masculine distractors, which usually are the unmarked or default from of gender agreement. However, it is unclear whether such a pattern is observed. Tucker and colleagues (2016) suggest some evidence for it in their investigation of Arabic comprehension. Yet, many others have failed to exhibit this in other languages, in either production or comprehension studies, including Deutsch and Dank (2011) in their study of Hebrew verbal agreement.[17]

4.1.3 Agreement attraction in reflexive pronouns

Agreement attraction has also been investigated (separately) in reflexive pronouns. In the relatively recent literature targeting this topic, we can find very different approaches to such effects. Specifically, researchers debate the extent to which such production and processing errors should occur in anaphoric dependencies such as those involving reflexive pronouns.

One the one hand, one may claim that reflexives would exhibit more attraction effects than verbs, since pronouns in general exhibit more permissive agreement patterns. As the well-known Agreement Hierarchy (Corbett, 1979; Wechsler & Zlatic 2000) indicates, pronouns are more likely than verbs to manifest agreement with the notional number or gender features of their antecedents. This might suggest that verbal agreement is more syntactically restricted, while pronouns (as a general category) show flexibility in their agreement relations. This could in principle incite incorporation of more agreement sources and thus lead to higher rates of agreement attraction overall in pronouns, including in reflexive ones. In line with this hypothesis, Bock, Nicol, and Cutting (1999) exhibited higher rates of agreement attraction in pronouns compared to verbs. Bock and colleagues compared production of verbs (35a), reflexive pronouns (35b) and tag pronouns (35c) in an agreement attraction design. Among other findings, they found that in the classic attraction environments as in (35), pronouns were even more susceptible than verbs to plural distractors. This was true for both pronoun types.

(35) a. The actor in the soap operas were/was popular
b. The actor in the soap operas watched himself/themselves
c. The actor in the soap operas rehearsed, didn't he/they?

[17] No reliable effects for the markedness of the distractor were detected in production and corpus studies of Romance gender agreement (Vigliocco & Franck, 1999; Igoa, García-Albea, & Sánchez-Casas, 1999). Similarly, in Hebrew markedness has been shown to affect production of number attraction errors but not gender (Deutsch & Dank, 2011). Markedness of the distractor seems to influence production in Slavic languages with a tripartite system, i.e. Slovak (Badecker & Kuminiak, 2007) and Russian (Slioussar & Malko, 2016). Yet, even in Russian the markedness asymmetry failed to appear in comprehension (Slioussar & Malko, 2016).

A different perspective, however, would suggest that agreement attraction should exhibit the same patterns in reflexives and in verbs. This approach leans on the classic models of agreement attraction. On the Marking and Morphing tradition, different agreement patterns in verbal and pronominal agreement (i.e. the Agreement Hierarchy) reflect differential contribution of semantic factors at the Marking stage. However, it is only in the Morphing stage that attraction arises. Thus, attraction on reflexive pronouns and verbs should be comparable. Bock, Eberhard, and Cutting (2004) exhibited the dissociation between semantic Marking and attraction in the production of verbs and tag pronouns. Across five experiments, they failed to detect an interaction between agreement type (verbal or pronominal) and the susceptibility to the distractor's number.

Similarly, on the Cue-Based Retrieval tradition, subject-verb dependencies and reflexive-antecedent ones should exhibit similar attraction rates since they both utilize the same morphological and syntactic features as cues. In line with this, Jäger, Mertzer, Van Dyke, and Vasishth (2020) report a large-scale eyetracking study, where comparable attraction effects are detected in total reading times for verbal and reflexive dependencies.

Lastly, it was also suggested that reflexive pronouns should be more restrictive than verbs in their agreement patterns (Dillon et al., 2013). As Dillon and colleagues point out, the grammatical function of agreement is different in reflexives and verbs. In reflexive pronouns but not in subject-verb dependencies, agreement reflects co-reference, overseen by binding principles. Thus, morphological features might not be directly invoked in the case of reflexives. If these features are not used as retrieval cues (or are given less weight in the retrieval process), comprehension of reflexive pronouns should rely mostly on structural cues and therefore would be less vulnerable to attraction effects. Indeed, several studies failed to find an effect for a structurally irrelevant antecedent on the processing of a reflexive pronoun (Sturt, 2003; Cunning & Strut, 2014), even when closely comparing verbal agreement and reflexive binding (Dillon et al., 2013).

In later research, it was suggested that reflexive pronouns might be less susceptible to attraction, but not entirely resilient to it. Parker and Phillips (2017) argued that an attraction effect can be detected for reflexives with a more severe mismatch between the reflexive and the target, and Sloggett (2017) claimed that attraction can arise depending on the availability of different nouns in the sentence as antecedents for logophoricity (expressed using the same self-pronouns as reflexive binding). Yet, some did find agreement attraction effects in the comprehension of reflexive pronouns without manipulating the extent of the feature mismatch or logophoric prominence (Cunning & Felser, 2013; Patil, Vasishth, & Lewis, 2016; Jäger et al., 2020). Overall, I take the various findings to suggest that agreement attraction arises and is observable in the processing of reflexive pronouns.

4.1.4 The Hebrew agreement system

Hebrew exhibits an overt dual gender system with masculine and feminine marking on all nouns. Most animate nouns have a feminine and a masculine version (like *actor* and *actress* in English). These nouns follow relatively regular morphological marking where the feminine form is derived by an addition of a suffix (either a, or et/it) on the (default/unmarked) masculine

form. For example, the masculine form of a student can be *talmid* or *student* (for higher education) and their feminine counterparts are *talmida* and *studentit*. Inanimate nouns (and on rare occasions some nouns referring to animate entities) have grammatical gender associated with them. In those cases, where grammatical gender does not rely on natural gender, gender-suffix correspondence is slightly less transparent (with various exceptions in the singular or plural forms).

Gender agreement has to be established with various constituents in Hebrew: attributive and predicative adjectives, personal, reflexive, and resumptive pronouns, numerals, and verbs. Crucially, Hebrew (like other Semitic languages, e.g. Arabic and Amharic) manifests gender agreement with most verb forms. Verbal gender agreement is regular yet much more intricate in its conjugation, relative to the manifestation of gender in alternating nouns (Table 4.1). Outside attraction configurations, violation of gender agreement in Hebrew verbs gives rise to increased reading times, depending on explicit marking of the erroneous feature on the verbal form (Deutsch & Bentin, 2001), and the distance between the subject and the verb (Deutsch, 1998).

	Past	Present	Future
1st person, singular	same form (katavti)		same form (extov)
2nd person, singular	katavt**a** vs. katavt	kotev vs. kotev**et**	tixtov vs. tixtev**i**
3nd person, singular	katav vs. katv**a**		**y**ixtov vs. **t**ixtov
1st person, plural	same form (katavnu)		same form (nixtov)
2nd person, plural	katavte**m** vs. katavte**n**	kotv**im** vs. kotv**ot**	tixtev**u** vs. tixtov**na***
3nd person, plural	same form (katvu)		**y**ixtev**u** vs. **t**ixtov**na***

Table 4.1. Example of the contrast between masculine and feminine forms in Hebrew: Conjugation of the verb *katav* ('wrote'). Within each pair, the masculine from appears on the left, and the feminine on the right.

Note: Traditionally 2nd and 3nd person plurals have a distinct feminine form in the future tense. Yet these feminine forms are rarely used by Modern Hebrew speakers in either speech or writing. The unmarked masculine form is mostly accepted for both genders, but the feminine form can be found in some legal or formal settings.

The reflexive form also inflects in gender, number and person (see Table 4.2). This might seem trivial in comparison to English reflexives. Yet reflexive pronouns in many languages are in fact deficient in terms of agreement marking. Reuland (2018, p. 82) notes, in a review of reflexivity, that reflexives often "lack a specification for gender and number (e.g., Dutch *zich*, Norwegian *seg*); in some languages, they also lack a specification for person (e.g., Russian *sebja*, reflexive clitics in other Slavic languages)". Thus, Hebrew allows a unique environment which enables us to test the interaction between multiple agreement dependencies in a sentence.

	Masculine	Feminine
1st person, singular	acmi	
2nd person, singular	acmex**a**	acmex
3nd person, singular	acm**o**	acm**a**
1st person, plural	acmenu	
2nd person, plural	acmexe**m**	acmexe**n**

3nd person, plural	acma**m**	acma**n**

Table 4.2. The contrast between masculine and feminine reflexive pronouns in in Hebrew.

4.1.5 What happens when two constituents need to agree with the same noun?

Overall, as described above, research has revealed evidence for gender attraction effects in predicative agreement and in anaphora. Interestingly, both constituent types, verbs and reflexive pronouns, need to agree with the same target (the head of the subject phrase). Are these processes independent? Does the early computation and/or retrieval of the subject's agreement at the verb result in a long-lasting effect on agreement representations? I consider several possible answers for this question.

First, the verb could pull the subsequent retrieval (at the reflexive) towards the target item or the distractor, by means of reactivation. The logic of this hypothesis should be as follows: The availability of agreement features on the verb should decrease (in the case of grammatical verbs) or increase (with ungrammatical verbs) the likelihood of erroneously retrieving the distractor. Retrieving an item from memory should facilitate access to it further downstream due to its heightened activation levels (see Vasishth & Lewis, 2006, for such a claim in the context of anti-locality effects). Namely, increasing or decreasing the activation of distractors at the verb may affect their retrieval likelihood at the reflexive pronoun. Put more simply, it could be that attraction in an early part of the sentence (the verb) increases the chances of another attraction later on (at the reflexive pronoun). This seems to be most in line with the Cue-Based Retrieval model, yet it seems to be the least efficient strategy in terms of avoiding errors in agreement processing.

On the other hand, it is possible that the parser directly retrieves the subject both at the verb and at the pronoun, utilizing the same agreement representation each time. This should allow agreement attraction at the reflexive pronoun to occur independently of verbal agreement properties. This strategy could be beneficial since it avoids lingering misrepresentations of agreement relations. Namely, if attraction occurs at the verb, it would be better for it not to propagate and incite additional errors. Moreover, such a mechanism could be in line with the fact that pronouns and verbs do not depend on exactly the same agreement features. According to the agreement hierarchy, pronouns tend to reflect notional features more freely than verbs. Thus, we can observe distinct features on verbs and pronouns which supposedly should agree with the same noun. Bock, Nicol, and Cutting (1999) cite the naturally occurring example in (36), where the verb matches the grammatical (singular) number of the subject, while its notional plurality is reflected in the form of the possessive pronoun. Computing agreement independently at each site could facilitate comprehension in such sentences.

(36) Tonight, on MTV, Bill Clinton faces the generation that <u>holds</u> the future in <u>their</u> hands

Lastly, verbal agreement could also directly affect the agreement associated with the reflexive or the subject. We could hypothesize that it is most efficient for the parser to consult agreement information at additional (intermediate) points of the sentence, rather than at the target only. This strategy could help avoiding an error-prone long-distance agreement. This, in my view, could be done in one of two ways. One option would be updating the representation of the subject while revisiting it for subject-verb agreement. At the verb site, the parser can use

verbal agreement to modify or stabilize the representation of the subject. Thus, the agreement dependency between the subject and the verb can support the feature representation of the subject (which has to be accessed later for the anaphoric relation). Namely, accessing the representation of the subject's agreement could fix its value, or add weight pulling the subject's feature representation in some direction or other. This would facilitate subsequent retrievals of the subject within that clause, and protect such retrieval processes from attraction.

Alternatively, the parser may aim at establishing local agreement between the verb and the reflexive. Checking the features at the verb, instead of retrieving the subject directly, is another way in which the parser might utilize the additional agreement information of the verb. Since the verb should mirror the features of the subject, the verb site functions as a good "shortcut" for agreement processing. Verifying verb-reflexive agreement could allow the parser to altogether avoid the possibly noisy retrieval of the subject. Thus, when the verb carries agreement information, a reflexive would be considered acceptable (as least in initial processing) when it matches the verb, whether or not the verb actually matches the subject in that sentence.

Both these options would be in line with observations that agreement on different parts of speech often go hand in hand. Antón-Méndez, Nicol, and Garrett (2002) suggested that computing number agreement for an auxiliary verb and a following predicative adjective is a unitary process. When testing attraction in production of Spanish, they found that preambles like the one in (37) could provoke erroneous completions where both the verb and the subsequent adjective agreed with the distractor (37a), but not completions where one constituent agrees with the target noun and the other does not (37b-c)

(37) *la vista de los playas*
the view of the beaches

a. *son bonitas*	b. *es bonitas*	c. *son bonita*
are pretty.PL	is pretty.PL	are pretty.SG

In addition, Molinaro, Kim, Vespignani, and Job (2008) probed the processing of reflexive pronouns which followed a subject-verb mismatch (as in 38). The results of their ERP study suggested a P600 effect at the reflexive pronoun mismatching the verb (and agreeing with the subject), as canonically occurs for ungrammaticality (relatively to a fully grammatical control). However, no effect was detected when the pronoun matched the verb (i.e. did not agree with the subject). This suggests that verbal agreement is used to determine the acceptable agreement of a subsequent reflexive or to modify the agreement representation of the subject.

(38) The famous dancer were nervously preparing herself/themselves to face the crowd

To sum up, we could hypothesize that (i) verbal agreement indirectly affects the accessibility of the distractor; (ii) verbal agreement is independent of subject-reflexive agreement (does not affect it at all); or (iii) verbal agreement directly modifies the representation of the subject (as a "stabilizer") and/or alters retrieval at the reflexive (as a "shortcut" to the subject). The first option should result in cascading attraction effects; the second option reduces the possibility of double attraction; and the last option could eliminate the need for long-distance retrievals and may employ representations that are less "noisy".

4.1.6 The current study

The current study tests the vulnerability of reflexive pronouns to gender attraction in three possible environments: Following verbs that do not carry agreement information (as attested in previous studies of reflexive attraction), when the verb mirrors the correct gender features of the subject noun, and when the verb manifests agreement cues which mismatch the features on the subject (i.e. following ungrammatical verbs).

Such an experimental design can contrast the three hypotheses laid out above. If verbal agreement indirectly influences distractor accessibility via activations, we should observe more prominent agreement attraction following ungrammatical verbs. This would follow from the classic fluctuating activation concept of Cue-Based Retrieval. The Cue-Based Retrieval model suggests that repeated retrievals facilitate access to the reactivated items. Therefore, agreement attraction at the verb should increase the likelihood of additional mis-retrievals, that is the rate of reflexive attraction. We know that agreement attraction is more likely with ungrammatical verbs (relative to grammatical verbs, matching the target subject) in number attraction (see Wagers et al., 2009 and Hammerly, Staub, & Dillon, 2019, for two different interpretations of this) and in gender attraction (Tucker et al., 2016). Thus, if the rate of attraction at the verb should increase the rate of attraction at the reflexive, we should expect more prominent effects when the reflexive follows an ungrammatical verb.

Alternatively, if agreement computations throughout the sentence are independent, we should observe similar rates and magnitudes of reflexive attraction in the three verbal environments. Namely, on the second hypothesis we should not expect an interaction between verbal agreement and reflexive attraction effects.

Finally, the last hypothesis suggests that verbal agreement features are used on or instead of subsequent retrievals of the subject. For the purpose of the current experiments, I group together using verbal agreement as a "shortcut" to the subject's agreement or as a high weight "stabilizer" of the subject's representation. These hypotheses suggest that reflexive attraction should be more prominent when the verb does not bear agreement cues. Let us consider why. On the "shortcut" interpretation, when agreement cues are available at the verb, the subject itself is not retrieved. This eliminates the chance of mis-retrieving it and reduces susceptibility to attraction. This strategy of establishing local agreement should naturally apply regardless of the verb's actual features, and thus no attraction is expected whether the verb is grammatical or not. On the "stabilizer" interpretation, we do re-access the subject, but its representation has been tweaked to match the features of the verb. With grammatical verbs the additional confirmation of the subject's agreement makes the distractor less prominent. On the other hand, with ungrammatical verbs the agreement could be treated as a cue to modify the representation of the subject, if verbal agreement is given enough weight in stabilizing an agreement feature representation. Thus, in both cases, the availability of verbal agreement should reduce the attraction effects at the reflexive pronouns. Thus, we should observe more prominent attraction at the reflexive, when the preceding verb does not carry agreement information.

I test these contrasting hypotheses using two different methods and two different syntactic structures in Experiments 5 and 7. It should be noted that testing processing downstream from an ungrammaticality might be prone to floor effects. Participants may strategically move to a

shallow processing strategy once they realize something has gone awry with the form of the sentence. Moreover, the ungrammaticality arises from gender agreement, exactly the same issue we are testing at the reflexive pronoun. This could be problematic. Even if readers try to construct the sentence interpretation, it could be that they take the early mismatch as a cue to neglect further agreement information in that sentence. This again would confound our findings. Experiment 6 tackles this issue.

4.2 Experiment 5: Verbal agreement and reflexive attraction in comprehension

4.2.1 Methods

Participants

Participants were 78 native speakers of Hebrew (according to self-report) (mean age: 24.85, range: 19-35). Four participants were bilingual of Hebrew and either Russian or English, and the rest were monolingual.

Materials

The experiment included 30 item sets of six conditions, manipulating subject-verb agreement and the agreement of a distractor noun (see example set in Table 4.3). All reflexive pronouns mismatched the subject in gender, and were therefore ungrammatical. In the grammatical verb conditions, thus, the verb matched the subject, but mismatched the reflexive pronoun in gender. The ungrammatical verb conditions produced the opposite situation – the verb mismatched the subject in gender, but matched the (ungrammatical) reflexive pronoun. For constructing an environment where no agreement features are available at the verb (no verbal agreement conditions), I used the Hebrew possessive structure, where the predicate *yeš* does not bear agreement with any noun in the sentence. I assume the possessor in these structures functions as the subject of the clause following Shlonsky (1987).[18]

Since the findings regarding the strength of agreement attraction are mixed, I wanted to allow the best conditions for attraction to arise, as a first approximation of the effect that verbal agreement might have on it. I thus did not use nouns which are morphologically or orthographically ambiguous/misleading, neither as the subject nouns nor as the distractors (see Dank & Deutsch, 2010, for effects of surface form on agreement attraction errors). All reflexive pronouns were of the feminine form, and all the subject head nouns were of the masculine (unmarked) form. This was done to accommodate the possibility that a markedness asymmetry affects gender attraction too.

In addition, all reflexive pronouns in this experiment were embedded within a picture NP (e.g. *paintings of herself*, *recordings of herself*, *stories about herself* etc.). The self-forms in such NPs are considered exempt anaphors, part of a well-known class of exceptions to the binding conditions, which may allow under some circumstances an antecedent outside the local subject position for the reflexive (consider for example the English *John heard that Mary likes this picture of himself*, in contrast to *John heard that Mary likes himself*). I used picture NP reflexives based on a previous experiment which exhibited more prominent reflexive attraction in Hebrew in such environments (Keshev, Bassel, & Meltzer-Asscher, 2018). Slogget (2017) suggested attraction in reflexive pronouns is affected by logophoricity, namely the usage of *self*-pronouns as referring to the perspective holder, even if this NP is not in the local subject

[18] Shlonsky (1987) suggests that the *yeš* in Hebrew is ambiguous between an existential 'be' interpretation (including the locative realizations of *yeš*), which takes a single argument, and a possessive 'have' interpretation which takes an accusative complement (the possessed argument) and a dative argument (the possessor), which in such sentences is the subject of the clause.

position (showing exempt anaphora). Hebrew does not license logophoric pronouns and exempt anaphors as freely as English and other languages (Bassel, 2016), but it does license exempt anaphors in picture NPs. I therefore used such NPs, since it is possible that attraction would be more likely to arise there, on a par with English logophors. However, note that the distractor in my materials is not a perspective holder and could not be interpreted as a grammatical antecedent for the reflexive pronoun.

Condition	Sentence
No verbal agreement, +distractor	*la-talmid šel ha-mexanexet ha-xadaša yeš karikaturot šel* to-the-student.M of the-teacher.F the-new.F be caricatures of ***acma** ba-maxberet ha-išit* **herself** in.the-notebook the-personal 'The student.M of the new teacher.F has caricatures of herself in the personal notebook'
No verbal agreement, -distractor	*la-talmid šel ha-mexanex ha-xadaš yeš karikaturot šel* to-the-student.M of the-teacher.M the-new.M be caricatures of ***acma** ba-maxberet ha-išit* **herself** in.the-notebook the-personal 'The student.M of the new teacher.M has caricatures of herself in the personal notebook'
Grammatical verb, +distractor	*ha-talmid šel ha-mexanexet ha-xadaša ciyer karikaturot šel* the-student.M of the-teacher.F the-new.F drew.M caricatures of ***acma** ba-maxberet ha-išit* **herself** in.the-notebook the-personal 'The student.M of the new teacher.F drew.M caricatures of herself in the personal notebook'
Grammatical verb, -distractor	*ha-talmid šel ha-mexanex ha-xadaš ciyer karikaturot šel* the-student.M of the-teacher.M the-new.M drew.M caricatures of ***acma** ba-maxberet ha-išit* **herself** in.the-notebook the-personal 'The student.M of the new teacher.M drew.M caricatures of herself in the personal notebook'
Ungrammatical verb, +distractor	*ha-talmid šel ha-mexanexet ha-xadaša ciyra karikaturot šel* the-student.M of the-teacher.F the-new.F drew.F caricatures of ***acma** ba-maxberet ha-išit* **herself** in.the-notebook the-personal

	'The student.M of the new teacher.F drew.F caricatures of herself in the personal notebook'
Ungrammatical verb, no distractor	*ha-talmid šel ha-mexanex ha-xadaš ciyra karikaturot šel* the-student.M of the-teacher.M the-new.M drew.F caricatures of ***acma ba-maxberet ha-išit*** **herself** in.the-notebook the-personal 'The student.M of the new teacher.M drew.F caricatures of herself in the personal notebook'

Table 4.3. Example set from Experiment 5.
The reflexive is marked in bold; F = feminine grammatical gender; M = masculine grammatical gender.

The experimental items were distributed in a Latin square design across six lists. Within each list, the target sentences were combined with 60 grammatical filler sentences, for a total of 90 sentences. The filler items included 18 sentences with grammatical reflexive pronouns (within picture NPs or as direct objects), and 10 additional items with a grammatical non-reflexive pronoun in a possessive position, similar to the one the reflexive took in the experimental materials. Filler items utilized various syntactic structures, including possessive constructions similar to the ones used in the experimental conditions lacking verbal agreement (20 items).

Procedure

The experiment was a lab-based self-paced reading experiment. Fifty percent of the trials were followed by a 'yes/no' comprehension question.

Data analysis

Participants were excluded from analysis if their performance on comprehension questions of experimental items was lower than 70% (resulting in the removal of two participants), or if their average RT in experimental items was more than 2.5 SDs above the group's average (resulting in the removal of three additional participants). For the remaining 72 participants, RTs higher than 2.5 SDs above the individual's average RT were trimmed to that cutoff (affecting 0.08% of the data). RTs shorter than 120ms were excluded (affecting another 0.08% of the data).

I analyzed reading times at the critical region and at the spillover region (the first word after the reflexive, which began a prepositional phrase), in separate models. I used a treatment (dummy) coding scheme for the verbal agreement manipulation (to compare reading times in the grammatical and ungrammatical conditions against the no verbal agreement condition), and sum coding for the distractor manipulation (to collapse over both +/-distractor conditions). In addition, to detect attraction effects in each case, three pairwise comparisons were carried out (for every region), contrasting the +/- distractor levels in each case. Therefore, on the frequentist analysis, pairwise comparisons were corrected by Bonferroni correction to three comparisons. Due to convergence failure on the frequentist analysis, slopes were removed from the random effect structure. Thus, the final model for the spillover region included only one slope (corresponding to the main effect of grammatical verb conditions) on by-subject effects,

and no slopes (only intercept) on by-item effects. At the critical region, the final model included only the main effect of distractor for by-subject effects, and that slope in addition to one interaction term for by-item effects.

4.2.2. Results

Word-by-word reading times of the different experimental conditions and by-condition numerical means at the spillover region are presented in Figure 4.1.

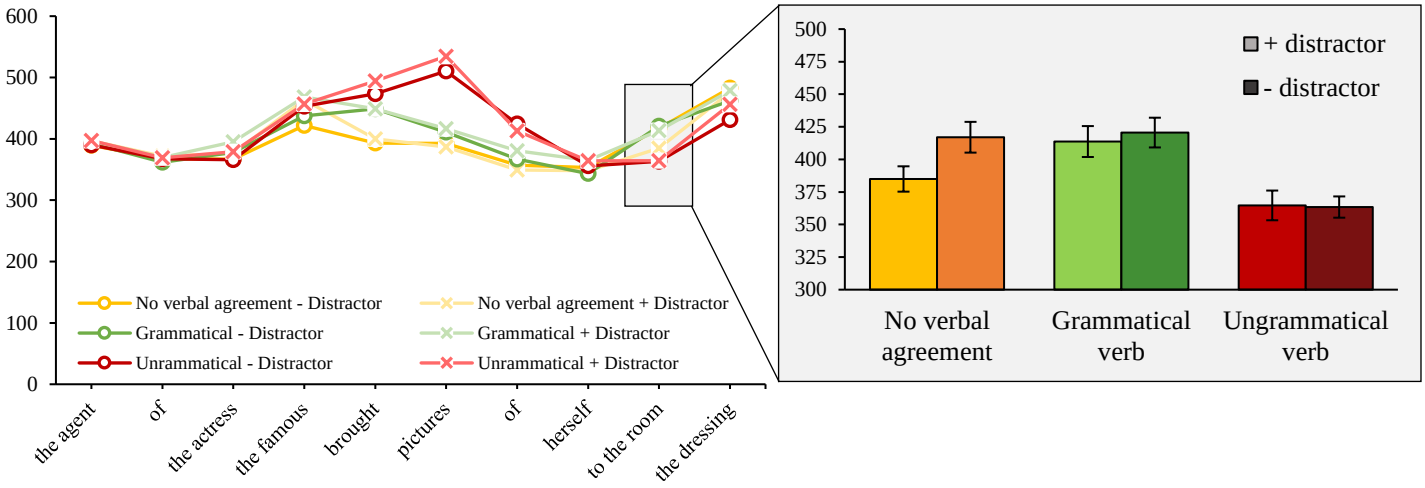

Figure 4.1. Word-by-word RT means (ms) by condition, Experiment 5. Error bars represent +/-SE.

The reflexive pronoun

No reliable effects were observed on the reflexive pronoun. The Bayesian analysis indicated several effects. However, none of them were reliable on the frequentist analysis (see Table 4.4 for the effects detected on the Bayesian analysis and their frequentist counterparts). The only effect which had some resonance in the frequentist analysis (with marginal significance) was an interaction between verbal agreement and the distractor's presence, for the grammatical verb case. This seemed to reflect in pairwise comparisons as an increase in reading times when the distractor matched the reflexive (and mismatched the subject), only in the grammatical verb conditions.

	Frequentist analysis				Bayesian analysis	
	Estimate	SE	*t*	*p*	Posterior Mean [CrI]	Posterior beyond zero
Main effect of Grammatical - No agreement	0.009	0.013	0.70	.48	9 [0, 19]	97%
Main effect of Ungrammatical - No agreement	0.010	0.013	0.72	.47	16 [5, 26]	> 99%

Interaction Grammatical:Distractor	0.022	0.013	1.68	.093	10 [1, 20]	98%
+/- distractor contrast, within grammatical verb conditions	0.044	0.020	2.24	.075	20 [5, 35]	> 99%

Table 4.4. Effects detected at the reflexive pronoun in Experiment 5, on the Bayesian analysis (posterior means, 95% credible intervals, and the probability of the parameter being beyond zero), and their frequentist counterparts (estimates, standard-errors, t-values, and p-values).

The spillover region

The results of the frequentist and Bayesian analyses for the spillover region are summarized in Table 4.5 and 4.6, correspondingly. Both analyses detected a main effect of verbal agreement, reflecting overall faster RTs on the ungrammatical verb conditions relatively to the no verbal agreement conditions (collapsing over +/- distractor cases). The analyses also detected an effect of distractor, with faster reading times when the distractor matched the reflexive pronoun. Given the contrast coding, this signifies the basic attraction effect in the baseline conditions, where gender cues were not available on the verb.

Pairwise comparisons revealed a reliable attraction effect, namely a decrease in RTs for the +distractor condition, only when no agreement cues were available at the verb. In addition, the Bayesian analysis suggested an interaction between verbal agreement and distractor, such that the effect of the distractor was less prominent in the ungrammatical verb case, relatively to the no verbal agreement conditions. Yet this interaction was not significant (marginal) on the frequentist analysis. No other effects were detected by any of the analyses.

	Estimate	SE	*t*	*p*
Main effects and interactions:				
Main effect of Grammatical – No agreement	0.020	0.017	1.13	.26
Main effect of Ungrammatical – No agreement	-0.076	0.015	4.93	< .001
Main effect of Distractor	-0.032	0.011	2.93	.003
Interaction Grammatical:Distractor	0.016	0.015	1.05	.29
Interaction Ungrammatical:Distractor	0.026	0.015	1.89	.058
Attraction effects (+/- distractor contrasts):				
No verbal agreement	-0.064	0.021	2.93	.01
Grammatical verb	-0.032	0.023	1.45	.44
Ungrammatical verb	-0.005	0.022	0.26	> .99

Table 4.5. Results of the frequentist analysis of the spillover region in Experiment 5: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero

Main effects and interactions:		
Main effect of Grammatical – No agreement	13 [-2, 28]	95%
Main effect of Ungrammatical – No agreement	-22 [-34, -10]	> 99%
Main effect of Distractor	30 [24, 36]	> 99%
Interaction Grammatical:Distractor	9 [-3, 22]	93%
Interaction Ungrammatical:Distractor	13 [0, 26]	98%
Attraction effects (+/- distractor contrasts):		
No verbal-agreement	-22 [-38, -6]	> 99%
Grammatical verb	-5 [-23, 14]	70%
Ungrammatical verb	2 [-16, 21]	58%

Table 4.6. Results of the Bayesian analysis of the spillover region in Experiment 5: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

4.2.3. Discussion

The results of Experiment 5 suggest some evidence for an effect of verbal agreement on attraction rates at a subsequent reflexive pronoun. Agreement attraction was observed only in cases where the verb did not manifest any agreement cues. This is in line with the hypothesis about the "shortcut" or the "stabilizer" roles of verbal agreement, which were laid out in the introduction to this study. These hypotheses suggest that attraction should be more prominent when the verb does not provide agreement features. Only in that case, agreement has to be evaluated directly from the subject phrase, a process which is affected by interference. When the reflexive can use the help of a proximal verbal cue or if the verbal agreement can pull the representation of the subject to match it, the distractor noun is given overall lower weight in the verification of the reflexive's agreement.

These hypotheses are also in line with the general pattern observed, where the reading times of the ungrammatical verb condition were relatively fast, while the reading times on the grammatical verb conditions were relatively slow. On these accounts, verbal agreement has high weight in determining the featural representation of the subject and/or the expected form of the reflexive. Hence, the reflexive pronoun is perceived as grammatical or ungrammatical not according to its agreement with the head of the subject noun, but relatively to the features on the verb. Thus, the reflexive in this experiment, which mismatches the subject across all conditions, is perceived as acceptable when it matches the verb (i.e. when the verb mismatches the subject as well). This results in fast reading times at the reflexive in ungrammatical verb conditions. On the grammatical verb conditions, the opposite occurs - the reflexive creates processing difficulty since it mismatches the verb.

Although the overall pattern of the results seems to be in line with only one of the possible predictions, these findings should not be considered conclusive. The interactions in this study, required for deducing any difference between the attraction patterns in the different agreement conditions, were not reliable. This makes it hard to draw any conclusions from these data. One of the interaction effects, suggesting that the distractor had less effect on the ungrammatical

verb conditions relative to the no verbal agreement case, was observed on the Bayesian analysis. Yet the above interpretation requires both interaction terms, namely it also requires the grammatical conditions to differ from the no verbal agreement conditions, an interaction which was not observed. Additionally, the frequentist analysis failed to detect a significant effect even for the case of ungrammatical verbs. Therefore, the results of this experiment are inconclusive by themselves.

In addition, there seems to be another possible explanation for the missing attraction effect following ungrammatical verbs, and the fast reading times on these conditions. This explanation does not appeal to fallibility of memory mechanisms, but to task-specific strategies. It is possible that once readers identify the ungrammaticality of the verb, they give up on interpreting the sentence as a whole, or specifically neglect subsequent agreement relations. On this account, the fast reading times in the ungrammatical verb conditions represent participants' lack of commitment to the sentence's structure rather than a preference for matching agreement features with the verb. If participants indeed "give up" on the sentence, this should also create a floor effect by which attraction is not detectable, since retrieval and agreement computations are put on hold. This interpretation would not be in line with results of Molinaro et al. (2008) discussed above, where an ungrammaticality effect (a P600 component) was observed on a reflexive downstream from an ungrammatical verb. However, since that study used a different method (ERP), language (English), and agreement feature (number), it is possible that Molinaro's findings are not directly relevant in this case. In addition, If one would take seriously the lack of attraction in Grammatical Verbal Agreement conditions, despite the lack of significant interaction, this is not predicted by the task-specific strategy account (though this is not a strong claim naturally).

Lastly, it should be noted that an unpredicted effect was observed on the reflexive pronoun, namely increased reading times for +distractor in the grammatical verb conditions. This effect could, in principle, reflect spillover from the verb region, namely slowdown when the verb matches both the subject and the distractor. This interpretation could be reasonable given that in Arabic gender attraction was observed only on late regions. However, since the effect was not mirrored in the ungrammatical verb conditions, and since it appears on the third word after the verb, this interpretation should be considered carefully.

Given all this, and in order to validate the results, I conducted two additional experiments. Experiment 6 aimed to explicitly test the processing of reflexive pronouns following ungrammatical verbs, and check if the parser neglects the agreement of the reflexive when the sentence is ungrammatical already at the point of the verb. Experiment 7 tested again the same effects and predictions I presented with regards to Experiment 5, with a different task and different sentential structure. Thus, Experiment 6 aims to establish the validity of testing the ungrammatical verb conditions, and Experiment 7 returns to the main question of this chapter and tries to determine if and how verbal agreement affects the vulnerability of the reflexive to attraction effects.

4.3 Experiment 6: Reflexive preferences following a subject-verb mismatch

Experiment 6 tests for readers' processing of reflexives matching or mismatching the subject, following grammatical and ungrammatical verbs. This experiment had two aims. First, I aimed to verify that readers do not neglect additional subsequent agreement relations when some mismatch or ungrammaticality appears early in the sentence. This is crucial for our ability to make deductions from the ungrammatical verb conditions in Experiments 5 and 7. Testing the processing strategy taken at reflexive pronouns after verbs which mismatch the subject, assumes that results from such conditions reflect normal parsing mechanisms rather than a strategic shift to shallow processing. This needs to be verified, in the current experiment. Second, the experiment aimed to establish an interaction between the agreement of a reflexive and that of the verb, similar to the one reported by Molinaro, et al. (2008), only for gender features. Previous experiments which investigated the relation between multiple agreeing entities within one sentence (Antón-Méndez et al., 2002; Molinaro, et al., 2008, reviewed above in subsection 4.1.4) used methods different from those used in the current study (production and ERP). More importantly, their findings relate to number agreement, which, as established in subsection 4.1.1, give rise to different patterns than those observed in gender agreement. Thus, such extension is important for interpreting the results of the current study.

4.3.1 Methods

Participants

Participants were 60 native speakers of Hebrew (according to self-report) (mean age: 24.78, range: 18-33). Three participants were bilingual of Hebrew and Russian, and the rest were monolingual.

Materials

The experiment included 28 item sets of six conditions, manipulating subject-verb agreement and the agreement of a distractor noun (see example set in Table 4.10). The experiment included 84 filler items similar to the ones used in Experiment 5.

Condition	Sentence
Grammatical verb, Grammatical reflexive	*ha-talmid šel ha-mexanexet ha-xadaša ciyer karikaturot* the-student.M of the-teacher.F the-new.F drew.M caricatures *šel* ***acmo*** *ba-maxberet ha-išit* of **himself** in.the-notebook the-personal 'The student.M of the new teacher.F drew.M caricatures of himself in the personal notebook'
Grammatical verb, Ungrammatical reflexive	*ha-talmid šel ha-mexanexet ha-xadaša ciyer karikaturot* the-student.M of the-teacher.F the-new.F drew.M caricatures

	šel ***acma*** *ba-maxberet ha-išit* of **herself** in.the-notebook the-personal 'The student.M of the new teacher.F drew.M caricatures of herself in the personal notebook'
Ungrammatical verb, Grammatical reflexive	*ha-talmid šel ha-mexanexet ha-xadaša ciyra karikaturot* the-student.M of the-teacher.F the-new.F drew.F caricatures *šel* ***acmo*** *ba-maxberet ha-išit* of **himself** in.the-notebook the-personal 'The student.M of the new teacher.F drew.F caricatures of himself in the personal notebook'
Ungrammatical verb, Ungrammatical reflexive	*ha-talmid šel ha-mexanexet ha-xadaša ciyra karikaturot* the-student.M of the-teacher.F the-new.F drew.F caricatures *šel* ***acma*** *ba-maxberet ha-išit* of **herself** in.the-notebook the-personal 'The student.M of the new teacher.F drew.F caricatures of herself in the personal notebook'

Table 4.10. Example set from Experiment 6.
The reflexive is marked in bold; F = feminine grammatical gender; M = masculine grammatical gender.

Procedure

The experiment was a web-based self-paced reading experiment. Fifty percent of the sentences were followed by a 'yes/no' comprehension question.

Data analysis

I examined the reflexive pronoun and one word after it in separate models. The experimental fixed effects in this analysis, the grammaticality of verb and that of the reflexive, were sum coded. In addition, within each model, I conducted pairwise comparisons contrasting the two versions of the reflexive pronoun (one for reflexives which followed an ungrammatical verb, and one for those following a grammatical verb). Thus, *p*-values on the frequentist analysis were corrected using Bonferroni correction for two comparisons. Due to convergence failure, the final model of the spillover region included only one slope (corresponding to the main effect of reflexive form) on by-subject random effects, and no slopes (only intercept) on by-item random effects. At the critical reflexive, the final model included only intercepts (and no slopes) for both by-subject and by-item random effects.

Exclusion criteria resulted in removal of seven participants (three for low accuracy in comprehension questions and four for abnormal reading times). For the remaining 53 participants, trimming of high and low RTs affected 2.06% and 0.37% of the data, correspondingly.

4.3.2 Results

Word-by-word reading times of the different experimental conditions, and by-condition numerical means at the spillover region, are presented in Figure 4.2. The results of the frequentist and Bayesian analyses are summarized in Table 4.8 and 4.9, correspondingly.

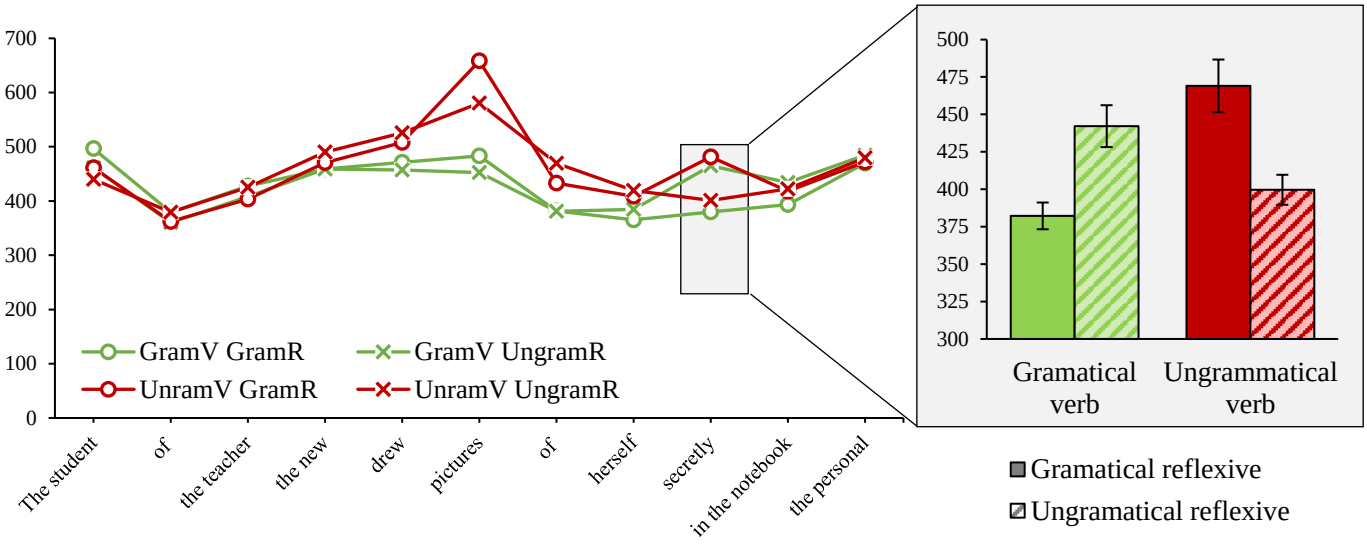

Figure 4.2. Word-by-word RT means (ms) by condition, Experiment 6. Error bars represent +/-SE.

The reflexive pronoun

At the reflexive pronoun, the only effect detected was a main effect of verb grammaticality. On the Bayesian analysis, reading times were higher for pronouns in the ungrammatical verb conditions, relative to their counterparts in the grammatical verb conditions (Posterior mean [CrI]: -11 [-18, -4]; Probability of the parameter beyond zero was over 99%). This effect failed to reach significance in the frequentist analysis (Estimate = -0.013, SE = 0.007, t = 1.88, p = .06). None of the other effects or contrasts of interest was detected by either of the analyses.

The spillover region

At the spillover region, I observed an interaction of verb grammaticality and reflexive grammaticality. Pairwise comparisons, contrasting the grammatical and ungrammatical reflexive forms for each verb type, confirmed that the reflexive form which matched the subject was read faster than the ungrammatical reflexive following a grammatical verb. In contrast, after an ungrammatical verb, the reflexive matching the subject (and mismatching the verb) incurred slower reading times relative to the ungrammatical reflexive (matching the ungrammatical verb). The interaction and the two pairwise comparisons were reliable on both Bayesian and frequentist analysis (see Tables 4.11 and 4.12). None of the other effects were reliable on either of the analyses.

	Estimate	SE	t	p
Main effects and interactions:				

Main effect of verb grammaticality	-0.004	0.008	0.46	.64
Main effect of reflexive grammaticality	0.000	0.009	0.05	.96
Interaction	-0.047	0.008	6.02	< .001
Attraction effects (reflexive form contrasts):				
Following a grammatical verb	-0.095	0.024	3.92	< .001
Following an ungrammatical verb	0.093	0.024	3.85	< .001

Table 4.11. Results of the frequentist analysis of Experiment 6: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Main effects and interactions:		
Main effect of verb grammaticality	-6 [-14, 2]	93%
Main effect of reflexive grammaticality	-1 [-9, 7]	61%
Interaction	-18 [-26, -9]	> 99%
Attraction effects (reflexive form contrasts):		
Following a grammatical verb	-37 [-59, -16]	> 99%
Following an ungrammatical verb	33 [8, 58]	99%

Table 4.12. Results of the Bayesian analysis of Experiment 6: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

4.3.3 Discussion

The results of Experiment 6 show that following an ungrammatical verb, readers prefer a reflexive pronoun matching the verb rather than one matching the subject. This pattern suggests that readers do not neglect subsequent gender agreement relations after detecting an earlier error of this sort. Namely, readers do not abandon the attempt to reach grammaticality or interpretability, but rather switch their agreement preferences (to prefer matching with the verb rather than with the subject). This is in line with ERP results from Molinaro and colleagues (2008), which indicated a similar pattern for English number agreement (discussed in subsection 4.1.4). This extension of Molinaro's findings to the self-paced reading paradigm and to the gender agreement realm is important for the current study since it supports the case for interaction between verbal and reflexive agreement. These findings thus license further investigations as to how verbal agreement intervenes in the process of associating the reflexive with its antecedent and its vulnerability to agreement attraction.

4.4 Experiment 7: Verbal agreement and reflexive attraction in production

Experiment 7 aimed to re-examine the effect of explicit verbal gender cues on attraction at a subsequent reflexive. This experiment followed the 2-by-3 manipulation of distractor and verbal agreement as in Experiment 5, with some changes to the task and the materials. First, I switched the method to force-choice sentence completion, in the hope that binary data will give rise to more conclusive results. Second, I amended certain characteristics of the reflexive's syntactic environment to avoid possible uncontrolled factors (see details in the Materials subsection).

4.4.1 Methods

Participants

Participants were 57 native speakers of Hebrew (according to self-report) (mean age: 24.80, range: 19-31). Two participants were bilingual speakers of Hebrew and either Russian or English, and the rest were monolingual.

Materials

The experiment included 24 item sets of six conditions, manipulating subject-verb agreement and the agreement of a distractor noun (see example set in Table 4.7). For constructing an environment where no agreement features are available at the verb (no verbal agreement conditions), I used embedded non-finite clauses in object control constructions. This was done in order to avoid dative subjects in possessive *yeš* constructions, which were used in Experiments 5-6. Dative subjects could be deviant in terms of the subject's availability to subsequent processes, due to the unusual prepositional marking. In addition, reflexive pronouns in Experiment 7 appeared in direct object position, rather than inside picture NPs, to exclude the possibility of interference from the picture noun itself (which also carries agreement features), and to allow better comparison with other literature regarding reflexive pronouns.

Condition	Sentence
No verbal agreement, +distractor	*hayom darašnu me-ha-talmid šel ha-mexanexet ha-* today we.demanded from-the-student.M of the-teacher.F the- *xadaša lehosif le-rešimat ha-tfuca et* [*acmo* | *acma*] new.F to.add to-list the-circulation ACC [himself | herself] 'Today we required the student.M of the new teacher.F to add [himself|herself] to the mailing list'
No verbal agreement, no distractor	*hayom darašnu me-ha-talmid šel ha-mexanex ha-* today we.demanded from-the-student.M of the-teacher.M the- *xadaš lehosif le-rešimat ha-tfuca et* [acmo | acma] new.M to.add to-list the-circulation ACC [himself | herself]

	'Today we required the student.M of the new teacher.M to add [himself|herself] to the mailing list'
Grammatical verb, +distractor	*hayom darašnu še-ha-talmid šel ha-mexanexet ha-* today we.demanded that-the-student.M of the-teacher.F the- *xadaša yosif le-rešimat ha-tfuca et* [*acmo* | *acma*] new.F will.add.M to-list the-circulation ACC [himself | herself] 'Today we required that the student.M of the new teacher.F will add.M [himself|herself] to the mailing list'
Grammatical verb, no distractor	*hayom darašnu še-ha-talmid šel ha-mexanex ha-* today we.demanded that-the-student.M of the-teacher.M the- *xadaš yosif le-rešimat ha-tfuca et* [acmo | acma] new.M will.add.M to-list the-circulation ACC [himself | herself] 'Today we required that the student.M of the new teacher.M will add.M [himself|herself] to the mailing list'
Ungrammatical verb, +distractor	*hayom darašnu še-ha-talmid šel ha-mexanexet ha-* today we.demanded that-the-student.M of the-teacher.F the- *xadaša tosif le-rešimat ha-tfuca et* [*acmo* | *acma*] new.F will.add.F to-list the-circulation ACC [himself | herself] 'Today we required that the student.M of the new teacher.F will add.F [himself|herself] to the mailing list'
Ungrammatical verb, no distractor	*hayom darašnu še-ha-talmid šel ha-mexanex ha-* today we.demanded that-the-student.M of the-teacher.M the- *xadaš tosif le-rešimat ha-tfuca et* [acmo | acma] new.M will.add.F to-list the-circulation ACC [himself | herself] 'Today we required that student.M of the new teacher.M will add.F [himself|herself] to the mailing list'

Table 4.7. Example set from Experiment 7.
The forced-choice alternatives appear in square brackets. ACC = accusative case marker; F = feminine grammatical gender; M = masculine grammatical gender.

The experimental items were distributed in a Latin square design across six lists. Within each list, the target sentences were combined with 56 grammatical filler sentence fragments, for a total of 80 sentences. The filler items included sentences with embedded structures, with similar proportions of non-finite (18 items), feminine (19 items), and masculine (19 items) verbs. In 16 of the filler items, the grammatical pronoun choice was of a feminine reflexive, in order to balance the experimental materials (where a reflexive matching the subject was male). The remaining 40 filler items were to be completed with regular (non-reflexive) pronouns, such that there should be no overall bias in the experiment to feminine or masculine pronouns.

Procedure

The experiment was built using Ibex Farm. Participants pressed a key to begin each trial. Then, the preamble of the sentence was presented using rapid visual serial presentation (RSVP) at a rate of 400ms per word (with no inter-stimulus interval). After each preamble, participants were instructed to choose one of two pronominal forms which appeared on the screen as a continuation for the sentence they had just read. The feminine form consistently appeared on the right side of the screen, and the masculine form on the left. These pronouns were reflexive for all experimental items and some of the fillers, and non-reflexive pronouns in the accusative form (*oto* vs. *ota*, 'him' and 'her', correspondingly) for other filler items. Before they began the experiment, participants underwent a practice block of six items. The order of presentation was randomized for each participant.

Data analysis

Participants were excluded from the analysis if their accuracy on filler items was lower than 85% (resulting in the removal of nine participants). For the remaining 48 participants, trials in which participants took over 5000 milliseconds to choose the appropriate pronoun were excluded from analysis (affecting 0.08% of the data), to tap into the more intuitive pronoun preferences.

I analyzed the pronoun choice (matching or mismatching the subject) in a logistic regression. As in Experiment 5, I used a treatment (dummy) coding scheme to for the verbal agreement manipulation (to compare reading times in the grammatical and ungrammatical conditions against the no verbal agreement condition), and sum coding for the distractor manipulation (to collapse over both +/-distractor conditions). In addition, to detect attraction effects in each case, three pairwise comparisons were made, contrasting the +/-distractor levels in each level of verbal agreement. Therefore, on the frequentist analysis, pairwise comparisons were corrected by Bonferroni correction to three comparisons. Due to repeated convergence failures, the final model included only intercepts (and no slopes) on both by-subject and by-item random effects.

4.4.2. Results

Pronoun choice rates are presented in Figure 4.3, partitioned by experimental condition. The results of the frequentist and Bayesian analyses are summarized in Table 4.8 and 4.9, correspondingly.

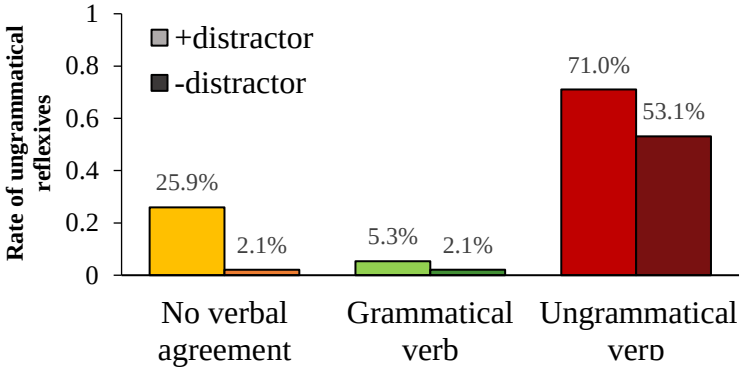

Figure 4.3. By-condition rates of choosing the (feminine) reflexive from mismatching the (masculine) subject, Experiment 7.

The analyses detected a main effect of verbal agreement, with higher rate of choosing the reflexive form matching the subject following verbs with grammatical agreement (relative to verbs with no agreement). Another reliable main effect reflected higher rates of choosing the reflexive form mismatching the subject following verbs with ungrammatical agreement (relative to verbs with no agreement). The analyses also detected an effect of distractor, such that a distractor (mismatching the subject) elicited more choices of the ungrammatical reflexive form (mismatching the subject and matching the distractor). Yet note that given the contrast coding, this signifies the basic attraction effect in the baseline conditions, where gender cues were not available on the verb.

One of the interaction terms was reliable on both frequentist and Bayesian analyses: both analyses indicate that attraction (i.e. the effect of the distractor) is more prominent when the verb does not provide agreement cues relatively to the case of ungrammatical verbs. In addition, the frequentist analysis (but not the Bayesian one) suggested an interaction between verbal agreement and distractor, such that the effect of the distractor was less prominent in the grammatical verb case, relatively to the no verbal agreement conditions. No other effects were detected by any of the analyses.

Pairwise comparisons revealed a reliable attraction effect, namely likelier choice of the reflexive form which mismatched the subject for the +distractor condition, in two out of the three cases: when no agreement cues were available at the verb, and when the verbal agreement was ungrammatical (mismatching the subject), but not when the verb had grammatical (subject matching) gender features.

	Estimate	SE	*t*	*p*
Main effects and interactions:				
Main effect of Grammatical – No agreement	1.00	0.38	2.60	.009
Main effect of Ungrammatical – No agreement	-2.98	0.28	10.81	< .001
Main effect of Distractor	-1.31	0.24	5.38	< .001
Interaction Grammatical:Distractor	0.77	0.38	2.00	.045
Interaction Ungammatical:Distractor	0.93	0.27	3.48	< .001
Attraction effects (+/- distractor contrasts):				
No verbal-agreement	-2.61	0.49	5. 38	< .001
Grammatical verb	-1.08	0.59	1.81	.21
Ungrammatical verb	-0.75	0.23	3.32	.002

Table 4.6. Results of the frequentist analysis of Experiment 7: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Main effects and interactions:		
Main effect of Grammatical – No agreement	1.54 [0.73, 2.47]	> 99%
Main effect of Ungrammatical – No agreement	-3.08 [-3.66, -2.52]	> 99%
Main effect of Distractor	-1.17 [-1.65, -0.72]	> 99%
Interaction Grammatical:Distractor	0.59 [-0.19, 1.35]	93%
Interaction Ungammatical:Distractor	0.70 [0.14, 1.25]	99%
Attraction effects (+/- distractor contrasts):		
No verbal-agreement	-2.34 [-3.30, 1.45]	> 99%
Grammatical verb	-1.17 [-2.67, 0.24]	94%
Ungrammatical verb	-0.93 [-1.83, -0.08]	98%

Table 4.7. Results of the Bayesian analysis of Experiment 7: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

4.4.3. Discussion

The results of Experiment 7 are in line with those of Experiment 5. First, Experiment 7 indicated agreement attraction on reflexive pronouns when these follow a verb with no agreement cues. This attraction rate seems comparable to that found in other production experiments. Targeting number agreement, Bock and colleagues (1999) found erroneous plural reflexives on 2% of productions in the absence of a plural distractor, and on 17% of productions in a plural distractor condition. In production of tag pronouns, Bock and colleagues (1999, 2004) found between 0-8% errors without the distractor, and between 17%-39% with it. Overall, these numbers seem comparable to the 2% and 26% error rate in the current study.

More importantly, Experiment 7 provides further support to the hypothesis that verbal agreement may act as a "shortcut" or a "stabilizer" when the parser requires re-access to the subject's features. The results suggest that reflexive attraction occurs at a lower rate when the verb manifests some agreement cues. This is in line with the hypothesis that verbal agreement gets high priority in determining other agreement relations in the sentence. Specifically, I suggested in the introduction to this chapter and in the discussion of Experiment 5 that verbal agreement sets either the subject's long-lasting representation or the reflexive's agreement.

In line with this, the results of Experiment 7 also exhibited a preference for using the reflexive form matching the verb rather than the subject, when there is a mismatch between the latter two. In Experiment 7, participants picked the ungrammatical reflexive matching at a higher rate when this produces a match with the (ungrammatical) verb. Note that this result is unpredicted under Cue-Based Retrieval, where cues of the reflexive should not directly interact with those of the verb. In Cue-Based Retrieval the effect of verbal agreement is mediated through activation of the distractor. Therefore, this model cannot account for the results in the condition where there is no available distractor (when the intervening noun mismatches the verb and the reflexive). In addition, this high rate of using reflexives mismatch the subject across the ungrammatical verb conditions cannot be interpreted as copying of features on the

previous word, since three additional words - a two-word adverbial and the accusative marker - appeared after the verb and before the reflexive choice was prompted. These should have allowed decay of the verbal form. In addition,

It should be noted that in this experiment the reflexive preferences following ungrammatical verbs seem to be close to chance level. This pattern may contrast with previous results (Molinaro et al., 2008, and Experiment 6 on the current study), where a reflexive matching the ungrammatical verb was perceived more like reflexives in fully grammatical sentences. This could suggest some contrast between comprehension and production. It could be that when participants are required to choose a completion for the sentence, and are even explicitly directed between two agreement forms, they monitor agreement relations more closely and are less forgiving towards errors in the input. Future research could test that using more closely matched comprehension and production experiments. Additional ways of tackling this question could be comparing these results with productions on a paradigm which does not explicitly point participants to notice agreement relations (e.g. free sentence completion).

4.4 General discussion

Does the fallibility of agreement in one part of the sentence increase fallibility later on? Can the parser resist such feedforward interference by using independent agreement processing? Can the parser altogether avoid the interference effects of long-distance agreement through intermediate steps which validate agreement features? The current study tried to answer these questions by examining the susceptibility of reflexive pronouns to agreement attraction following verbs which provide congruent, incongruent or no agreement cues. This manipulation was enabled through a unique feature of Semitic languages like Hebrew, where gender agreement regularly appears in both pronominal agreement and verbal agreement.

Results from both sentence completion and reading times (and from reflexives in two different syntactic positions) seem to be in line with the same conclusion - gender cues on the verb reduce the effect of a distractor noun on the processing of reflexive pronouns. We observed more prominent attraction when the reflexive followed a verb with no agreement marking. Verbal agreement thus seems to help the parser avoid attraction in subsequent sites which require agreement with the subject, in this case reflexive pronouns. This might occur through intermediate updating the representation of the subject, and giving high weight to the verb's features, or through bypassing direct retrieval of the subject and establishing agreement only with the local verb.

Similar interpretations were suggested by Molinaro and colleagues (2008). These authors suggested that when the verb and the subject mismatch, the subject's features can be coerced to match those of the verb, or get dismissed for subsequent agreement relation within the same verb phrase. I suggest these mechanisms should be expanded to sentences with grammatical verbs as well. The resistance to attraction following verbs with grammatical agreement features suggests either reinforcement of the correct subject representation (extension of Molinaro's coercion proposal), or bypassing subject retrieval which in turn prevents possible misretrievals (extension of Molinaro's independent verb phrase proposal). To further understand the possibilities, we should consider them within the two prominent theories of establishing agreement attraction - Cue-Based Retrieval and Marking and Morphing.

4.4.2. Interpretation within models of agreement attraction

As presented in the first chapter, agreement attraction may be attributed to representation or retrieval mechanisms. In representational accounts of attraction, as offered in the Marking and Morphing framework, the grammaticality illusion arises due to an unstable representation of the agreement features on the subject phrase. In that case morphosyntactic cues on the distractor contribute to the agreement representation of the subject phrase. The Marking and Morphing model, initially suggested for number agreement, describes number marking as a value on a continuous scale between "unambiguously plural" and "unambiguously singular" (i.e. the number is a gradient feature value rather than a binary one). This can apply similarly to gender agreement with a scale between "unambiguously feminine" and "unambiguously masculine". The feature value of the subject phrase is calculated by combining the value from the head noun, and weighted activations which spread from the morphological marking of other items

(e.g. the distractors). When the pooled value is set to an intermediate level on the scale, agreement attraction may arise.

On such an approach, it seems very natural that the agreement morphology of the verb contributes to the overall representation of the subject's features. If the spreading activation from the verb is weighed in determining the gender value of the subject phrase, this can counter the effect of distractors. Indeed, the verb could probably overpower the effect of a distractor since the verb is closely associated with the subject through predication, and constitutes a prominent element. This would also mean that the verb can pull the gender value of the subject in either direction. Namely, the verb may coerce the subject's representation to align with its mismatching feature, in the presence or absence of distractors. Therefore, the Marking and Morphing model is in line with the idea that agreement information on the verb is taken into account in establishing the agreement features of the subject.

This model might also be compatible with the other alternative proposed above, according to which the parser avoids retrieval of the subject altogether. However, the "shortcut" mechanism does not follow from the model's assumptions. Therefore, such an interpretation requires an additional implementation of shallow processing or a bias for local coherence. On this note it should be mentioned however that subject-verb agreement does not seem to be sensitive to such proximity bias. The linear distance between the verb and the distractor does not seem to directly affect the rate of attraction (Vigliocco & Nicol, 1998; Frank, Vigliocco, & Nicol, 2002; Gillespie & Pearlmutter, 2011). Yet it could be that such a strategy is adopted specifically for reflexive pronouns, as they participate in dependencies spanning longer distances.

On the other hand, I suspect that the Cue-Based Retrieval model does not accommodate either of the above suggestions (the "shortcut" or the "stabilizer" roles of verbal agreement). Three findings specifically are at odds with the Cue-Based Retrieval model: the effect of grammatical verbal agreement on attraction rate (Experiment 5 and 7), the direction of the effect of ungrammatical verbal agreement on attraction rate (Experiment 5 and 7), and the high rate of choosing reflexive forms which mismatch the subject in absence of a distractor (Experiment 7). The last point could arguably be explained by task effects like a confusion resulting in chance-level performance. Therefore I focus on possible ways in which Cue-Based Retrieval could (or could not) apply a "shortcut" or a " stabilizer" mechanism to account for the first two issues.

As explained in Chapter 1, under Cue-Based Retrieval, attraction arises due to multiple partial matches. The agreement cue which ungrammatical verbs/reflexives provide, and the structural cue, specifying the position which the target should take, result in the activation of two different items. The race between the target (matching only the structural cue) and the distractor (matching only the agreement cue) can lead to erroneous retrievals of the distractor. In addition, this also predicts faster processing (on average) than in sentences with only one partial match (ungrammatical dependency with no distractor). In absence of a distractor (and a race), the activation of the target is not time-bound by a parallel alternative activation of a distractor (Vasishth, Nicenboim, Engelmann, & Burchert, 2019). With this in mind, let us consider again the results of this study.

In the Cue-Based Retrieval model, the activation of previous constituents rapidly decays, and retrieval of previous items is executed based on cues that the current trigger provides. These two crucial assumptions seem to argue against the "shortcut" hypothesis. It seems unlikely that at the point of the reflexive pronoun, across two additional words (*picture of* on Experiment 5, and an adverbial phrase in Experiment 7), the verb's activation is still high enough to affect the processing of the reflexive without retrieval. To affect processing of the reflexive, the verb would have to be retrieved. However, retrieval of the verb at the reflexive does not seem plausible at all. In my understanding, the reflexive pronoun initiates retrieval of the subject using cues which guide the search to nominal phrases. Hence, the verb should not be retrieved at that point and could not affect the agreement of the reflexive pronoun. One could argue that the verb could be erroneously retrieved due to a partial match with the reflexive's agreement cue (like nominal distractors are retrieved). However, this should occur only when the verb is ungrammatical as well. Thus, this mechanism cannot account for the failure to detect reflexive attraction following grammatically inflected verbs in the current study.

As for the "stabilizer" interpretation, this would also be hard to accommodate within a Cue-Based Retrieval model. Retrieval should not affect the representation of the retrieved items but only bring them into the focus of attention. Thus, it is unclear how the verb could serve to coerce some agreement features on the subject. Moreover, even if we assume that retrieval also allows some feature overwriting, this should apply only in the case of ungrammatical verbs. Therefore, this mechanism also fails to account for the lack (or decrease) of reflexive attraction following grammatically inflected verbs.

Lastly, as mentioned in the introduction to this study, the common implementation of Cue-Based Retrieval incorporates the retrieval history of an item as a part of its activation level. This predicts that items which were recently retrieved would more likely distract subsequent retrievals. Hence, this model predicts that accessing the distractor at the verb (i.e. occurrence of agreement attraction upon reading the verb) would facilitate reactivation of the distractor at the reflexive pronoun. Namely, attraction at the reflexive pronoun should be more pronounced in conditions where verbal attraction usually arises – following ungrammatical verbs. This prediction is not borne out in the current datasets.

Given these challenges, I believe the pattern results is not consistent with a Cue-Based Retrieval account. However, the study was not designed to tease apart the predictions of these two attraction implementations. Therefore, these conclusions should be taken with a grain of salt. Moreover, crucial parts of the above interpretation are based on the assumption that attraction is not evident in reflexive pronouns which follow grammatical verbs, yet the statistical reliability of this conclusion is questionable. Deducing from null pairwise comparisons is problematic, especially given the relatively high posterior distribution on Experiment 7. Moreover, an interaction which would suggest that attraction is less prominent in these conditions was not detected at all in Experiment 5 and was not robust enough on the Bayesian analysis of Experiment 7. Therefore, the proposed conclusions from this study are provisional at this point.

4.4.1. Directions for future research

The interpretations presented here seem to naturally give rise to some to additional research questions. First, it seems crucial to distinguish between the two possible interpretations of the verb's contribution here, which are very different in their approach to processing. The "shortcut" idea assumes some sort of shallow processing, or proximity concord, while the "stabilizer" interpretation assumes a fluid representation of morphosyntactic features. These have different consequences for conceiving how memory limits our language comprehension. The current study cannot distinguish between these two mechanisms, but they are not impossible to tease apart. For example, the two interpretations seem to give rise to different predictions regarding the activation level of the target subject. On the "shortcut" account, the subject is fully retrieved only when the verb does not bear agreement cues. Thus, on a probe recognition task immediately after reading the sentence, response to the subject noun should be slower when verbal agreement is unavailable. On the other hand, this effect is not predicted if verbal agreement enhances the representation of the subject. In fact, it could even be suggested that by stabilizing the subject's agreement, the inflected verb could make the subject more accessible also for the probe recognition task.

Other possible follow-ups may include testing other types of agreeing elements. Attributive adjectives, for example, appear closely together with target nouns, but have a less prominent role in the sentence as a whole. Moreover, they do not constitute a single phrasal projection with subsequent verbs or reflexive pronouns. Thus, it could be interesting to see if an agreement mismatch between the head subject noun and its modifying adjective induces a similar change in later agreement relations.

4.4.2. Conclusions

In the context of this dissertation, the results of the current study suggest another way in which the parser utilizes the available linguistic information and adopts parsing strategies which help us keep errors under control. The results exhibit that different agreement relations interact during incremental processing, and thus help readers avoid interference. Verbal agreement, which usually confirms the feature representation of the subject, can help later processing stages which require re-accessing that subject. In principle, support could arise at the point of the verb (enhancing and anchoring the features of the subject phrase) or at the point of the second retrieval (providing a more local delegate which the reflexive can verify agreement with). Whatever parsing strategy underlies it, the attraction pattern that this study identifies may suggest that comprehenders have tools to avoid lapses of their own parsing mechanism, like agreement attraction effects. Comprehenders might be able to skip over the pitfalls of unstable representation traces or risky memory access, by assigning high weight to intermediate agreement sites (in languages where such sites are available).

Yet it should be noted that these conclusions necessitate replication and additional investigation on two grounds. First, some of the reported effects were not statistically robust enough. Second, the parallels between comprehension and production should be considered more closely. So far, I have treated the evidence from comprehension and production as converging. However, distinguishing and closely comparing the two processes is crucial for

reaching a conclusion about the optimality of the above policy. I suggested that a parser which is prone to interference may benefit from verifying agreement on intermediate verbal positions. However, while this strategy may help comprehenders avoid interference errors originating from their own parsing mechanism, it makes comprehension more vulnerable to cascading an error of the producer. If the comprehender relies on the agreement at the verb rather than that of the subject itself, a production error in verbal agreement (as in my ungrammatical verb conditions) would cause the parser consistently to choose the "wrong" agreement for the reflexive (instead of erring only when a distractor interferes). Does the benefit of reducing attraction on grammatical verb cases exceed the offset caused by ungrammatical verb cases? This requires additional investigation as to the likelihood of production errors (and specifically verb agreement errors with intact reflexivity). Only with this, we will be able to compare the different possible parsing strategies on the grounds of their overall error rate.

Chapter 5
Expecting possible errors rather than rare linguistic structures

5.1 Introduction

On this last study, I ask more directly whether incremental parsing decisions take into account the possibility of noisy input. Namely, I test for noisy-channel inference, as defined by recent models of rational (Bayesian) processing. The framework of Rational Noisy-Channel processing aims to explain how listeners and readers manage to understand what their interlocutor is trying to convey, even though many utterances are corrupted by noise. In this framework, the addressee weighs the probability distribution of the different structures against the probability that noise corrupted the input into its current form (Levy, 2008b). Thus, comprehenders take into account the possibility of speech and print errors, as well as perception errors, and maintain some uncertainty with regard to the input. As a result, they may be pulled towards "near-neighbour", extrapolated interpretations, when the literal interpretation of the input is rare or implausible.

Most studies investigating the Noisy-Channel hypothesis (e.g. Futrell & Levy, 2017; Gibson, et al., 2017; Ryskin, et al., 2018) have utilized the paradigm developed by Gibson and colleagues (2013) for testing noisy-channel inference. Using comprehension questions, Gibson and colleagues (2013) probed the rate of unfaithful interpretations in different settings and structures. Unfaithful readings were chosen at a higher rate when this achieved semantically more plausible interpretations and required fewer edits of the input. Namely, this paradigm focuses on semantic plausibility as the prior probability manipulation, and different types of extrapolated "errors" which can restore a congruent meaning (deletion, insertion and exchange). Thus, the frequency of syntactic structures (rather than their semantic meaning) is rarely manipulated (apart from the study of Levy, 2011. See the Introduction chapter, section 1.3.4 for review of their findings). Moreover, Gibson et al. (2013) and the papers which followed it measure reader's interpretation using comprehension questions, an offline task. The extent to which noisy-channel inference affects incremental processing decisions is again studied only in Levy et al. (2009) and Levy (2011).

In the current study, I investigate how infrequent a structure has to be in order for readers to assume that it was corrupted by noise, namely that it is a noisy version of another structure.

I aim to exhibit the effect of the prior probability of the structure required for the faithful parse. I will try to argue that sometimes readers do not interpret a mismatch in agreement features as evidence for a different grammatical structure. Namely, the markedness of the alternative can modulate how likely readers are to treat the mismatch as a cue for interpretation shift. To this end, I use unique sentence structures in Hebrew which are temporarily ambiguous between object relative clauses and subject relative clauses. Therefore, before approaching the experimental design, we take a short detour to consider the rich literature covering the preference of subject over object relative clauses. Then I present the crucial information regarding word order in Hebrew (subsection 5.1.2) and the logic of the current study (subsection 5.1.3).

5.1.1 The subject/object relative clause asymmetry

One of the most well-known findings in the sentence processing literature is the difference in difficulty of two types of relative clauses: Object-extracted and subject-extracted relative clauses. The distinction between the two relative clause types concerns the syntactic position of the gap within the embedded clause. Thus, in object relatives (39b), the gap is in the object position and the filler is interpreted as the object of the relative clause verb, while in subject relatives (39a) the embedded subject contains the gap and refers to the filler.

(39) a. Here is a filler [which _ has a corresponding gap at the embedded subject position]
b. Here is a filler [which we interpret _ as the object in the bracketed clause]

Subject relatives (SRs) are preferred over object relatives (ORs), as was shown with various measures.[19] They are read faster (King & Just 1991; Traxler, Morris, & Seely, 2002; Staub, 2010), recalled better (Baird & Koslick, 1974), found to be less taxing in dual task paradigms (Ford, 1983; Wanner & Maratsos 1978), acquired earlier (de Villiers, Flusberg, Hakuta, & Cohen, 1979; Tavakolian 1981; O'Grady, 1997; Diessel & Tomasello 2005; among others), and retained at a higher rate in language disorders (Caplan & Futter, 1986; Grodzinsky, 1989; Hickok, Zurif, & Canseco-Gonzalez, 1993). This asymmetry holds across different languages (Dutch - Mak, Vonk, & Schriefers, 2002; German - Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; French - Holmes & O'Regan, 1981; Spanish – Betancort, Carreiras, & Sturt, 2009; Japanese – Miyamoto & Nakamura, 2003; Korean - Kwon, Lee, Gordon, Kluender, Polinsky, 2010). For Hebrew, the asymmetry was tested mainly in language acquisition research (Arnon, 2010; Friedmann & Novogrodsky 2004; Friedmann, Belletti, & Rizzi, 2009).

The general ease of processing SRs in comparison to ORs can be attributed to various factors. The different accounts include costs of the distant filler-gap association, frequency considerations, and overall structure complexity. In languages where the filler precedes the embedded clause, and word order adheres the subject-verb-object scheme (like English and Hebrew), fewer constituents intervene between the filler and the gap in SRs than in ORs. The

[19] Yet it should be mentioned that several factors modulate this subject/object relative asymmetry. These factors include the animacy of the filler noun (Gennari & MacDonald, 2008; Traxler, Williams, Blozis, & Morris, 2005), the referential type of intervening elements (Gordon et al., 2004; Reali & Christiansen, 2007), the verb type (Gennari & MacDonald, 2009), the relative order of the relative clause and its head (Wagers, Borja, & Chung, 2018), and word order within the relative clause (Levy et al., 2013).

additional distance in ORs can interfere with processing in different ways. Thus, the additional processing load of ORs could be attributed to longer maintenance of the filler, or to more challenging retrieval or integration at the gap site. A filler requires a gap position for interpretation and syntactic licensing. Thus, the parser might hold information about the filler actively or sustain the expectation of the gap (Wanner & Maratsos, 1978; Frazier, 1987a; Just and Carpenter, 1992). In ORs, increased distance between the filler and the gap site would entail prolonged maintenance costs (Wanner & Maratsos, 1978), and a higher number of computations which occur simultaneously to the maintenance process (Hawkins, 1999). Alternatively, on the retrieval view, the filler has to be re-activated at the gap site for semantic integration and syntactic licensing. Accessing the filler in memory would be more straining in ORs due to interference from the subject of the relative clause (Gibson, 2000; Lewis & Vasishth, 2005).

On the other hand, experience-based accounts suggest that the preference for SRs over ORs should be attributed to prior language exposure. This idea could be framed in various ways. First, SRs are more frequent that ORs. This observation comes from various corpus studies of English (Fox & Thompson, 1990; Reali & Christiansen, 2007; Roland, Dick, & Elman, 2007), and applies also to other languages, including Hebrew (Arnon, 2010). Therefore, the SR reading would be favored in early stages of interpretation (Mitchell, Cuetos, Corley, & Brysbaert, 1995; Traxler, Morris, & Seely, 2002) and its processing would encompass less surprisal than that of ORs, as it fits better with probabilistic expectations (Levy, 2008a). Another way in which frequencies could bias the parser towards SRs is entropy reduction (Hale, 2006). Comprehenders may face less ambiguity and uncertainty when reading a SR, since it reveals the relativization site immediately. When the relative clause starts with a lexical subject, more possible relativization sites need to be entertained (and accordingly more possible thematic relations, Gennari & MacDonald, 2008). Lastly, expectation can also be formulated in terms of similarity to the canonical word order patterns in main clauses (MacDonald & Christiansen, 2002): In English (and Hebrew), SRs maintain the surface form of subject-verb-object word order, while ORs present a constituent order different from the common one.

Another family of accounts for the SR/OR asymmetry focuses on the syntactic complexity of their final interpretation. Structural theories (Keenan & Comrie, 1977; O'Grady, 1997) propose that SRs are easier to comprehend due to greater accessibility of the subject extraction site: The subject position is hierarchically higher in the sentence structure, and presents a more obligatory constituent.

Finally, it could be that more than one of these factors contributes to the SR preference. Staub (2010) presented an eyetracking study focusing on the timing of processing difficulty in ORs. He demonstrated two distinct behavioral patterns in object relatives, one consistent with the violation of expectations and one with memory retrieval. This may suggest that both factors are at play, working together rather than providing mutually exclusive explanations. Similarly, Levy, Fedorenko and Gibson (2011), found support for both memory-based and expectation-based theories in the processing of Russian relative clause, where the flexible word order allows to disentangle predictions of the competing theories. Recently, Pizzaro-Guerva (2020) observed a response time advantage in favor of SRs across various structures in Tagalog, which

is unpredicted by expectation-based and memory-based accounts. Therefore, while the overall pattern of his results rules out a purely structural account, Pizzaro-Guerva argues that structural complexity should to be entertained in any account mixture.

Either of the above hypotheses (or any conjunction of them) predicts that in Hebrew, like in English and many other languages, SRs would be preferred over ORs and cases with more embedded relativization sites. In the next subsection I describe how the grammatical properties of Hebrew may allow some ambiguity in verb-initial relative clauses. This ambiguity will be utilized in the current study to test whether, and under what circumstances, the preference for SRs can motivate readers to disregard other grammatical information.

5.1.2 The occasional ambiguity in Hebrew relative clauses

The default word order in Hebrew is SVO (Ravid, 1977; Berman, 1980). However, there are some cases where verb-initial sentences are licensed. First, subject-verb inversion is possible in certain registers (mostly in formal or literary contexts). In such cases (40), the subject would follow the verb (and usually a clause-initial prepositional phrase). This inversion could be similar to the locative or directive inversion in English (41), and is associated with the subject being discourse new (for English - Birner, 1996; for Hebrew – Maschler, 2015). However, this structure probably appears slightly more freely in Hebrew (Kuzar, 2012).[20]

(40) *ba-dugma ha-ba'a,* ***mofia*** *lefeta ha-po'al lifney ha-nose*
In.the-example the-next, **appears** suddenly the-verb before the-subject

(41) On a gathering storm / **comes** a tall handsome man
From the song "Red Right Hand" by Nick Cave and the Bad Seeds

Second, Hebrew allows various sentence constructions where an overt grammatical subject is lacking (Berman, 1980). The most productive among these subjectless constructions, are third-person plural impersonals (42). These constructions are similar to passives in their usage as they allow the speaker not to specify who was the agent of the event or action. These impersonals can also function in generic claims (like *In Paris, they eat frogs*) where the predicate is meant as a group characteristic (see Siewierska & Papastathi, 2011, for an elaborate review of third-person plural impersonals and their usage across different languages).

(42) **mistaklim** alay kol ha-zman / ma hem rocim / ma ixpat li be-e'ecem
look.PL on-me all the-time / what they want / what cares me actually
'People look at me all the time / what do they want / actually why do I care'
From the song "kol ha-kesem" ('All the magic') by Nimrod Lev

These Hebrew structures which allow verb-initial clauses can create local ambiguity between subject and object relatives (see example 43). Namely, a relative clause starting with a verb could be interpreted either as a canonical subject relative (43a), or as an object relative whose subject was removed: either omitted as an impersonal (43b), or dislocated in an inversion (43c). At the verb, we can eliminate the option of an impersonal null subject only if the verb is not in

[20] Subject-verb inversion is also possible and more common with specific verb types like unaccusatives and possessive *yeš* constructions.

the correct form, that is when the verb is singular. The option of a post-verbal subject construction would be degraded when there is no initial prepositional phrase, which supposedly "triggers" the inversion (Shlonsky & Doron, 1992).

(43) *ha-yeladim she-be-sof ha-yom, mac'u…*
The-kids that-in-end the-day, found.PL

a. **Subject relative:**
ha-yeladim$_i$ *she-be-sof ha-yom,* __$_i$ *mac'u et ha-derex habayta*
The-kids$_i$ that-in-end the-day, __$_i$ found.PL ACC the-way home
'The kids who, at the end of the day, found the way home'

b. **Object relative with an impersonal null subject:**
ha-yeladim$_i$ *she-be-sof ha-yom, Ø mac'u* __$_i$ *be-ezrat kol toshvey ha-shxuna*
The-kids$_i$ that-in-end the-day, Ø found.PL __$_i$ in-help all residents the-neighbourhood
'The kids who, at the end of the day, were found thanks to the neighbourhood residents'

c. **Object relative with a post-verbal subject:**
ha-yeladim$_i$ *she-be-sof ha-yom, mac'u ha-horim* __$_i$ *yeshenim al ha-sapa*
The-kids$_i$ that-in-end the-day, found.PL the-parents __$_i$ sleeping on the-sofa
'The kids who, at the end of the day, the parents found sleeping on the-sofa'

Upon reading the verb, in sentences like (43), the initial interpretation should exhibit the SR/OR asymmetry. Ambiguous relative clauses, in Dutch for example, seem to be exhibit the SR preference. Thus, Dutch sentences which are fully ambiguous between object and subject relatives, are interpreted as SRs over 70% of the times (Frazier, 1987a). Similarly, in German, processing disruption arises when locally ambiguous wh-questions are disambiguated to the object extraction reading (Meng & Bader, 2000). Therefore, in cases like (43) I expect the SR reading to be preferred over either of the OR structures.

5.1.3 The current study: Prior probability of SR and OR analyses and noisy agreement

In this study, I use sentences which are temporarily ambiguous between subject and object relatives in order to test noisy-channel inference. I manipulate the agreement matching between the filler and the verb in order to test whether readers compromise subject-verb agreement (which is required for the SR reading) to avoid the less frequent OR structures. Namely, I test whether the parser acts upon a top-down preference for SRs over ORs, even when the verb's features do not licence grammatical subject-verb (i.e. filler-verb) agreement. If an SR analysis is preferred in such cases, this would present evidence for noisy-channel inference in incremental sentence processing.

Why should it be rational for the parser to compromise the grammaticality of subject-verb agreement? I suggest that agreement is prone to errors, and thus can provide high-likelihood competitors for the faithful parse. Agreement errors are frequent in everyday communication due to their orthographic, linguistic and psycholinguistic status. First, agreement marking on both verbs and nouns is usually established orthographically by manipulation of only one or two letters. This should make it a likely target of typing errors. Second, the agreement system is commonly acquired inaccurately by L2 learners. Non-native speakers may produce various

agreement errors in both speaking and writing, and these utterances form part of the everyday language input of native speakers. Finally, subject-verb agreement is malleable and systematically vulnerable to interference, in both production and comprehension (Bock & Miller, 1991; Eberhard, Cutting, & Bock, 2005; Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau & Phillips, 2009, among others; for evidence from Hebrew see - Dank & Deutsch, 2010; Dank, Deutsch, & Bock, 2015; Deutsch & Dank, 2009; 2011). These should make agreement errors relatively frequent in the input of the parser, thus is should make sense for comprehenders to adapt the weight they give to agreement in formation in parsing accordingly. Given the considerations above, in choosing between a faithful, rare structure and a slightly noisy but more common one, a rational inference framework predicts that a noisy SR would be formed. Namely, readers may dismiss the agreement mismatch as an error, to avoid constructing a low probability structure.

In this context, it should also be noted while testing the SR preference in Dutch, Frazier (1987a) found a lingering footprint of the SR bias even when agreement was supposed to disambiguate the structure (44). When the verb did not agree with the filler (pointing to the OR reading, as in 44b), readers identified the filler as the subject of the relative clause in 31% of the cases. The opposite pattern of errors was much less likely - only 3.7% of the trials were analyzed as an OR when agreement suggested otherwise (in sentences like 44a). This provides evidence that readers may compromise a mismatch between a filler and a verb in order to construct an unmarked SR.

(44) *Ik schreef aan de vriend die mijn tantes ...*
I wrote to the friend who my aunts ...

a. **Subject relative:**
Ik schreef aan de vriend$_1$ [die __$_1$ mijn tantes heeft bezocht].
I wrote to the friend$_1$ [who __$_1$ my aunts has visited].
'I wrote to the friend who has visited my aunts.'

b. **Object relative:**
Ik schreef aan de vriend$_1$ [die mijn tantes __$_1$ hebben bezocht].
I wrote to the friend$_1$ [who my aunts __$_1$ have visited].
'I wrote to the friend who my aunts have visited.'

Interestingly, manipulating the availability of the two OR strategies in Hebrew, the post-verbal subject and impersonal subject, will allow us to test the willingness to compromise agreement when different prior probabilities are at stake: Post-verbal subjects are rare and marked, while impersonal null subjects are frequent (though presumably not as frequent as simple SRs). Opting for the noisy SR interpretation could be rational only if the alternative, grammatical OR analysis has a very low prior probability. Thus, I test whether ORs with low and intermediate prior probability produce the same noisy-channel inference.

Full data on the frequency of the different structures in Hebrew are missing, since corpus studies of Hebrew, like experimental ones, are relatively rare. However, there are a few observations on this matter. Maschler (2015) found only 57 tokens of post-verbal subjects in a Hebrew corpus which comprises approximately 6.5 hours of conversation among 396 different

speakers. In relative clauses the proportion of post-verbal subjects might be even smaller, as Givón (1976) suggests that the post-verbal subject strategy is suppressed in such environments for complexity considerations. In contrast, impersonal null subjects are frequent in different registers and styles of Israeli Hebrew (Taube, 2007), including both written and colloquial language. For example, Berman (2011) identifies such uses "in nearly every transcript" in child-directed speech.

On the model of rational noisy-channel inference, such differences in the prior probability of the pristine OR structure analysis should affect readers' tendency to consider the corrupt SR interpretation and to ignore the agreement mismatch. If readers take these probabilities into account, we should observe a stronger bias for a corrupted SR analysis, when the grammatical alternative is an OR with a post-verbal subject. On the other hand, when they have the option of impersonal null subjects, comprehenders might abandon the corrupted SR interpretation.

Thus, in Experiment 8, I use cases where the only faithful analysis is that of a OR with the rare post-verbal subject. In Experiment 9, I test cases where the more common OR interpretation (with an impersonal null subject analysis) is available. Finally, the two structures are directly contrasted in Experiment 10, which also examines comprehenders' interpretations more directly, using a sentence completion paradigm. To disclose the results early, the manipulation of prior probability indeed affects how readers treat the agreement mismatch. Experiments 8A-B exhibit that, faced with the choice between noisy SR and an OR with a rare word order, readers opt for the SR analysis even though it includes an agreement mismatch. However, when the OR option does not involve a syntactic structure of very low probability (Experiments 9A-B), the grammatical OR analysis is preferred over a noisy SR. This preference pattern is also mirrored in production data of Experiment 10.

5.2 Experiment 8: Choosing a noisy structure over an extremely rare one

As explained above, Hebrew licenses post-verbal subjects (mostly in higher registers) when the clause begins with an adjunct or an adverbial phrase (Shlonsky, 1990; Shlonsky & Doron, 1992). Therefore, when a relative clause (with a singular head) begins with an adverbial, followed by a (singular) verb, the structure is ambiguous between a SR and an OR with a post-verbal subject (exemplified again in 45 for convenience).

(45) *ha-yeled she-be-sof ha-yom, maca…*
The-child that-in-end the-day, found.SG

a. **Subject relative:**
ha-yeled$_i$ *she-be-sof ha-yom,* __$_i$ *maca et ha-derex habayta*
The-child$_i$ that-in-end the-day, __$_i$ found.SG ACC the-way home
'The child who, at the end of the day, found the way home'

b. **Object relative with a post verbal subject:**
ha-yeled$_i$ *she-be-sof ha-yom, maca ha-dod* __$_i$ *mul ha-televizia*
The-child$_i$ that-in-end the-day, found.SG the-uncle __$_i$ in.front.of the-TV
'The child who, at the end of the day, the uncle found in front of the TV'

Disambiguation arises at the first NP following the verb. If this is a definite NP preceded by an accusative case marker (*et*, as in 45a), or an NP that does not agree with the verb, this NP is the object, meaning that the clause is a subject relative clause.[21] If the post-verbal NP is a definite NP which is not preceded by an accusative case marker nor by a preposition, and it agrees with the verb, this NP is the subject, meaning that the clause is an object relative clause, with a post-verbal subject.

It is most likely that, faced with this local ambiguity in, comprehenders would initially prefer the SR reading. Therefore, a subject NP following the verb (e.g. *the uncle* in 45b), would be inconsistent with the initial parsing preferences and should give rise to a processing difficulty. The two experiments reported in this subsection build on such a reanalysis effect as evidence for the SR interpretation. I test whether readers compromise agreement and endorse the SR structure, by measuring reading times on the post-verbal subject in OR clauses, when the verb matches the filler (as in 45b, confirming the basic SR bias) and when it mismatches it (as in 46).

(46) *ha-yeladim*$_i$ *she-be-sof ha-yom, maca* ***ha-dod*** __$_i$ *mul ha-televizia*
The-kids$_i$ that-in-end the-day, found.SG **the-uncle** __$_i$ in.front.of the-TV

If disruption on the post-verbal subject is evident when the verb mismatches the filler, this would suggest that readers overlook the agreement mismatch and form a SR, in line with Rational Noisy-Channel processing. However, if readers take the mismatch as indication that

[21] In Hebrew, the distinction between nominative and accusative case is not marked for indefinite nouns; thus, a post-verbal indefinite NP will preserve the OR/SR ambiguity, if it agrees with the verb in number, gender and person.

a subject gap is impossible, they should predict a post-verbal subject. In that case, reading times on the post-verbal subject should align with those in a baseline condition including no dependency (see below), as in both cases it should be consistent with the parser's prediction.

5.2.1 Experiment 8A

5.2.1.1 Methods

Participants

Participants were 52 native speakers of Hebrew (according to self-report) (mean age: 24.18, range: 18-35). Nine participants were bilingual of Hebrew and either Russian or English, and the rest were monolingual.

Materials

The experiment included 30 item sets of three conditions: temporarily ambiguous relative clauses where the filler matched or mismatched the embedded verb, and a baseline condition (see example set in Table 5.1).

The experimental conditions were constructed such that at the verb, relative clauses can be interpreted as ORs with a post-verbal subject but not at ORs with an impersonal null subject. Thus, the embedded clauses started with a preposition phrase which supposedly allows the subject-verb inversion (Shlonsky & Doron, 1992). The verb was always in the singular form to eliminate the possibility of the impersonal subject reading. In the Match condition, the filler and verb were therefore both singular. In the Mismatch condition, the verb was singular, but the filler was a noun phrase marked as plural.[22]

Following the verb, there appeared a post-verbal subject. This NP, filling the assumed subject gap, should give rise to processing costs (prolonged RTs) if a SR was constructed, as the structure should be reanalyzed to accommodate a post-verbal subject instead of the subject gap. To avoid contamination of the critical region (the post-verbal NP) from the verb, I used an adverb as a buffer between the verb and the subject. To make sure that the NP following the verb is interpreted as the subject, I used definite NPs, which agreed with the features on the preceding verb, and were not preceded by the accusative case marker. The relative clauses in the experimental items were ultimately resolved with an accusative resumptive pronoun (grammatical in Hebrew) within the clausal complement of the verb. Thus, all the sentences in the experiment were eventually grammatical.

As the experimental design aims to trace increased processing difficulty at the post-verbal subject, I used sentences with an unambiguous post-verbal subject as the baseline condition. This choice of baseline was designed to allow relatively similar word order in the experimental

[22] I assume that if readers edit the agreement information on the verb to plural form, they do so to construct the subject relative reading, rather than the object relative one with an impersonal null subject. See results of Experiments 9 and 10 for evidence that readers do not prefer the impersonal object relative reading when the verbal agreement allows both grammatical subject relative analysis and an impersonal subject analysis.

and baseline conditions, in line with the typical baseline conditions in the ambiguity resolution (and filled-gap effects) literature.[23] To create this unambiguous baseline, I used adjunct clauses, without a filler-gap dependency. In such cases, since there is no option to assume a gap or any subject before the occurrence of the verb, a post-verbal subject is predicted, and should not disrupt processing. Since, in the baseline sentences, subject-verb inversion must be assumed upon reading the verb, if a SR is initially constructed in the relative clause cases (either Match or Mismatch conditions), the reading of the post-verbal subject should be slower in relative clauses relative to the reading times in the baseline condition.

Condition	Sentence
Match (Exp. 1A&B)	*avarnu leyad ha-pakax$_1$ še-be-mahalax ha-pkak* we+passed near the-inspector-M.SG$_1$ that-during the-traffic jam *himšix lixora **ha-nahag** ha-alim litkof oto$_1$* continued-3.M.SG allegedly **the-driver**.M.SG the-violent to-attack him$_1$ *bli šum buša* without any shame. 'We went past the inspector who the violent driver continued to attack shamelessly, during the traffic jam'.
Mismatch (one feature) (Exp. 1A&B)	*avarnu leyad ha-pakaxim$_1$ še-be-mahalax ha-pkak* we+passed near the-inspectors-M.PL$_1$ that-during the-traffic jam *himšix lixora **ha-nahag** ha-alim litkof otam$_1$* continued-3.M.SG allegedly **the-driver**.M.SG the-violent to-attack them-M$_1$ *bli šum buša.* without any shame. 'We went past the inspectors who the violent driver continued to attack shamelessly, during the traffic jam'.
Baseline (Exp. 1A&B)	*avarnu bezman še-be-mahalax ha-pkak himšix* we+passed while that-during the-traffic.jam continued-3.M.SG *lixora **ha-nahag** ha-alim litkof et ha-pakax bli* allegedly **the-driver**.M.SG the-violent to-attack ACC the-inspector without *šum buša.* any shame.

[23] Note that using a subject relative as the baseline here would not have provided a parallel region to compare against for the critical words. In addition, Gennari and MacDonald (2008, p. 162-163) present an argument against choosing subject relatives as a baseline for comparison for comprehension difficulty in object relatives. They suggest that subject and object relatives may use the same lexical content and are identical in that they include a filler gap dependency, "but the two sentence types differ in their word order and in final sentence meaning. This choice contrasts with typical baseline conditions in ambiguity resolution studies in which unambiguous structures are chosen to have similar meanings to those in ambiguous conditions, though the sentences may have minor differences in number of words or word choices."

	'We went past while the violent driver continued to attack the inspector shamelessly, during the traffic jam'.
Two-feature mismatch (Exp. 1B only)	*avarnu leyad ha-pakaxiyot$_1$ še-be-mahalax ha-pkak* we.passed near the-inspectors-F.PL$_1$ that-during the-traffic jam *himšix lixora **ha-nahag** ha-alim litkof otan$_1$* continued-3.M.SG allegedly **the-driver**.M.SG the-violent to-attack them-F$_1$ *bli šum buša.* without any shame. 'We went passed the inspectors who the violent driver continued to attack shamelessly, during the traffic jam'.

Table 5.1. Example set from Experiments 8A & 1B.
The critical NP is marked in bold; ACC = accusative case marker; F = feminine grammatical gender; M = masculine grammatical gender; SG = singular; PL = plural.

The experimental items were distributed in a Latin square design across three lists. Within each list, the target sentences were combined with 45 grammatical filler sentences, for a total of 75 sentences. The filler items included 10 sentences with no filler-gap dependency, 15 unambiguous subject relatives, and 20 long filler-gap dependencies (10 object relatives and 10 relative clauses with an embedded subject resolution).

Procedure

This experiment ran as a lab-based self-paced reading experiment. Seventy-five percent of the trials were followed by a 'yes/no' comprehension question.

Data analysis

Participants were excluded from analysis if their performance on comprehension questions of experimental items was lower than 70% (resulting in the removal of five participants), or if their average RT in experimental items was more than 2.5 SDs above the group's average (resulting in the removal of two additional participants). For the remaining 45 participants, RTs higher than 2.5 SDs above the individual's average RT were trimmed to that cutoff (affecting 2.6% of the data). RTs shorter than 120ms were excluded (affecting 0.03% of the data).

I analyzed RTs from the critical region (the post-verbal subject noun), as well as from the spillover word (the adjective following the post-verbal noun).[24] I used a treatment (dummy) coding scheme to detect an increase in RT for each condition relative to the baseline condition. Namely, for every region, two comparisons were made: Match vs. Baseline and Mismatch vs. Baseline. In addition, due to an unpredicted contrast between the pattern of the effects at the critical and the spillover region, I fit another model, using the data of both regions, to test for main effects over both regions and for the interaction between region and condition. In this

[24] I did not analyze the verb region, since I believe there is no strong prediction for the RT pattern there. Under rational noisy-channel processing, the agreement mismatch should be detected at the verb. Yet the low probability of the grammatical alternative pushes the parser to the SR reading. Thus, it could be that the mismatch induces processing disruption or dissolves without a prominent increase in RT.

model, treatment (dummy) coding was used for the condition factor and sum coding was used for the region factor. In the unified model, four pairwise comparisons were conducted, comparing the critical conditions (Match and Mismatch) with the baseline in each position (critical and spillover word). Thus, the reported *p*-values for this model are corrected for four comparisons.

5.2.1.2 Results

Word-by-word reading times of the different experimental conditions and by-condition numerical means of the tested regions are presented in Figure 5.1. The results of the frequentist and Bayesian analyses are summarized in Table 5.2 and 5.3, correspondingly.

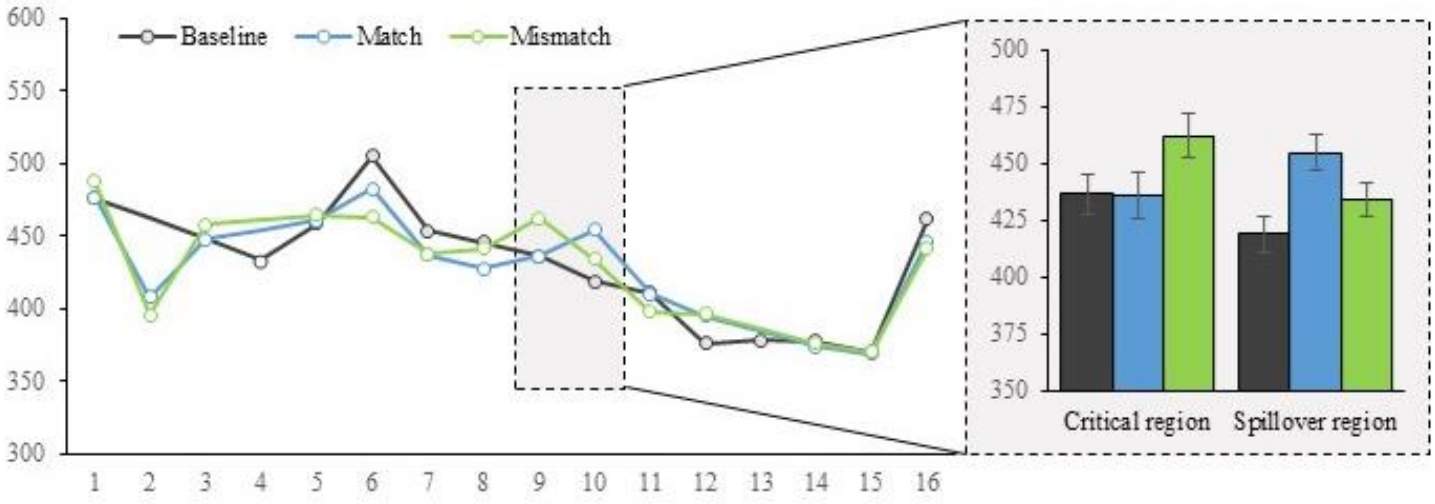

Figure 5.1. Word-by-word RT means (ms) by condition, Experiment 8A. Error bars represent +/-SE.

Region mapping: |We passed|$_1$ |near|$_2$ |the inspector/s|$_3$ (|while|$_4$) |that during|$_5$ |the traffic jam|$_6$ |continued|$_7$ |allegedly|$_8$ |the driver|$_9$ |the violent|$_{10}$ |to attack|$_{11}$ |ACC|$_{12}$ |the inspector|$_{13}$ |without|$_{14}$ |any|$_{15}$ |shame|$_{16}$
The critical region is shaded gray.

I observed an unexpected contrast in the timing of the effects, when modeling the critical and the spillover regions separately (this can also be seen by visual inspection of the results): processing difficulty in the Match condition only arises at the spillover region, whereas difficulty arises already in the critical region for the Mismatch condition. In order to validate the contrast in the timing of the effects, I conducted another analysis using data from both regions. This analysis tested for an interaction between experimental condition and sentential position. Since results of both versions of the analysis were compatible, I report here only the results of the unified model (results of the by-region models are available in Appendix C).

The analyses detected a main effect for the Mismatch condition, namely a general increase in RT in these sentences, relative to baseline. The models also detected an interaction between region and condition, for the Match-Baseline contrast, but not for the Mismatch-baseline contrast. This suggests a differential pattern in the critical and post-critical positions. Pairwise contrasts revealed a reliable increase in reading times at the critical region of the Mismatch condition, and at spillover region of the Match condition (relative to the Baseline condition).

Two additional effects were detected yet were not reliable on both analyses. On the frequentist analysis there was a significant main effect for the overall comparison of the Match and Baseline conditions (collapsing over both regions). This failed to reach the required level on the Bayesian analysis (with 94% of the posterior distribution beyond zero). On the other hand, the Bayesian analysis detected an effect of region in the model. Given the contrast coding, this should reflect a comparison of the critical and the post-critical words on the baseline condition. However, this effect is not of theoretical significance for the current study, and was not mirrored in the frequentist analysis.

	Estimate	SE	*t*	*p*
Main effects and interactions:				
Main effect of Match - Baseline	0.054	0.015	3.70	< .001
Main effect of Mismatch - Baseline	0.053	0.014	3.85	< .001
Main effect of Region	0.006	0.013	0.48	.631
Interaction Match:Region	-0.043	0.016	2.70	.011
Interaction Mismatch:Region	0.011	0.018	0.62	.538
Pairwise comparisons:				
Match - Baseline, critical region	0.011	0.021	0.53	> .99
Mismatch - Baseline, critical region	0.064	0.023	2.72	.026
Match - Baseline, spillover region	0.096	0.022	4.36	< .001
Mismatch - Baseline, spillover region	0.042	0.022	1.93	.214

Table 5.2. Results of the frequentist analysis of Experiment 8A: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Main effects and interactions:		
Main effect of Match – Baseline	10 [-3, 24]	94%
Main effect of Mismatch – Baseline	14 [1, 27]	98%
Main effect of Region	8 [0, 16]	97%
Interaction Match:Region	-17 [-28, -6]	> 99%
Interaction Mismatch:Region	3 [-10, 16]	67%
Pairwise comparisons:		
Match - Baseline, critical region	-7 [-24, 10]	80%
Mismatch - Baseline, critical region	17 [-2, 36]	96%
Match - Baseline, spillover region	29 [11, 47]	> 99%
Mismatch - Baseline, spillover region	11 [-7, 30]	89%

Table 5.3. Results of the Bayesian analysis of Experiment 8A: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

5.2.1.3 Discussion

The results exhibit increased processing costs at the disambiguating post-verbal subject both in the Match and Mismatch conditions, relative to an unambiguous subject-inversion structure. I interpret these increased RTs as reanalysis costs, required if readers constructed a SR at the verb. The results suggest that readers tend to disregard the mismatch between the filler and the verb and construct a SR in order to avoid the (possible yet rare) subject-verb inversion structure. Note that the tendency to dismiss the agreement mismatch cannot be attributed to a strategy developed throughout the experiment. Since all experimental materials and filler items were grammatical, there should be no reason to expect ungrammaticality as part of the experiment.

It should be noted that while the reanalysis effect in the Match condition was shifted and was found only at the spillover region, in the Mismatch condition, reanalysis costs were evident on the post-verbal subject itself. This could suggest that the detection of the required reanalysis was faster in the mismatching condition relative to sentences with a matching verb. Since this result was unexpected, I conducted another experiment to replicate it.

5.2.2 Experiment 8B

In Experiment 8B I aimed to replicate the pattern of results obtained in Experiment 8A. In addition, I wanted to examine how robust the preference for discarding agreement is, by extending my manipulation to more pronounced filler-verb agreement mismatches. I thus added another condition, in which the filler and the verb were differentiated by the two features, with discrepancy in both gender and number (rather than number only).

5.2.2.1 Methods

Participants

Participants were 68 native speakers of Hebrew (according to self-report) (mean age: 25.48, range: 21-35). Three participants were bilingual of Hebrew and English, and the rest were monolingual.

Materials

The experiment included 28 sets out of the sets in Experiment 8A. The experiment comprised the three conditions of Experiment 8A, and an additional condition, the Two-feature mismatch condition, in which the filler and verb differed in both gender and number features: the filler was a feminine plural noun phrase, while the verb was inflected for masculine singular (see the last condition in Table 5.1). Items were distributed into four lists in a Latin square design. Filler items were also the same as in Experiment 8A.

Procedure

The procedure was the same as in Experiment 8A.

Data analysis

Data analysis followed the same protocol as in Experiment 8A, only this time I fit the data directly with the unified model (testing for an interaction between region and condition). Six pairwise comparisons were conducted, comparing the critical conditions (Match, Mismatch, and 2-Mismatch) with the baseline in each position (critical and spillover word). Thus, the reported *p*-values for this model are corrected for six comparisons. Due to convergence failure in the frequentist analysis, the slopes were gradually removed from the random effect structure. The final model lacked one interaction term on by-subject random effects and included only one interaction term (with no main effects) on by-item random effects.

Exclusion criteria resulted in the removal of six participants (four for low accuracy in comprehension questions and two for abnormal reading times). For the remaining 60 participants, trimming of high and low RTs affected 2.6% and 0.06% of the data, correspondingly.

5.2.2.2 Results

Word-by-word reading times of the different experimental conditions and by-condition numerical means of the tested regions are presented in Figure 5.2. The results of the frequentist and Bayesian analyses are summarized in Table 5.4 and 5.5, correspondingly.

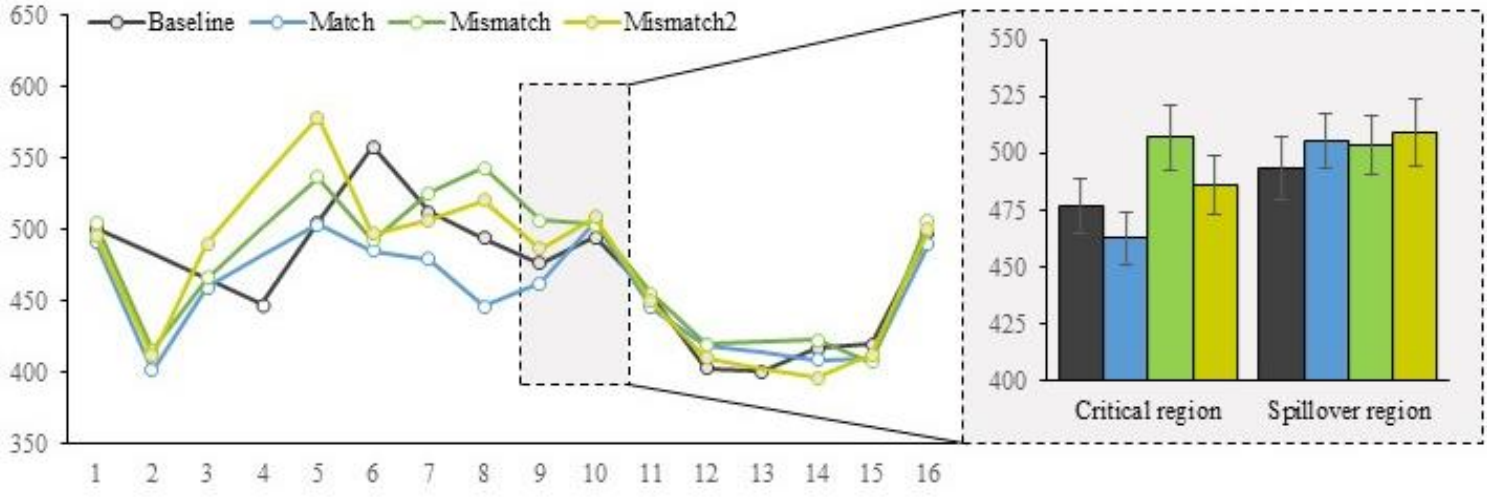

Figure 5.2. Word-by-word RT means and means in the tested regions (ms) by condition, Experiment 8B. Error bars represent +/-SE.

Region mapping: |We passed|$_1$ |near|$_2$ |the inspector/s|$_3$ (|while|$_4$) |that during|$_5$ |the traffic jam|$_6$ |continued|$_7$ |allegedly|$_8$ |the driver|$_9$ |the violent|$_{10}$ |to attack|$_{11}$ |acc.|$_{12}$ |the inspector|$_{13}$ |without|$_{14}$ |any|$_{15}$ |shame|$_{16}$

The critical region is shaded gray.

The analyses detected a main effect for the Mismatch condition, namely a general increase in RT in these sentences, relative to baseline. The models also detected an interaction between region and condition, for the Match-Baseline contrast, but not for the Mismatch-baseline contrast. This suggests a differential pattern in the critical and post-critical positions. Pairwise contrasts revealed a reliable increase in reading times at spillover region of the Match condition (relative to the Baseline condition).

The contrast between the Mismatch and the baseline conditions, at the critical region specifically (observed in the pairwise comparisons of Experiment 8A), was detected only in

the Bayesian analysis this time. However, since there were increased RTs in this condition when collapsing over both regions (the main effect of Mismatch which both analyses detected, as mentioned above), I take the results to generally indicate the expected reanalysis costs.

Overall, no reliable effects were observed for the 2-feature mismatch condition. The pairwise comparisons for that condition did not detect an effect on either analysis. The frequentist analysis reflected a main effect, suggesting increased RTs for that condition relative to baseline when collapsing over both regions. However, the Bayesian analysis did not detect a comparable effect. Several additional effects appeared only in the frequentist analyses, namely a main effect of the Match condition and a main effect of the region factor.

	Estimate	SE	*t*	*p*
Main effects and interactions:				
Main effect of Match - Baseline	0.035	0.016	2.22	.028
Main effect of Mismatch - Baseline	0.045	0.016	2.83	.005
Main effect of 2-Mismatch - Baseline	0.039	0.015	2.55	.011
Main effect of Region	-0.026	0.012	-2.20	.028
Interaction Match:Region	-0.036	0.016	-2.27	.024
Interaction Mismatch:Region	0.007	0.017	0.40	.689
Interaction Mismatch2:Region	-0.006	0.015	-0.37	.715
Pairwise comparisons:				
Match - Baseline, critical region	-0.001	0.022	-0.06	> .99
Mismatch - Baseline, critical region	0.052	0.023	2.25	.145
Mismatch2 - Baseline, critical region	0.033	0.022	1.54	.735
Match - Baseline, spillover region	0.070	0.022	3.16	.009
Mismatch - Baseline, spillover region	0.038	0.023	1.68	.559
Mismatch2 - Baseline, spillover region	0.045	0.022	2.06	.235

Table 5.4. Results of the frequentist analysis of Experiment 8B: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Main effects and interactions:		
Main effect of Match – Baseline	2 [-14, 17]	58%
Main effect of Mismatch – Baseline	15 [-1, 32]	97%
Main effect of Mismatch2 – Baseline	10 [-6, 26]	88%
Main effect of Region	-4 [-11, 4]	82%
Interaction Match:Region	-17 [-30, -3]	> 99%
Interaction Mismatch:Region	4 [-10, 18]	71%
Interaction Mismatch2:Region	-2 [-15, 12]	58%
Pairwise comparisons:		

Match - Baseline, critical region	-15 [-35, 5]	93%
Mismatch - Baseline, critical region	19 [-2, 41]	96%
Mismatch2 - Baseline, critical region	8 [-13, 30]	78%
Match - Baseline, spillover region	19 [-2, 41]	96%
Mismatch - Baseline, spillover region	11 [-10, 33]	85%
Mismatch2 - Baseline, spillover region	11 [-10, 33]	84%

Table 5.5. Results of the Bayesian analysis of Experiment 8B: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

5.2.2.3 Discussion

Results of this experiment replicated those of Experiment 8A, in that we observed increased RTs in the Mismatch condition (see Figure 5.3 for a comparison of the effects in both experiments). These results suggest that when the relative clause is ambiguous between a corrupted SR and a OR with a post-verbal subject, readers prefer constructing a noisy SR. In addition, Experiment 8B replicated the temporal pattern observed in Experiment 8A, such that in the Match condition, the reanalysis effect was only observed in the post-critical position. In this experiment, the effect in the Mismatch condition was not detected in one specific region but seemed to be spread over the two regions. I suggest that this pattern may reflect earlier detection of the required reanalysis in the Mismatch condition relative to the Match condition, and/or an easier reanalysis in this condition, as a result of the earlier origin of uncertainty (or lower stability of the representation, in terms of dynamical self-organizing parsers). Namely, I propose that readers maintain some uncertainty regarding the correct structure after encountering the mismatching verb in the Mismatch condition. This in turn facilitates the reanalysis procedure at the post-verbal subject. This is in line with findings from the processing of filler-gap dependencies. For instance, Pickering and Traxler (1998, 2003) suggest that semantic implausibility of the filler-verb association reduces the parser's commitment to the dependency and thus lowers reanalysis costs.

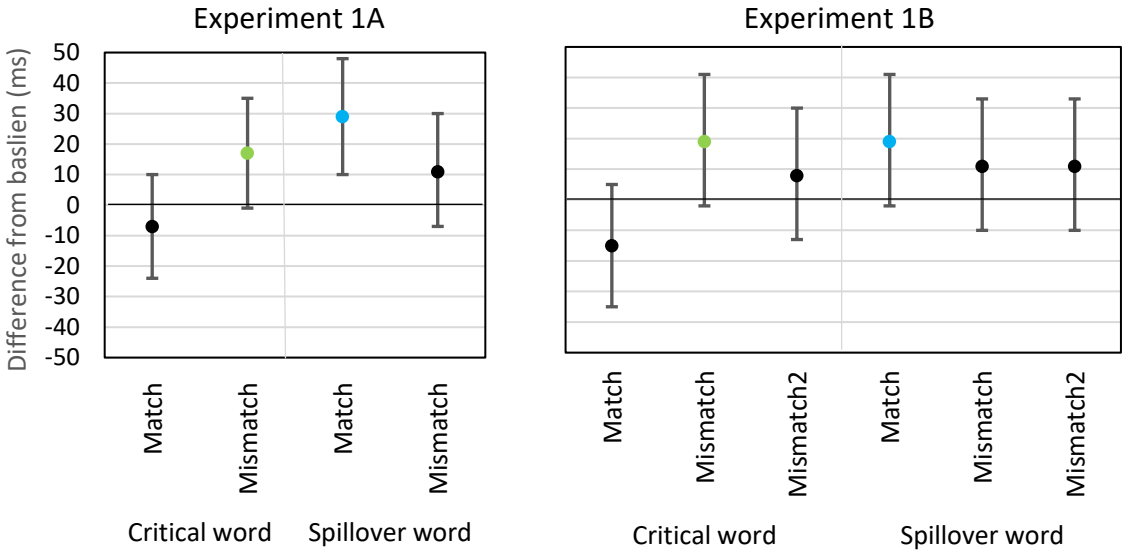

Figure 5.3. Results of the Bayesian analysis of Experiments 8A&B: The posterior means of the tested contrasts with 95% CrI. The coloured points mark the critical contrasts.

For the Two-feature mismatch condition, results were inconclusive. We detected significant processing costs in the frequentist analysis, but these were not apparent in the Bayesian analysis. I therefore would like to refrain from speculating whether readers are willing to compromise a double-feature mismatch, or rather take the mismatch as an indication for the rare subject-verb inversion. It should be noted that in most sets, the additional mismatching feature did not affect the number of orthographic edits which are required for restoring agreement. Future research should examine more closely whether manipulating an abstract feature mismatch, without affecting the likelihood of low-level errors (in visual perception or in typing), can affect readers' parsing strategy. Ryskin and her colleagues (2018) recently suggested that the comprehender's noise model treats some types of errors as analogous and some as categorically distinct, possibly differentiating typing errors and planning errors. If indeed readers make inferences regarding the producer's planning errors, they should consider abstract/linguistic similarity, rather than only form similarity, when calculating the likelihood of different errors.

Overall, the results replicate the findings from Experiment 8A and suggest that readers are willing to compromise agreement in order to construct a structure of higher prior probability. These findings are in line with informal judgments provided in Givón (1976). In his review of the distribution of post-verbal subjects in Hebrew, Givón suggests that post-verbal subjects might be less common in ORs, due to the ambiguity they introduce. He states that when the filler mismatches the verb's features, there seems to be a "tendency" to interpret the sentence as a SR with an agreement error.

Yet, it could be suggested that negligence of agreement information does not reflect a rational processing strategy, but a heuristic of good-enough processing (Ferreira & Patson, 2007), which does not take into account elaborate probabilistic information. For example, English passives were suggested by Ferreira (2003) to be systematically misinterpreted, based on good-enough processing and an agent-first heuristic. It is possible that the tendency to interpret the relative clauses of Experiment 8 as SRs is based on shallow processing, general underspecification of agreement information, and/or consistent dismissal of agreement in favor of shorter filler-gap dependencies. In order to test this, in Experiment 9 I use another case where the SR/OR ambiguity might arise, namely sentences which license impersonal null subjects, where the prior probability of the OR analysis is not as low as that of post-verbal subjects.

5.3 Experiment 9: Choosing a slightly infrequent structure over a noisy one

In Experiment 8, I used the rare post-verbal subject structure to create temporary ambiguity in relative clauses. In Experiment 9, I use a different structure which can create ambiguity in Hebrew relative clauses: impersonal null subjects. In Hebrew, verbs with plural third person agreement licence empty subjects (Shlonsky, 1997), which are interpreted as referring to a generic (47), unknown, or immaterial entity (48) (see Berman, 2011, for the semantic interpretation of such impersonals).

(47) *ganvu et ha-ofana'im šeli etmol.*
stole-3.PL ACC the-bicycle mine yesterday.
'Someone stole my bicycle yesterday'.
(48) *be-avar, axlu yoter yerakot.*
in-past, ate-3.PL more vegetables.
'In the past, people ate more vegetables.

Therefore, in a relative clause with a plural head, a third person plural verb (with no previous subject), would have two possible readings, as in (49a-b). One possible analysis (49a) posits the gap in subject position (i.e. forming an SR). The second (49b) posits it in the object position (i.e. forming an OR) and assumes an impersonal null subject (marked here by *Ø*).

(49) *ele ha-yeladim* [*še-mac'u*…
These the-kids [that-found-3.PL …

a. **Subject relative:**
ele ha-yeladim$_1$ [*še-*__$_1$*-mac'u et ha-ofana'im šeli ha-boker*].
These the-kids$_1$ [that-__$_1$-found-3.PL ACC the-bicycle mine the-morning].
'These are the kids who found my bicycle this morning.'

b. **Object relative:**
ele ha-yeladim$_1$ [*še-Ø-mac'u* __$_1$ *im ha-ofana'im šeli ha-boker*].
These the-kids$_1$ [that-found-3.PL __$_1$ with the-bicycle mine the-morning].
'These are the kids who someone found with my bicycle this morning.'

As in Experiment 8, I test the parser's preference in the face of this ambiguity, manipulating the agreement features of the filler to allow or disallow subject-verb agreement, required under a SR reading. In Experiment 8, the low frequency of the alternative (object relative with a post-verbal subject) pushed readers towards adopting the SR reading, despite the agreement mismatch. However, I expect that the current experiments, using the plural verb form and therefore introducing the alternative impersonal OR analysis, will produce different results, due to the higher prior probability of this structure. As mentioned in subsection 5.1.3, while ORs with post-verbal subjects are exceptionally rare, impersonal ORs are quite frequent (although presumably still less probable than simple SRs).

If the prior probability of the alternative structure affects the interpretation of filler-verb agreement mismatch, the results of Experiment 9 might diverge from those of Experiment 8.

While SRs should still be favoured over ORs in general, the frequency of the OR with impersonal subjects might not be low enough to tip the balance and make the noisy SR analysis rank higher than the faithful OR analysis. In contrast, if rational noisy-channel processing is not behind the findings of Experiment 8, and agreement is categorically ignored in favour of any probable SR structure, since these are easier to process, the results of Experiment 9 should align with those of Experiment 8, even though an OR with an impersonal subject is not as rare as an OR with a post-verbal subject.

To assess whether the parser adopts an OR representation, I test for reanalysis costs when a lexically realized argument appears in the direct object position. If a SR analysis is constructed in the relative clause sentences (like in the previous experiments), an object NP is predicted and should not cause any difficulty. When both the filler and the verb are plural, this structure will present the basic SR/OR ambiguity exemplified in 49. If in this case the object NP will not entail processing costs, this would suggest that readers prefer the SR interpretation, when both options (SR and impersonal OR) are grammatical.

As for the case where the verb mismatches the agreement features of the filler, predictions are as follows: If it is the case that agreement mismatches are categorically ignored in favour of constructing an SR (due to general good-enough processing), then an object NP should not cause processing disruption in this condition as well (in line with the SR preference in the match condition). However, the framework of Noisy-Channel processing predicts another second pattern. When the relatively probable reading of an OR with an impersonal null subject is possible, rational readers should take the mismatch as indication that a subject gap is impossible and construct an OR. If readers construct an OR (using the null subject option), a realized direct object should disrupt processing, as a gap was posited in object position, and reanalysis is required (i.e. the "filled-gap" effect, cf. Stowe, 1986).

5.3.1 Experiment 9A

5.3.1.1 Methods

Participants

Participants were 69 native speakers of Hebrew (according to self-report) (mean age: 25.12, range: 18-35). Seven participants were bilingual of Hebrew and either Russian, English, or Spanish and the rest were monolingual.

Materials

The experiment included 24 experimental items of four conditions (see example set in Table 5.6): temporarily ambiguous relative clauses where the filler matched or mismatched the embedded plural verb, a classic "filled-gap" condition, and a baseline condition. The experimental conditions were constructed such that at the verb, the Match and Mismatch conditions can be interpreted as ORs with an impersonal null subject. Thus, the verb always

appeared in the plural form (in the Match condition, the filler and the verb were both plural; in the Mismatch condition, the verb was plural, but the filler was a singular noun phrase).

Following the verb, an object NP appeared, marked by the accusative marker. This lexical NP fills the potential object gap position. If an object relative was constructed, then encountering this NP would require reanalysis of the sentence structure. The dependency resolved grammatically later, with a possessive resumptive pronoun matching the filler's agreement.[25] In addition to the Match and Mismatch conditions, I included a "standard" filled-gap effect condition. In this condition, the filler was plural, but a first-person pronoun was cliticized to the verb, so a subject gap would not be considered.[26] Therefore, in this condition, readers are predicted to construct an object relative, and engage in reanalysis upon encountering the verb's object.

As the baseline condition, I use an adjunct clause with a realized subject (similarly to the baseline in Experiment 8, and in line with no-dependency baselines in the filled-gap paradigm – e.g. Stowe, 1986). In such an unambiguous baseline, no relative clause structure is involved, and thus no gap could be posited in the verb's object position at any point. The allows a measure of reading times at the object NP, against which we can compare the predicted processing disruption of the other conditions.

Condition	Sentence
Match	*ha-saparit dibra im ha-me'acvim*$_1$ *še- baxru larov* the-hairdresser talked with the-designers-M.PL$_1$ that-chose-3.PL mostly ***et ha-dugmaniyot*** *ha-rašiyot šelahem*$_1$ *ba-rega ha-axaron.* **ACC the-models** the-main of.them-M$_1$ in.the-moment the-last. 'The hairdresser talked with the designers who usually chose their leading models at the last moment.'
Mismatch	*ha-saparit dibra im ha-me'acev*$_1$ *še- baxru larov* the-hairdresser talked with the-designer-M.SG$_1$ that- chose-3.PL mostly ***et ha-dugmaniyot*** *ha-rašiyot šelo*$_1$ *ba-rega ha-axaron.* **ACC the-models** the-main of.him$_1$ in.the-moment the-last. 'The hairdresser talked with the designer whose leading models were usually chosen at the last moment.'
Filled-gap effect	*ha-saparit dibra im ha-me'acvim*$_1$ *še-baxarnu larov* the-hairdresser talked with the-designers-M.PL$_1$ that-we.chose mostly ***et ha-dugmaniyot*** *ha-rašiyot šelahem*$_1$ *ba-rega ha-axaron.*

[25] In the Match condition the resolution is ambiguous between a SR, with grammatical agreement between the filler and the verb (the more likely interpretation), and an analysis with an impersonal null subject and a possessive resumptive pronoun (as in the Mismatch and FGE conditions).

[26] In principle, it would be possible for readers to "edit" the verb form in this condition as well, changing it to the third person plural form in order to adopt a subject relative analysis. However, we assume that readers are much less likely to do so in the FGE condition, relative to the Mismatch condition, since in the latter, a SR can be established by merely assuming misreading of the filler.

	ACC the-models the-main of.them-M_1 in.the-moment the-last. 'The hairdresser talked with the designer whose leading models we usually chose at the last moment'
Baseline	*ha-saparit dibra im ha-me'acvim*$_1$ *mikeyvan še-hem*$_1$ the-hairdresser talked with the-designers-M.PL$_1$ because that-they-M *baxru larov* ***et ha-dugmaniyot*** *ha-rašiyot šelahem*$_1$ *ba-rega ha-axaron.* chose mostly **ACC the-models** the-main of.them-M_1 in.the-moment the-last. 'The hairdresser talked with the designers because they usually chose their leading models at the last moment'

Table 5.6. Example set from Experiment 9A.
The critical NP is marked in bold; ACC = accusative case marker. PL = plural.

The experimental items were distributed in a Latin square design across four lists. Within each list, the target sentences were combined with 51 grammatical filler sentences, for a total of 75 sentences. The filler items included 27 sentences with no relative clauses, 12 unambiguous SRs with a singular head and verb, and 12 unambiguous ORs. It should be noted that in this experiment, the final structure unambiguously contained impersonal null subjects in the Mismatch condition (the Match condition is globally ambiguous between a SR and an OR with an impersonal subject, see footnote 24). Thus, each experimental list included only 6-12 instances of the tested structure, while in Experiments 8A-B all experimental sentences ultimately contained the tested structure (a post-verbal subject). To minimize between-experiments differences in adaptation to the tested structures, I included 10 globally unambiguous impersonal subjects within the filler items of this experiment.

Procedure

The procedure was the same as in the previous experiments.

Data analysis

Since there was no prediction or evidence for a differential pattern across regions (critical and post-critical words), I returned to separate by-region models. Within each model, I conducted pairwise comparisons between each condition and the baseline condition, and thus *p*-values on the frequentist analysis were corrected, within each model, using Bonferroni correction for three comparisons. Due to convergence failure, the analysis of the critical word excluded one slope (corresponding to the effect on the match condition) from by-subject random effects, and included only the intercept (with no slopes) on by-item random effects. In addition, at the spillover region the final model excluded the same slope from the by-subject effects structure (but retained all slopes on the by-item effects structure).

Exclusion criteria resulted in removal of nine participants (seven for low accuracy in comprehension questions and two for abnormal reading time averages). For the remaining 60

participants, trimming of high and low RTs affected 2.65% and 0.03% of the data, correspondingly.

5.3.1.2 Results

Word-by-word reading times of the different experimental conditions and by-condition numerical means of the tested regions are presented in Figure 5.4. The results of the frequentist and Bayesian analyses are summarized in Table 5.7 and 5.8, correspondingly.

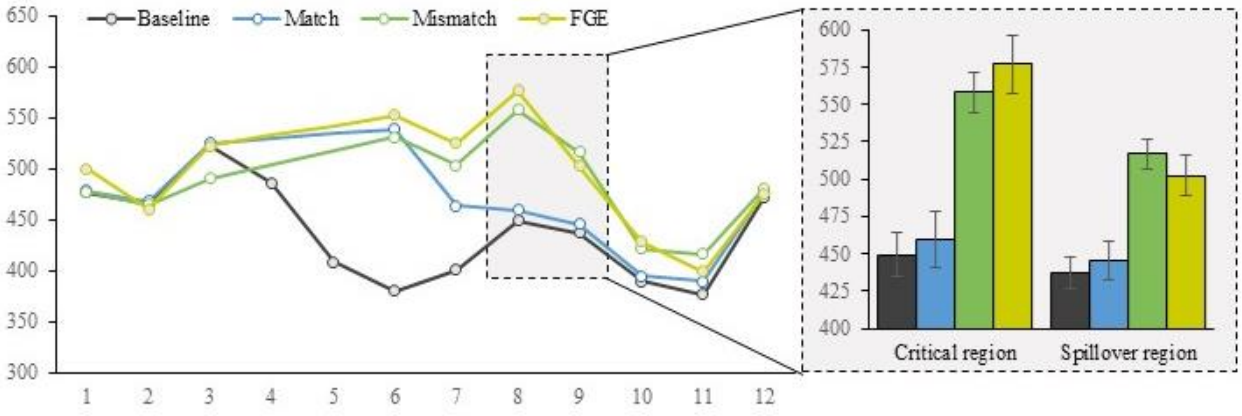

Figure 5.4. Word-by-word RT means (ms) by condition, Experiment 9A. Error bars represent +/-SE.

Region mapping: |The hairdresser|$_1$ |talked|$_2$ |with the designers|$_3$ (|because|$_4$) (|that they|$_5$) |(that) chose|$_6$ |mostly|$_7$ |ACC the models|$_8$ |the-main|$_9$ |of them|$_{10}$ |in the moment|$_{11}$ |the last|$_{12}$

The critical region is shaded gray.

At the critical region (the direct object position), both the frequentist and Bayesian analyses indicated increased RTs in the Mismatch and in the Filled-gap conditions, relative to the baseline condition. At the spillover region, the frequentist analysis failed to detect significant contrasts. In the Bayesian analysis, on the other hand, the contrast between the baseline and both the Mismatch and the Filled-gap conditions extended to the spillover region.

No reliable contrast was observed between the Match condition and the baseline. It should be noted that, in the Bayesian analysis, 90% of the posterior distribution was above zero for the Match-Baseline contrast at the critical region. This is a relatively high probability, yet as it does not reach the required level of reliability (95%), I treat it as equivocal evidence. At any rate, this possible effect is clearly smaller than that observed in the Mismatch and Filled-gap conditions: the contrast between the Match and baseline conditions produced a smaller estimate than the other two contrasts, such that its credible interval falls clearly outside those of the latter two conditions (Figure 5.6).

	Estimate	SE	*t*	*p*
Critical region:				
Match - Baseline	-0.049	0.025	-1.94	.157
Mismatch - Baseline	0.081	0.029	2.83	.016

FGE – Baseline	0.111	0.032	3.48	.002
Spillover region:				
Match - Baseline	-0.033	0.022	-1.48	.430
Mismatch - Baseline	0.077	0.031	2.48	.055
FGE – Baseline	0.056	0.028	2.01	.151

Table 5.7. Results of the frequentist analysis of Experiment 9A: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Critical region:		
Match – Baseline	16 [-8, 42]	90%
Mismatch – Baseline	80 [52, 110]	> 99%
FGE – Baseline	99 [68, 133]	> 99%
Spillover region:		
Match - Baseline	10 [-8, 30]	86%
Mismatch - Baseline	62 [36, 90]	> 99%
FGE - Baseline	53 [30, 76]	> 99%

Table 5.8. Results of the Bayesian analysis of Experiment 9A: Posterior means of the tested contrasts (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

5.3.1.3 Discussion

In this experiment, we observed increased reading times at the direct object in the Mismatch condition relative to an unambiguous baseline (with no filler-gap dependency), on a par with the Filled-gap condition. I interpret this as evidence that readers constructed an OR in the Mismatch condition, and were therefore required to perform reanalysis at the object position. I suggest that when the filler mismatches the verb, and an impersonal null subject analysis is available, readers interpret the input as a faithful OR with an impersonal subject, rather than assume a noisy SR.

No comparable effect was detected for the Match conditions. This is interesting since readers could in principle resort to the impersonal null subject reading in the Match condition as well. Such a strategy would have allowed them to quickly locate all obligatory arguments of the verb (with an impersonal subject and an object gap). However, this strategy was not adopted by the readers, at least not as consistently as in the Mismatch condition. Therefore, it is unlikely that an OR analysis is categorically adopted when the impersonal form is available. This is an important observation since, in the absence of corpus data, we cannot verify that SRs are still more frequent than ORs with null subjects. Thus, the lack of effect in the Match condition serves to corroborate the assumption that, when agreement conditions allow this, the SR reading is preferred over the OR one.

As predicted, the results contrast with the results from Experiment 8: In the current experiment the mismatch results in an OR analysis, while in Experiment 8 a noisy SR analysis was preferred over the OR one when the filler mismatched the verb. This aligns with the

hypothesis that the prior probability of the alternative structure affects the interpretation of the agreement mismatch, and that the canonical SR reading is not applied across the board as a heuristic. The faithful analysis in Experiment 9A is of a higher probability than that in Experiment 8, allowing for higher ranking of this OR analysis. Namely, I suggest that the contrast between the results of Experiments 8A-B and 9A should be interpreted as evidence for Rational Noisy-Channel processing, and specifically for the effect of the prior probability (of the alternative structure) on the interpretation of possibly noisy input.

An alternative explanation for the contrast between the results of Experiment 9A and those of Experiment 8 could attribute it to the distance between the filler and the verb. Recall that, in Experiment 8, a temporal or locative adjunct opened the relative clause and served as a buffer between the filler and the verb. This was done in order to somewhat increase the probability of a post-verbal subject structure (Shlonsky & Doron, 1992). However, it might have also contributed to the decay of the filler's representation and to an increase in the uncertainty regarding its agreement features by the time the verb is processed. The sentences in Experiment 9A lacked this adjunct "buffer" and thus, reliance on the filler's features during dependency formation might have been greater in this experiment. By this logic, agreement features are dismissed when the distance between the filler and the verb is long (as in Experiment 8) but can be weighed higher in shorter dependencies (Experiment 9A). In order to test this hypothesis, I conducted Experiment 9B.

5.3.2 Experiment 9B

To rule out the possibility that the contrast between Experiments 8A-B and 9A can be reduced to dependency length, I ran a modified version of Experiment 9A. In Experiment 9B, an adjunct phrase, of the same length as those in Experiments 8A-B, was inserted at the beginning of the relative clause.

5.3.2.1 Methods

Participants

Participants were 51 native speakers of Hebrew (according to self-report) (mean age: 25.96, range: 20-35). One participant was a bilingual speaker of Hebrew and English, and the rest were monolingual.

Materials

Materials were based on the 24 sets of Experiment 9A. I added a two-word adjunct at the beginning of the relative clause (in all conditions) and removed the condition of the "standard" filled-gap effect from the design (see example set in Table 5.9). Filler items were also identical to those used in Experiment 9A, except for a similar adjunct insertion in ten of the sentences.

Condition	Sentence

Match	*ha-saparit dibra im ha-me'acvim še-be-tahalix ha-sinun* the-hairdresser talked with the-designers that-in-process the-selection *baxru larov* ***et ha-dugmaniyot*** *ha-rašiyot šelahem ba-* chose-3.PL mostly **ACC the-models** the-main of.them in.the- *rega ha-axaron.* moment the-last. 'The hairdresser talked with the designers who usually chose their leading models at the last moment in the selection process.'
Mismatch	*ha-saparit dibra im ha-me'acev še-be-tahalix ha-sinun* the-hairdresser talked with the-designer that-in-process the-selection *baxru*-3.PL *larov* ***et ha-dugmaniyot*** *ha-rašiyot šelo ba-rega* chose-3.PL mostly **ACC the-models** the-main of.him in.the-moment *ha-axaron.* the-last. 'The hairdresser talked with the designer whose leading models were usually chosen at the last moment in the selection process'
Baseline	*ha-saparit dibra im ha-me'acvim mikeyvan še-be-tahalix* the-hairdresser talked with the-designers because that-in-process *ha-sinun hem baxru larov* ***et ha-dugmaniyot*** *ha-rašiyot* the-selection they chose-3.PL mostly **ACC the-models** the-main *šelahem ba-rega ha-axaron.* of.them in.the-moment the-last. 'The hairdresser talked with the designers because they usually chose their leading models at the last moment'

Table 5.9. Example set from Experiment 9B.
The critical NP is marked in bold; ACC = accusative case marker. PL = plural.

Procedure

The procedure was the same as in the previous experiments.

Data analysis

Data analysis followed the same protocol as in Experiment 9A, with separate models by region. Due to the lower number of conditions, *p*-values on the frequentist analysis were corrected, within each model, using Bonferroni correction for two comparisons. Due to convergence failure the final model of the critical region included only a partial random effects structure, with only one slope corresponding to the effect of the mismatch condition on by-subject effects, and no slopes on by-item effects. At the spillover region, the final random effect structure excluded only one slope (that corresponding to the effect of the match condition) from by-item effects, with a full random structure for by-subject effects.

Exclusion criteria resulted in removal of six participants (four for low accuracy in comprehension questions and two for abnormal reading time averages). For the remaining 45 participants, trimming of high and low RTs affected 2.57% and 0.02% of the data, correspondingly.

5.3.2.2 Results

Word-by-word reading times in the different experimental conditions and by-condition numerical means of the tested regions are presented in Figure 5.5. The results of the frequentist and Bayesian analyses are summarized in Table 5.10 and 5.11, correspondingly.

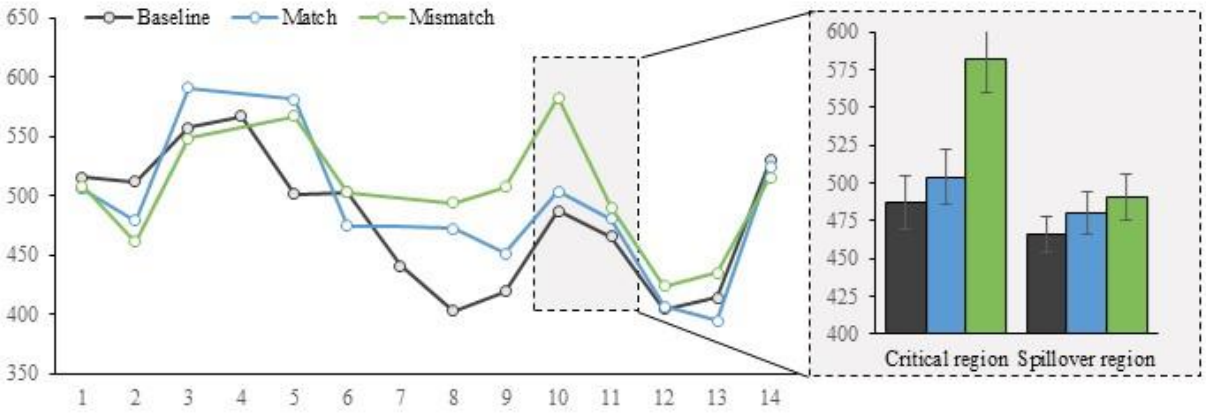

Figure 5.5. Word-by-word RT means (ms) by condition, Experiment 9B. Error bars represent +/-SE.

Region mapping: |The hairdresser|$_1$ |talked|$_2$ |with the designers|$_3$ (|because|$_4$) |that in the process|$_5$ |of selection|$_6$ (|they|$_7$) |chose|$_8$ |mostly|$_9$ |ACC the-models|$_{10}$ |the-main|$_{11}$ |of them|$_{12}$ |in the moment|$_{13}$ |the last|$_{14}$
The critical region is shaded gray.

At the critical region (the direct object position), I detected an increase in RTs in the Mismatch condition relative to the baseline condition (reliable on both frequentist and Bayesian models). There was no reliable evidence for increase in the Match condition relative to baseline, in either of the analyses. Finally, at the spillover region, both the Bayesian and frequentist analyses did not detect reliable effects.

	Estimate	SE	*t*	*p*
Critical region:				
Match - Baseline	-0.007	0.026	-0.28	> .99
Mismatch - Baseline	0.089	0.036	2.46	.033
Spillover region:				
Match - Baseline	0.005	0.023	0.21	> .99
Mismatch - Baseline	0.022	0.023	0.97	.67

Table 5.10. Results of the frequentist analysis of Experiment 9B: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Critical region:		
Match – Baseline	16 [-9, 41]	90%
Mismatch – Baseline	65 [28, 105]	> 99%
Spillover region:		
Match - Baseline, spillover region	8 [-13, 29]	78%
Mismatch - Baseline, spillover region	13 [-8, 36]	88%

Table 5.11. Results of the Bayesian analysis of Experiment 9B: Posterior means of the tested contrasts (with 95% Bayesian credible intervals) and the probability of the parameter being beyond zero.

5.3.2.3 Discussion

Experiment 9B replicated the findings of Experiment 9A. We observed increased reading times at the object position (i.e. a "filled-gap" effect) in the Mismatch condition, in line with the predictions. This suggests that the impersonal OR reading is adopted even with prolonged filler-verb distance. In addition, we did not observe reliable evidence for a corresponding filled-gap effect when the filler matched the verb's features. This suggests that the parser does not prioritize the OR reading whenever the impersonal null subject is possible, and that the balance between the SR and OR analyses depends also on the mismatch in agreement features.

It might be noted that the filled-gap effect in this experiment produces a numerically smaller estimate relative to that in Experiment 9A (see Figure 5.6). However, since there is also high overlap between the posterior distributions, this difference is not reliable. Yet, if I am to speculate on the source of this difference, it can be due to lower weighing of agreement information when the distance from the filler is greater. Namely, the certainty regarding the filler's features may decrease with distance, and readers may dismiss the mismatch more easily.

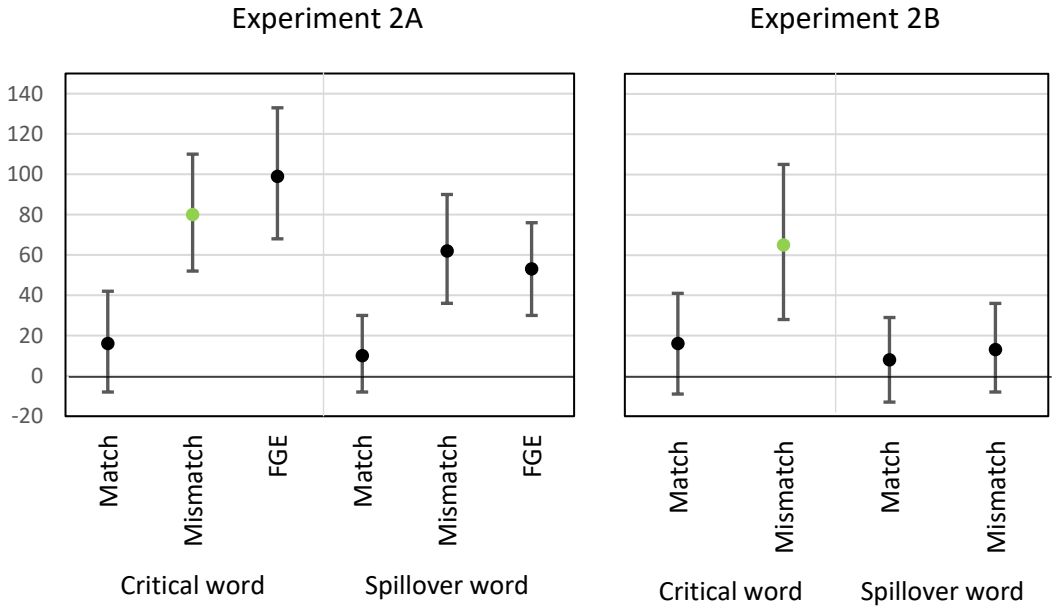

Figure 5.6. Results of the Bayesian analysis of Experiments 9A-B: The posterior means of the tested contrasts with 95% CrI. The coloured points mark the critical contrasts.

Another possibly interesting pattern suggested by this data is presented by the considerably larger effects we observed in Experiments 9A-B relative to Experiments 8A-B (see Figures 5.3 and 5.6). This difference in the magnitude of the effects might suggest that reanalysis of a subject gap is less costly than that of an object gap. It is known that filled-gap effects are harder to find in subject position (Stowe, 1986; Lee, 2004). However, this was previously attributed to the short distance between the initiation of the "active filler" strategy and the filled-gap position, or to the absence of a verb with which the subject gap or argument is integrated. In our case, neither could underlie the contrast in effect magnitude (in all experiments the reanalysis occurs after the verb). It could be that subject gaps are reanalyzed more frequently in everyday communication and thus their reanalysis creates less processing disruption.

Overall, the replication of Experiment 9A suggests that the contrast in readers' interpretation of agreement information in Experiments 8 and 9 cannot be reduced to distance effects only. I therefore interpret the results obtained in this experiment as evidence that the difference in the prior probability of the faithful OR alternative is the most likely source of the contrast between readers' strategies in Experiments 8 and 9. To provide additional support for this claim, in the next section I present my final experiment, which compares the two structures within one experimental design.

5.4 Experiment 10: Evidence from sentence completion

In Experiment 10, I wanted to test my interpretation of the results from the previous experiments, and of the contrast between them, within one experiment. I chose to use a different methodology – a sentence completion task – in order to tap into readers' preferences more directly, rather than deduce them from reanalysis costs.

I conducted a sentence completion experiment in which participants were asked to complete a preamble that was presented word-by-word in rapid serial visual presentation (to increase the likelihood of perception/memory errors). As the preamble, I used sentences truncated after the verb. The experimental design included four conditions, crossing the number agreement on the filler and the verb. I hypothesized that SR completions will be less common in those Mismatch conditions that allow an impersonal null subject OR. In addition, I manipulated presentation speed of the words in the preamble, to test the effect of perceptual uncertainty. I hypothesized that SR completions will be more common in the mismatch conditions when readers have shorter exposure time to the words of the preamble.

5.4.1 Methods

Participants

Participants were 132 native speakers of Hebrew (according to self-report) (mean age: 25.2, range: 18-44). Thirty-two participants were bilingual of Hebrew and either Russian, English, Spanish, French, Farsi or Hindi, and the rest were monolingual.

Materials

The experiment included 24 experimental items of four conditions, crossing the form of the verb (singular vs. plural) and its agreement with the filler (matching or mismatching). The sentences presented a beginning of a relative clause and were truncated after the embedded verb (see example set in Table 5.12). The materials included a clause initial adjunct, as a buffer between the filler and the verb (on a par with Experiment 8A-B and 9B). The experimental items were distributed in a Latin square design along with 12 OR filler sentences (to balance a possible bias in favor of SRs in my materials, originating in the Match conditions).

Condition	Sentence
Match -Verb.SG	*xibavnu et ha-talmida še-lamrot ha-xašašot mac'a …* we.liked ACC the-pupil.SG that-despite the-concerns found.SG …
Match -Verb.PL	*xibavnu et ha-talmidot še-lamrot ha-xašašot mac'u …* we.liked ACC the-pupils.PL that-despite the-concerns found.PL …
Mismatch -Verb.SG	*xibavnu et ha-talmidot še-lamrot ha-xašašot mac'a …* we.liked ACC the-pupils.PL that-despite the-concerns found.SG …

Mismatch -Verb.PL	*xibavnu et ha-talmida še-lamrot ha-xašašot mac'u …*	
	we.liked ACC the-pupil.SG that-despite the-concerns found.PL …	

Table 5.12. Example set from Experiment 10.

Procedure

The experiment was built using Ibex Farm. The preamble of the sentences was presented using rapid visual serial presentation (RSVP) at a rate of 400ms per word. To test the effect of perceptual uncertainty, I manipulated the duration of word presentation, such that items were presented either for 400ms with no inter-stimulus interval, or for 250ms with a 150ms inter-stimulus interval (maintaining the rate of 400ms per word). Word duration was manipulated within-participant and counterbalanced in a Latin Square design

After the end of the preamble, participants were instructed to complete the sentence fragment in a designated text box. Before they began the experiment, participants underwent a practice block of four items. The order of presentation was randomized for each participant.

Data analysis

Data from 15 participants were excluded as they did not understand the nature of the task (repeating the sentence fragment instead of completing it). Productions of the remaining 117 participants were coded for SR vs. OR completions. For example, a sentence like (50a) would be considered a SR, as it includes indication (via an accusative marked NP) that the object position does not contain a gap. On the other hand, a completion like (50b) is an OR production since it presents a post-verbal subject (see Appendix D for details about the coding scheme).

(50) *xibavnu et ha-talmidot še-mac'a …*
we.liked ACC the-students.PL that-found.SG …

a. *et ha-xatul bari ve-shalem.*
ACC the-cat healthy and-whole
'We liked the students who found the cat in a good shape.'

b. *ha-madrixa be-kce ha-maslul.*
the-guide in-edge the-trail.
'We liked the students who the guide found at the end of the trail.'

The data were analyzed using logistic mixed-effects models (Bayesian and frequentist). To test the effect of verb form (i.e. the possibility of a null subject) on the rate of SR productions I used treatment (dummy) coding for this factor, obtaining a measure comparing the singular and plural verb conditions, when the filler and the verb matched (regardless of the effects in the Mismatch conditions). Similarly, to test the base rate of the agreement mismatch effect, this factor was also treatment (dummy) coded, such that the main effect represents the contrast between the Match and Mismatch conditions of the singular verb version (independent of the effect of the impersonal null subject option in the plural verb conditions). Finally, I used sum coding of the word duration factor in order to average the above-mentioned effects over both

presentation rates. Due to repeated convergence failures the final analysis included a minimal random effects structure, with only the intercept of by-subject effects.

5.4.2 Results

The distribution of SR and non-SR productions across conditions is presented in Figure 5.7. The results of the frequentist and Bayesian analyses are summarized in Tables 5.13 and 5.14, correspondingly.

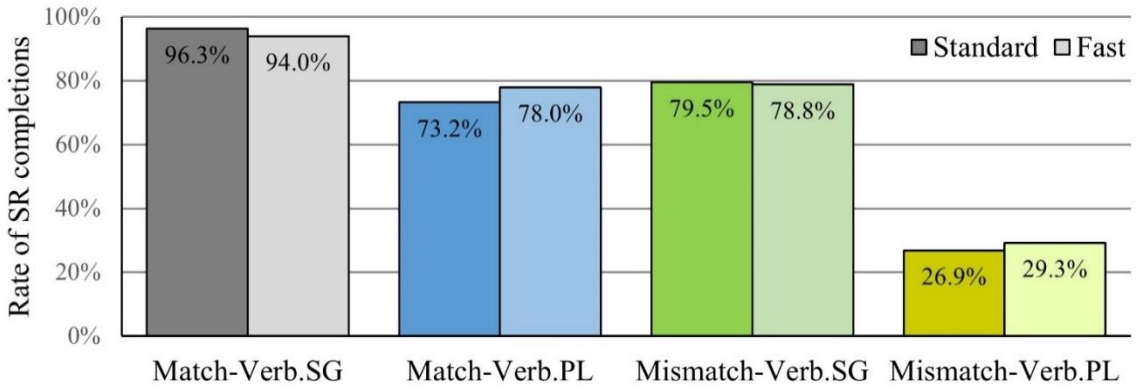

Figure 5.7. Rate of subject relative completions by condition, Experiment 10.

Both the frequentist and Bayesian analyses indicated several effects. First, we can observe an effect for the verb-form factor, such that the usage of a plural verb resulted in a decreased rate of SR productions. In addition, we can observe an effect of agreement, such that mismatch between the verb and the filler resulted in a decreased rate of SR productions. Crucially, we see evidence for an interaction effect, signifying that the decrease in SR production rate due to agreement mismatch was larger when the verb was plural.

We also observe an interaction between verb form and word duration, such that sentences with a plural verb were completed as SR in a higher rate in the speeded presentation conditions. The main effect of word duration and its other interactions did not reliably affect SR production rates. No other effects were reliable on either of the analyses.

	Estimate	SE	*t*	*p*
Effect of verb form	-1.950	0.204	-9.57	< .001
Effect of agreement match	-1.731	0.206	-8.41	< .001
Effect of word duration	0.260	0.181	1.44	.151
Interaction of verb form and agreement match	-0.536	0.243	-2.21	.027
Interaction of verb form and word duration	-0.399	0.203	-1.96	.050
Interaction of agreement match and word duration	-0.238	0.205	-1.16	.247
Three-way interaction of verb form, agreement match and word duration	0.313	0.241	1.29	.196

Table 5.13. Results of the frequentist analysis of Experiment 10: Estimate, standard-error, *t*-value and *p*-values.

	Posterior mean [CrI]	Posterior beyond zero
Effect of verb form	-1.65 [-2.21, -1.08]	> 99%
Effect of agreement match	-1.23 [-1.81, -0.65]	> 99%
Effect of word duration	0.24 [-0.10, 0.58]	91%
Interaction of verb form and agreement match	-1.73 [-2.38, -1.09]	> 99%
Interaction of verb form and word duration	-0.33 [-0.73, 0.07]	95%
Interaction of agreement match and word duration	-0.21 [-0.61, 0.19]	84%
Three-way interaction of verb form, agreement match and word duration	0.22 [-0.29, 0.73]	79%

Table 5.14. Results of the Bayesian analysis of Experiment 10: Posterior means on the log-odds scale (with 95% Bayesian CrI) and the probability of the parameter being beyond zero.

5.4.3 Discussion

In this experiment I tested the contrast between the two structures of Experiments 8 and 9 within one experimental design, which taps into readers' preferred interpretation more directly. The results suggest that the rate of SR interpretations depends on the availability of the impersonal null subject reading. The main effect of verb form suggests an overall increase in OR rate when the verb licenses an impersonal null subject. In addition, we also observed an overall decrease in SR rate for the Mismatch conditions, suggesting that, to some extent, readers use the mismatch itself as evidence against the SR structure. Crucially, the interaction between verb form and agreement suggests that the combination of the Mismatch cue and the availability of an impersonal null subject reading yields a change in strategy, over and above the effect of each of these factors independently. Both factors conspire to increase the probability of an OR structure, yet their interaction suggests that while a corrupted SR analysis is more likely than a faithful OR with a post-verbal subject, it is still less likely than that of a faithful OR with an impersonal null subject. Namely, the way readers interpret the agreement mismatch depends on the prior probability of the grammatical faithful alternative. Specifically, agreement is more likely to be compromised when the faithful alternative is one of low prior probability, like the post-verbal subject alternative in the Mismatch singular verb condition.

The presentation rate manipulation did not produce the predicted pattern of results. The only significant effect was an increase in the rate of SR productions following plural verbs in the speeded presentation condition, which cannot be attributed to noise compensation as it did not interact with filler-verb match. The lack of a three-way interaction could be due to the low sensitivity of the manipulation. Namely, it could be that the speeded condition was not sufficiently difficult and did not increase readers' uncertainty enough to allow detection of a contrast. Alternatively, the lack of effects could be attributed to other types of uncertainty that speeded presentation might give rise to. It could be that in the speeded presentation, readers are not only uncertain about the perceived word form, but also incorporate higher likelihood of

missing a whole word. In the current setting, these two types of uncertainty may pull the interpretation in different directions. If readers assume that they might be missing some of the words, they may interpret the mismatch between the filler and the verb as indication that they missed a word corresponding to the subject. Therefore, while the uncertainty regarding the perceived words may increase the rate of SR completions, the likelihood of missing a whole word may increase the rate of OR completions.

5.5. General discussion

5.5.1 Summary

In this study, I investigated the processing of Hebrew relative clauses that are ambiguous between subject relatives with an agreement error, and input-faithful, grammatical, but less common, object relative structures. I found that the agreement mismatch between the filler and the verb does not entirely rule out the possibility of a subject gap for comprehenders, but is treated differently in different environments. In Experiment 8, we observed reanalysis costs at the true (post-verbal) subject in the relative clause, suggesting that readers had preferred constructing a SR, dismissing the agreement mismatch, over adopting a rare OR structure with a post-verbal subject. However, in Experiment 9, when another, more common, OR reading was grammatically possible (containing an impersonal null subject), we observed reanalysis costs at the object position of the relative clause, suggesting that an OR was constructed. Importantly, we observed all the effects while using only globally grammatical sentences (in both experimental and filler items), suggesting that the consideration of a possibility of error is not a task-dependent strategy. Finally, Experiment 10 corroborated these findings in a sentence completion task. We observed mostly SR completions when only the rare OR structure (with a post-verbal subject) could restore grammaticality of the sentence, but mostly OR completions when the possibility of an impersonal null subject was introduced.

5.5.2 The effect of prior probability on the interpretation of agreement mismatches

Human communication is frequently corrupted by noise, originating either in the production system of the interlocutor or in perception or memory errors of the addressee. Yet, addressees often manage to recover the intended meaning from the noisy input with little effort. Inspired by recent models of noise inferences and uncertainty maintenance in sentence processing (Levy, 2008b; Gibson et al., 2013), my findings provide evidence for rational noisy-channel inference during incremental processing.

First, I demonstrate the readers treat a string which has only one grammatical analysis as ambiguous. Namely, I find that readers consider ungrammatical representations (in this case mismatching subject-verb agreement) during online processing, even though a semantically plausible and grammatically pristine analysis is available. Second, I show that assuming a corrupted structural dependency hinges on the prior probability of the faithful alternative. Readers adopted a corrupted analysis only when the grammatically pristine one was based on a very infrequent phenomenon in the language (the inverted word order in Experiment 8A-B). When another, relatively probable, structure was grammatically possible (Experiment 9A-B), the strategy shifted, and interpretation was pulled towards the faithful analysis. This suggests that the above strategy is aimed for forming more probable structures and does not represent a categorical negligence of agreement. I would like to stress that the idea of dismissing agreement information, based on rational noisy-channel processing, should not be taken to suggest that agreement is not computed or that the mismatch is not detected. On the contrary, I suggest that although the parser identifies the mismatch, low probability of the grammatical alternative can push the parser to the SR reading.

Overall, the results suggest that readers apply elaborate probabilistic knowledge regarding the distribution of structures in their language during online processing, and that they are willing to compromise some aspects of the input to arrive at a higher-probability interpretation. I demonstrate this in Hebrew, a language which has not yet been studied in this context. Previous studies of the Noisy-Channel model have focused on English, which is more limited in terms of manipulating word order, morphological marking and orthographic neighborhood sizes.

These results are direct predictions of Noisy-Channel inference models (Levy, 2008b), but it should also be mentioned that they could also be in line with models of Self-Organizing Sentence Processing (Tabor & Hutchins, 2004). The framework of Self-Organizing Sentence Processing suggests that the parser can sometimes choose an ungrammatical structure over a grammatical one which is difficult to construct (Villata, Sprouse, & Tabor, 2019). In this framework, new constituents interact with previous ones in all possible ways to form structure, and thus the various attachments dynamically compete. The system gradually stabilizes as attachments with a good feature match generally outcompete attachments with a poor feature match. However, the system allows generation of intermediate structures, where the attachment's fit is not perfect (graded on a harmony scale of 0 to 1) when no optimal bond is available, as in ungrammatical sentences and difficult garden paths (i.e. the system forces the sub-optimal attachments in cases of extreme difficulty). This could in principle, yield the SR representation with the mismatch in subject-verb agreement which readers resort to in this study. Similarly to the noisy-channel idea, here too one could claim that the rarity of the alternative word order makes the harmonious structure unavailable. Yet, more research is required model and test these sub-optimal attachments to understand how and why it should arise Self-Organizing Sentence Processing also could naturally explain the timing difference found in Experiment 1. The dynamical representation and its graded attachment coefficient suitable grounds of such effects.[27]

5.5.3 Alternative accounts

Here I address several possible alternatives to my account of the findings, and point out why I believe they cannot explain the entirety of the results I observed, unlike the noisy-channel interpretation, and specifically the hypothesis that the prior probability of alternative analyses affects the interpretation of agreement mismatches.

The costs of filler-gap dependency processing and the SR/OR asymmetry

One alternative account for my results would suggest that they can be explained based only on a general SR preference, or on the costs of infrequent structures. It is widely accepted that the

[27] In that experiment, the reanalysis cost in noisy conditions (moving from a non-agreeing SR interpretation to an OR one) was found on the first word of the subject phrase, while the normal reanalysis cost was delayed (i.e. when there was no agreement mismatch, the cost was traced on the second word of the subject). Under Self-Organizing Sentence Processing, a better fit before the appearance of the post-verbal subject in (matching relative to mismatching agreement conditions) could mean that the system would require more time to respond to the sudden realization that something is amiss. Namely, when there is already some initial instability due to the lack of subject-verb agreement in the SR representation, the system is already on the verge of faltering so it is quicker execute the reanalysis.

probability of an utterance affects its processing difficulty (e.g. Levy, 2008a; Smith & Levy, 2013), and that SR structures are more probable and simpler to process than ORs (e.g. King & Just, 1991; Traxler, Morris, & Seely, 2002; Grodner & Gibson, 2005; Gennari & MacDonald, 2008; Staub, 2010). These assumptions predict that a SR analysis would be preferred over an OR analysis, in line with the results from Experiment 8. Yet they do not account for the contrast between Experiments 8 and 9, or between the singular verb conditions and the plural verb conditions in Experiment 10. In Experiment 8, the agreement mismatch was indeed dismissed in favor of an SR analysis; However, in Experiment 9 the mismatch was taken as evidence for an OR analysis with a null subject, which is still less probable than a SR. Therefore, the account must include an interaction between the dismissal of agreement and the probability of the alternative structure.

A similar line of thought could suggest that the contrasts which I rely on, namely those between the processing of relative clauses and that of baseline sentences with no dependency, are confounded by the general costs of dependency processing. However, once again this does not seem likely given the results of Experiment 9, where reading times in the match condition aligned with reading times in the baseline condition, despite the existence of a dependency in the former. I maintain that the critical region is distant enough from the initiation of the dependency, and reading times of this region are therefore less affected by general dependency processing costs and more likely to reflect the abandoning of a gap that was previously postulated.

Likelihood of edits

The contrast between the results of Experiments 8 and 9 can, to some extent, be attributed to the likelihood of the corruption, since edits of different types are required for different agreement errors, and some may be more likely than others. In particular, deletion errors are more likely than insertion errors, according to the Bayesian "size principle" (Xu & Tenenbaum, 2007).[28] In line with this, it was observed that readers are more likely to adopt the "corrupted" analysis when it involves assuming deletion rather than the insertion (Gibson et al., 2013; Poppels & Levy, 2016; Ryskin, Futrell, Kiran & Gibson, 2018).

In the context of the current study, it should be noted that in Experiments 8A-B, the corrupted SR analysis requires assuming either an insertion error at the filler (the intended filler was singular, and it was corrupted to produce a plural form by adding the plural suffix) or a deletion error at the verb (the intended verb was plural, and the singular form is the result of deletion of the plural morpheme). In contrast, in Experiments 9A-B, the corrupted SR analysis requires assuming the opposite (a deletion error at the filler, editing it from plural to singular, or an insertion error at the verb, editing it from singular to plural).

If one assumes that readers consistently edit the verb's features, the deletion/insertion contrast may confound my results, as it pulls towards a SR interpretation to a greater extent in Experiment 8A-B (in line with my predictions). However, I believe that this interpretation of

[28] A deletion only requires a particular word to be randomly selected from the set of words in the sentence, whereas an insertion requires the selection of a specific word from the producer's vocabulary. Since vocabulary size is considerably larger than the number of words in a sentence, the Bayesian "size principle" suggests that insertion of a specific word has smaller likelihood than deletion of a specific word.

the results is not likely, for several reasons. First, I suggest that editing of the filler's features, rather than the verb's, is more likely. Given the non-cumulative presentation of the sentence, I believe that readers maintained higher uncertainty regarding the filler's features (rather than the verb's), as it was no longer available for re-reading at the point of mismatch detection, that is upon encountering the verb. Therefore, readers should be more likely to amend the filler's representation than that of the verb. If the filler's features are edited, errors are more likely in Experiments 9A-B, thus pulling the results in the direction opposite to my findings and predictions. The likelihood of deletion and insertion errors thus cannot account for the current findings.

Moreover, in Experiment 10 I aimed to better control for this possible confound by using feminine forms only, for which editing of the verb's features – either from singular to plural or from plural to singular - would always require readers to assume a substitution error. This eliminates the possible effect of the deletion/insertion contrast on the analysis of the verb. The fact that the results of Experiments 8 and 9 were still mirrored in Experiment 10 supports my conclusion that the different types of edits cannot account for the findings of the current study.

Avoidance of predicting new material

A different alternative account for my findings may attribute them to a general preference for avoiding prediction of a new discourse referent. In Experiment 8, in order to construct an OR interpretation at the verb, the reader has to predict an upcoming new discourse entity (the subject), but this is not the case in Experiment 9. This would produce the higher tendency to adopt SR analyses in Experiment 8 relative to 9.

However, this interpretation is not consistent with results from the match condition of Experiments 9 and 10. In the match condition with a plural filler and a plural verb, namely when an impersonal null subject is possible, full thematic assignment without need to predict any new discourse entities can be established immediately by assuming an OR reading, assigning the agent role of the verb to the null generic subject, and its theme role to the gap. If readers refrain from predicting new discourse reference, an OR should be consistently constructed in these cases, to avoid prediction of a yet-unknown complement to the verb. However, I did not find reliable evidence that such an OR analysis is adopted in the match condition of Experiments 9A-B. On the contrary, in Experiment 10, I observed mostly SR completions in that condition. Therefore, I argue that this account is not consistent with the overall pattern of results.

Spillover from the verb region

Lastly, since the design of Experiments 8 and 9 is based on observation of processing difficulty in the mismatch condition, the effects could be argued to be contaminated by spillover from the verb mismatch effect. It could be suggested that the increased reading times, which I interpret as reanalysis costs, merely reflect mismatch-detection costs that spilled over from the verb region.

However, I believe that this is not likely to be the case. First, note that in both Experiments 8 and 9, an adverb was used as a buffer between the verb and the critical region. This makes the spillover interpretation less likely as it would require the effect to migrate two words

downstream. Moreover, this interpretation of the results wrongfully predicts that processing costs would be greater when the mismatch is more prominent. Yet, a two-feature mismatch (in Experiment 8B) did not elicit greater processing disruption; If anything, the effect was smaller than that of the one-feature mismatch (in line with the prediction of Noisy-Channel processing). Finally, since the dependent measure in the last experiment was the produced structure, and not reading times, this account cannot explain the results of Experiment 10.

5.5.4 Directions for future research

Future research can take the above findings as a basis to look more closely into the prior and likelihood components of the noisy-channel inference. As for priors, the current study is limited by the lack of corpus investigations regarding the frequency of the different structures. Future research could get numeric frequency estimates for post-verbal and impersonal subjects with different verbs (or in different environments – with/without a clause-initial adjunct, with animate/inanimate arguments, etc.). This would allow testing the gradient effect of the prior more closely, and estimate the relation between frequency and the reader's interpretation and processing. In the realm of interpretation choices (e.g. in sentence completion data), the relation should follow a linear function, as the Bayes Theorem suggests. On measures of processing time (e.g. reanalysis costs) the relation may follow a logarithmic function as Surprisal theory suggests (Levy, 2008a; Levy & Smith, 2013).

Further investigations could target the likelihood component, the probability that some other intended utterance was corrupted into the form of the current input. Additional research could try to investigate the noise model readers use by manipulating the extent of the mismatch between the filler and the verb. For example, it could be that the mismatch is perceived in terms of featural mismatch, orthographic editing (letter deletion/insertion), or error in a selecting the wrong slot in the inflectional paradigm (neutralizing the size principle effect). It could also be illuminating to test whether different mismatches are experienced as more severe (and thus give rise to less subject relative interpretations). It could be that gender and number mismatches are given different weights, as misgendering probably has a higher social penalty to it. In addition, the linguistic prominence or corpus validity of different linguistic features could also affect the rate at which these features are compromised. Case marking, for instance, might be more prominent and thus could be compromised less easily than agreement features.

The likelihood component could also be investigated by manipulating the experimental environment and looking at readers' adaptation to it. For example, adding typing errors could decrease the rate of object relative interpretations in experimental designs like that of Experiment 9 (where an OR with an impersonal subject was favored over the corrupted SR form). On the other hand, high register items could decrease the rate of subject relative interpretations in experimental designs like that of Experiment 8 (where a corrupted SR was preferred over the rare OR form).

Lastly, the timing of noisy-channel inference and the association between such inference and perceptual (un)certainty can be further investigated using more fine-grained measures like eyetracking while reading. This would allow better understanding of the timing contrast in reanalysis costs (reported in Experiment 8). Moreover, it would be interesting to check if

readers' interpretations (preference for the corrupted SR or pristine OR analyses) are modulated by fixation rates directly on the agreement suffixes. The parser could take into account errors in parafoveal view and thus be more lenient in compromising agreement features when they were not the target of a previous fixation.

5.5.4 Conclusions

The results of this study suggest that the prior probability of alternative analyses modulates the interpretation of agreement. I show that during online processing, readers apply elaborate knowledge regarding the distribution of structures in their language, and they are willing to compromise subject-verb agreement to refrain from (grammatical but) highly improbable structures. I propose (i) that incoming input is integrated with existing knowledge about the probability of various linguistic structures, pulling the interpretation towards the more probable structure; and (ii) that in line with the framework of Rational Noisy-Channel processing, the bias towards more probable structures applies not only for ambiguous strings (local or global), but also for cases where only one grammatical reading is available, based on consideration of possible production and perception errors.

Chapter 6
Concluding remarks

This dissertation presented a series of studies, focused on the processing of different dependencies in Hebrew: filler-gap dependencies, anaphora, and subject-verb agreement. The goal of this work was to combine research from different aspects of sentence processing in order to discuss possible ways in which parsing strategies may compensate for the mishaps which arise in human language - prediction errors, memory fallibility, and misperception of the input or typing errors.

Human communication is frequently corrupted by noise, originating either in the production system of the interlocutor or in the perception and comprehension systems of the addressee. Yet, addressees routinely manage to recover the intended meaning with little effort. I speculated that to confront the possibility of errors, readers and listeners may apply elaborate probabilistic knowledge, employ differential strategies for obligatory and non-obligatory relations, and utilize various linguistic cues available in their language.

In Chapter 3, I looked into prediction of filler-gap dependencies and of discourse dependencies. Successful predictions, which turn out to be in line with the continuation of the sentence, have the benefit of facilitating the processing of subsequent input. However, unsuccessful predictions may be costly. Moreover, the preparatory operations may incur efforts by themselves. Therefore, prediction might not be cost-effective at all times, and the parser may modulate the degree to which it actively engages in prediction, in order to reduce reanalysis costs under certain circumstances. If the parser operates efficiently, we may identify anticipatory processes which are binding enough to result in reanalysis costs, and others which are not.

I investigated the reanalysis costs which filler-gap and discourse dependencies incur. I found evidence for three different degrees of prediction: Contextually available antecedents may incite a preference for co-reference, yet the prediction associated with them is weak and does not involve costly reanalysis; Pragmatic motivations (e.g. information structure considerations) can provoke predictive dependency formation whose disconfirmation is costly; Lastly, syntactic licensing pressures enhance the predictive process such that dependency formation is observed in earlier measures, with even higher reanalysis costs, and with fuller semantic consequences. As explained in Chapter 3, the results are not straightforwardly reducible to Surprisal effects.

I proposed that the parser establishes predictions with different levels of commitment, as a way of modulating the potential costs of prediction errors. This provides flexibility in the formation and transformation of linguistic representations during incremental processing. Thus, it may also facilitate accurate interpretation, reducing possible lingering of initial misanalyses. These findings also support the syntactic account of the Active Filler strategy.

In Chapter 4, I examined the parser's propensity to make agreement attraction errors. These systematic interference errors are a central test case for the parser's memory system. This line of research provides a unique window into the representations and operations which arise during incremental sentence processing. Therefore, agreement dependencies have an important role when considering questions like how we accommodate the possibility of errors in our parsing strategies, and how we establish proper communication despite such errors.

I argued that establishing agreement relations is not independent of previous relations which targeted the same phrase. I examined the interaction between verbal and pronominal agreement in Hebrew, a language with a complex conjugation system. I found that agreement computations at different parts of the sentence interact, and can thus help readers confront the fallibility of memory representation and retrieval. Specifically, verbal agreement, which confirms the feature representation of the subject, can help later processing stages which require re-accessing that subject. This can also prevent a cascade of attraction errors which draw on each other. Comprehenders might be able to skip over the pitfalls of unstable representation traces or risky memory access, through giving higher weight to intermediate agreement sites (when these are available). Substantially more work is required to establish the exact mechanisms which underlie these effects. Yet, this study also contributes a new vantage point on the debate regarding the way agreement errors occur, through retrieval interference or representational fluidity.

It should be noted that higher reliance on the recent agreement site may make processing of the comprehender more susceptible to errors cascading from the producer's agreement mismatches. For example, while it is clear how grammatical verbal agreement may help stabilizing the representation of the subject for subsequent retrievals, ungrammatical verbs may help readers avoid the occasional attraction errors, but would cause consistently erroneous agreement relations downstream. Therefore, this strategy might be beneficial for counteracting the memory fallibility of the parser, but increase the dependence of grammatical licensing on proper input. To evaluate whether such strategy minimizes agreement errors in practice, future studies should consider the probability of attraction errors in comprehension, compared with the rate of agreement errors in production (and their distribution).

Lastly, in Chapter 5, I presented new evidence for noisy-channel processing and for the way frequency of syntactic alternatives is considered when the parser makes such inferences. I utilized an intriguing case of structural ambiguity between subject and object relative clauses to test parsing preferences in the face of contrasting biases. I found that during online processing, readers are willing to compromise subject-verb agreement to refrain from (grammatical but) highly improbable structures. As proponents of the Noisy-Channel model suggest, this strategy is useful considering that errors are frequent in human communication and thus may conceal the utterance intended by the speaker/writer.

I proposed that incoming input is integrated with existing knowledge about the probability of various linguistic structures, pulling the interpretation towards the more probable structure, even when this means creating a corrupted representation (with some grammatical mismatch) over a pristine one. Yet, I argued that this tool is employed only when the grammatical alternative is exceptionally rare, making the choice of simple errors more rational. When bias against the grammatical alternative is not as strong, readers do utilize the agreement mismatch as a cue for constructing the dispreferred reading. Thus, readers apply elaborate knowledge regarding the distribution of structures in their language, yet are not oblivious to the pristineness and grammaticality of their representations.

Overall, I suggested that while constructing structural dependencies within the sentence, comprehenders balance processing strategies, probabilistic knowledge, grammatical and extra-grammatical constraints. This system of "checks and balances" helps us arrive at an interpretation close to that intended by our interlocutor.

The rationality of the parser, and the extent to which memory capacity limits our interpretation mechanisms, have long been debated in the sentence processing literature. This took form on early days as debates regarding the availability of different information types (interactive vs. modular processing) and of maintaining multiple representations in initial processing (parallel vs. serial processing). Some suggested that the parser is unable to map all the linguistic options and the extra-linguistic evidence in real time because it requires more memory capacity than that available in our cognitive system processing (Frazier, 1987b). Others argued that what facilitates comprehension are exactly those strategies of interpreting the input using multiple sources of information with gradience between one reading to another (MacDonald et al., 1994).

Later on, findings regarding robust misinterpretations (Ferreira, 2003; Christianson, Hollingworth, Halliwell, & Ferreira, 2003) and the effect of local coherence (Tabor, Galantucci, & Richardson, 2004) turned the tables. They seemed to suggest that the rationality and capacity of the parser are even more limited than proponents of either approach suspected. Good-Enough processing approaches (Ferreira, Ferraro, & Bailey, 2002; Ferreira, & Patson, 2007) then argued for a heuristic relation between online parsing and grammatical knowledge, rather than a direct relation, limited or unlimited in scope. Yet recently, it has been suggested that even supposedly incoherent or shallow interpretations might be rational in their own way (Levy et al., 2009; Gibson et al., 2013).

The studies reported in this dissertation provide additional perspectives on these classical debates. How can the parser be both limited in memory capacity and rational in interpretation strategies? Different researchers suggested that some parsing strategies can be both rational and memory efficient as they allow minimization of reanalysis costs (MacDonald et al., 1994) and formation of more local dependencies (e.g. Levy et al., 2009). In this dissertation, I illustrated additional cases where this logic can be relevant (in reanalysis in Chapter 3; and in local dependencies in Chapters 4 and 5), without depending on parallel processing assumptions, or reducing the language processing architecture to frequency monitoring. Thus, the research presented here may bring together different traditions of research, expose novel findings, and suggest intriguing insights. In my view, it sheds light on the sophistication and

the complexity of language comprehension and brings us a small step closer to understanding how come human communication proceeds adeptly in the face of manifold linguistic details and despite the frequent missteps of production, perception, memory and prediction errors.

References

Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (in press). Extraction from subjects: differences in acceptability depend on the discourse function of the construction. *Cognition*.

Almeida, D., Tucker, M., & Dhabi, N. A. (2017). The complex structure of agreement errors: Evidence from distributional analyses of agreement attraction in Arabic. In *Proceedings of the 47th Meeting of the North-Eastern Linguistic Society*.

Altmann, G.T.M. (1999). Thematic role assignment in context. *Journal of Memory and Language, 41*(1), 124-145.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518.

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

Aoshima, S., Phillips C., & Weinberg A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51*(1), 23-54.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 914.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science, 15*(9), 578-582.

Arnon, I. (2010). Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language, 37*(1), 27-57.

Atkinson, R.C., & Shiffrin R.M. (1968) Human memory: A proposed system and its control processes. In K.W. Spence, J.T. Spence (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 2, Academic Press, New York (1968), pp. 89-195.

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.

Baird, R., & Koslick, J. D. (1974). Recall of grammatical relations within clause-containing sentences. *Journal of Psycholinguistic Research, 3*(2), 165-171.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364-390.

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

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition, 21*(4), 477-487.

Bassel, N. (2018). *Anaphors in space.* MA Thesis, Tel Aviv University.

Bates, D., Maechler M., Bolker B., & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Berman, R. A. (1980). The case of an (S) VO language: Subjectless constructions in Modern Hebrew. *Language*, 759-776.

Berman, R. A. (2011). Revisiting impersonal constructions in Modern Hebrew: Discourse-based perspectives. *Impersonal constructions: A cross-linguistic perspective*, 323-356.

Betancort, M., Carreiras, M., & Sturt, P. (2009). Short article: The processing of subject and object relative clauses in Spanish: An eye-tracking study. *Quarterly Journal of Experimental Psychology*, *62*(10), 1915-1929.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.) *Cognition and the Development of Language.* New York: John Wiley and Sons.

Birner, B. J. (1996). *The Discourse Function of Inversion in English.* New York, NY: Garland Publishing.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45-93.

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., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, *23*(1), 45-93.

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, *51*(2), 251-278.

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, *40*(3), 330-346.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*(6), 774-806.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123-147.

Bourdages, J. S. (1992). Parsing complex NPs in French. In H. Goodluck & M. Rochemont (Eds.), *Island constraints: Theory, acquisition, and processing* (pp. 61-87). Dordrecht, The Netherlands: Kluwer.

Brennan, S. E., & Williams, M. (1995). The feeling of another′ s knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383-398.

Bürkner P. C. (2017). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software. 80*(1), 1-28.

Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language*, *27*(1), 117-134.

Carpenter, B., Gelman, A., Hoffman, M., 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), 1-32.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*(4), 368-407.

Chow, W. Y., & Chen, D. (2020). Predicting (in) correctly: listeners rapidly use unexpected information to revise their predictions. *Language, Cognition and Neuroscience*, 1-13.

Corbett, G. G. (1979). The agreement hierarchy. *Journal of Linguistics, 15*(2), 203-224.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105*(3), 658-668.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114.

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., & 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.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*(2), 292-314.

Dank, M., & Deutsch, A. (2010). The role of morpho-phonological factors in subject–predicate gender agreement in Hebrew. *Language and Cognitive Processes, 25*(10), 1380-1410.

Dank, M., Deutsch, A., & Bock, K. (2015). Resolving conflicts in natural and grammatical gender agreement: Evidence from eye movements. *Journal of Psycholinguistic Research, 44*(4), 435-467.

De Villiers, J. G., Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research, 8*(5), 499-518.

De Vincenzi, M., (1991). *Syntactic parsing strategies in Italian.* Dordrecht: Kluwer.

Deane, P. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics, 2*(1), 1-64.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Deutsch, A. (1998). Subject-predicate agreement in Hebrew: Interrelations with semantic processes. *Language and Cognitive Processes*, *13*(5), 575-597.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language*, *45*(2), 200-224.

Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language*, *60*(1), 112-143.

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.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, *81*(4), 882-906.

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.

Drummond, A. (2018). Ibexfarm. www.spellout.net/ibexfarm.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, *112*(3), 531-559.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Memory and Language*, *20*(6), 641.

Erteschik-Shir, N. (1973). *On the nature of island constraints.* Doctoral dissertation, Massachusetts Institute of Technology.

Falk, Y. N., Butt, M., & King, T. H. (2009). Islands: A Mixed Analysis. In *Proceedings of LFG09*, 261-81.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75-84.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164-203.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725-745.

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

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11-15.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, *23*(5), 407-434.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 203-218.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709-738.

Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 297-316.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*(3), 329-374.

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

Frazier, L. (1987a). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, *5*(4), 519-559.

Frazier, L. (1987b). Sentence processing: A tutorial review. In Coltheart M. (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Erlbaum.

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

Frazier, L., & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331-344.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye move-ments in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, *31*(3), 661.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, *119*(1), 67-88.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200-214.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293-301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews of Neuroscience*, *11*(2), 127-138.

Futrell, R., & Levy, R. (2017, April). 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* (pp. 688-698).

Garnsey, S. M., Tanenhaus M. K., & Chapman R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 51-60.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of Learning and Motivation, Vol. 9* (pp. 133-175). New York: Academic Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Third ed.). Boca Raton, FL: Chapman and Hall/CRC Press.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*(2), 161-187.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition, 111*(1), 1-23.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain*, (pp. 95-126). Cambridge, MA: The MIT Press.

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

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science, 28*(6), 703-712.

Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject–verb agreement production. *Cognition, 118*(3), 377-397.

Givón, T. (1976). On the VS word order in Israeli Hebrew: Pragmatics and typological change. In Cole P. (ed.) *Studies in modern Hebrew syntax and semantics: the transformational-generative approach* (pp. 153-183). Amsterdam: North-Holland Publishing Company.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

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. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*(1), 97-114.

Gorrell, P. (1995). *Syntax and parsing.* Cambridge, England: Cambridge University Press.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274-279.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science, 29*(2), 261-290.

Grodzinsky, Y. (1989). Agrammatic comprehension of relative clauses. *Brain and Language, 37*(3), 480-499.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 159–166.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive science, 30*(4), 643-672.

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.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43-61.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, *8*(3), 191-219.

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language, 75*(2), 244-285.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108*(3), 831-836.

Hickok, G., Zurif, E., & Canseco-Gonzalez, E. (1993). Structural description of agrammatic comprehension. *Brain and Language, 45*(3), 371-395.

Hofstadter, D., and Moser, D. (1989). To err is human, to study error-making is cognitive science. *Michigan Quarterly Review,* 28(2), 185-215.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 417-430.

Huang Y. & Ferreira F. (2020). Is lingering misinterpretation of garden-path sentences a result of incorrect syntactic representation? Poster at CUNY Conference on Human Sentence Processing (Amherst, MA).

Igoa, J. M., García-Albea, J. E., & Sánchez-Casas, R. (1999). Gender-number dissociations in sentence production in Spanish. *Rivista di Linguistica, 11*(1), 163-196.

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia, 136,* 107291.

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

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language, 111,* 104063.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The tenten corpus family. In *7th International Corpus Linguistics Conference CL,* pp. 125-127.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 98,* 122–149.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133-156.

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.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8*(1), 63-99.

Keshev, M., & Meltzer-Asscher, A. (2019). A processing-based account of subliminal wh-island effects. *Natural Language & Linguistic Theory, 37*(2), 621-657.

Keshev, M., & Meltzer-Asscher, A., (2017). Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language, 93*(3), 549-568.

Keshev, M., Bassel, N., & Meltzer-Asscher A. (2018). Objects and self-portraits: Contrasts in the processing of reflexive pronouns across languages and structures. Poster at CUNY Conference on Human Sentence Processing (Davis, CA).

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-38.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580-602.

Kuno, S. (1976). Subject, theme, and the speaker's empathy: A reexamination of relativizatron phenomena. In C. N. Li (ed.), *Subject and Topic* (pp. 4I7-444).

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology, 5*, 1252.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kuzar, R. (2012). *Sentence patterns in English and Hebrew*. Amsterdam: John Benjamins Publishing Company.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13).

Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in Korean. *Language, 86*(3) 546-582.

Lago, S., Gračanin-Yuksek, M., Şafak, D. F., Demir, O., Kırkıcı, B., & Felser, C. (2019). Straight from the horse's mouth: agreement attraction effects with Turkish possessors. *Linguistic Approaches to Bilingualism, 9*(3), 398-426.

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.

Landau, I. (2011). Predication vs. aboutness in copy raising. *Natural Language & Linguistic Theory, 29*(3), 779-813.

Laurinavichyute, A., & von der Malsburg, T. (submitted). Semantic attraction in sentence processing. *URL https://psyarxiv.com/hk9nc*

Lee, M. W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research, 33*(1), 51-73.

Levy R., (2008a). Expectation-Based Syntactic Comprehension. *Cognition, 106*(3),1126–1177.

Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 234-243).

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-Volume 1* (pp. 1055-1065).

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086-21090.

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, *69*(4), 461-495.

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

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

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

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

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1400.

Lowder, M. W., & Ferreira, F. (2019). I see what you meant to say: Anticipating speech errors during online sentence processing. *Journal of Experimental Psychology: General*, *148*(10), 1849.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22-60.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35-54.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676-703.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50-68.

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

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

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports, 8*(1), 1-9.

Maschler, Y. (2015). Word order in time. In A. Deppermann, & S. Günthner, (eds.), *Temporality in Interaction* (pp. 201-236).

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*(6), 648-652.

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

Mendia, J. A., Poole, E., & Dillon, B. (2018, October). Spurious NPI licensing and exhaustification. In *Proceedings of SALT 28,* 233–250

Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language, 41*(2), 281-301.

Mitchell, D. C., Cuetos, F., Corley, M. M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24*(6), 469-488.

Miyamoto, E., & Nakamura, M. (2003). Subject/Object asymmetries in the processing of relative clauses in Japanese. In *Proceedings of WCCFL 22,* 342–355

Molinaro, N., Kim, A., Vespignani, F., & Job, R. (2008). Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition, 106*(2), 963-974.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Mézière, D. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife, 7*, e33468.

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, 107427.

O'Grady, W. (1997). *Syntactic development.* Chicago: University of Chicago Press.

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92-101.

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC neuroscience, 8*(1), 89.

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

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

Paspali, A., & Marinis, T. (2020). Gender Agreement Attraction in Greek Comprehension. *Frontiers in Psychology, 11,* 717.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 280.

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

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 940-961.

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469-503.

Pizarro-Guevara, J. (2020). *When human universal meets language specifics.* Doctoral dissertation, University of California, Santa Cruz.

Poppels, T., & Levy, R. (2016). Structure-sensitive Noise Inference: Comprehenders Expect Exchange Errors. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 378–383).

Pritchett, B. L. (1989). Garden path phenomena and the grammatical basis of language processing. *Language, 64*, 539-576.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. University of Chicago Press.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/.

Raab, D. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5), 574–590.

Ravid, D., (1977). Aspects of constituent ordering in Modern Israeli Hebrew. *Hebrew Computational Linguistics Bulletin, 11*, 1-45. [In Hebrew.]

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Psychology*, 39A, 657–673.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504-509.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1-23.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373.

Reuland, E. (2018). Reflexives and reflexivity. *Annual Review of Linguistics* 4(1). 81– 107.

Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: The MIT Press.

Rizzi, L. (2004). Locality and left periphery. *Structures and beyond: The cartography of syntactic structures*, *3*, 223-251.

Rohde, D. (2003). Linger: a flexible platform for language processing experiments.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348-379.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition, 181*, 141-150.

Ryskin, R. A., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition, 147*, 75-84.

Salzmann, M. (2017). Prolepsis. In M. Everaert & H. van Riemsdijk. *The Wiley Blackwell Companion to Syntax*, 2nd edition.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382-386.

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth, L. & Konieczny (eds.), *German Sentence Processing* (pp. 65-93). Dordrecht: Kluwer.

Schütze, C. T., & Sprouse, J. (2014). Judgment data. In R. J. Podesva & D.Sharma (Eds.), *Research Methods in Linguistics* (pp. 27-50). Cambridge University Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109-147.

Shlonsky, U. (1987). *Null and displaced subjects.* Doctoral dissertation, Massachusetts Institute of Technology.

Shlonsky, U. (1990). Pro in Hebrew subject inversion. *Linguistic Inquiry*, *21*(2), 263-275.

Shlonsky, U. (1997). *Clause structure and word order in Hebrew and Arabic: An essay in comparative Semitic syntax*. Oxford: Oxford University Press.

Shlonsky, U., & Doron, E. (1992). Verb second in Hebrew. In *Proceedings of the West Coast Conference on Formal Linguistics* (Vol. 10, pp. 431-446).

Siewierska, A., & Papastathi, M. (2011). Towards a typology of third person plural impersonals. *Linguistics*, *49*(3), 575-610.

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science, 14*(1), 44-48.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*(2), 104-120.

Slioussar, N. (2018). Forms and features: The role of syncretism in number agreement attraction. *Journal of Memory and Language, 101*, 51-63.

Slioussar, N., & Malko, A. (2016). Gender agreement attraction in Russian: production and comprehension evidence. *Frontiers in Psychology, 7*, 1651.

Sloggett, S. (2017). *When errors aren't: How comprehenders selectively violate Binding Theory*. Doctoral dissertation, University of Massachusetts, Amherst.

Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject–verb number agreement. *Cognitive Science, 42*, 1043-1074.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25-38.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, *49*(1), 1-46.

Sprouse, J. (2008). The differential sensitivity of acceptability judgments to processing effects. *Linguistic Inquiry*, *39*(4), 686-694.

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). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71-86.

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, *26*(1), 340-346.

Stowe, L. A. (1986). Parsing WH-construction: Evidence for online gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

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

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477-488.

Sturt, P., & Crocker, M. W. (1997). Thematic monotonicity. *Journal of Psycholinguistic Research*, *26*(3), 297-322.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, *40*(1), 136-150.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143-163.

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.

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.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

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.

Taube, D. (2007). Impersonal and passive constructions in contemporary Hebrew. In Bar, T. & Cohen E. (eds.), *Studies in Semitic and General Linguistics in Honor of Gideon Goldenberg* (pp. 277-297). Münster: Ugarit-Verlag.

Tavakolian. S. (1981). The conjoined-clause analysis of relative clauses. In S. Tavakolian (Ed.), *Language Acquisition and Linguistic Theory*, pp. 167-187.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382-392.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1266.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69-90.

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

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285-318.

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.

Tucker, M. A., Idrissi, A., & Almeida, D. (2016). Attraction effects for verbal gender and number are similar but not identical: Self-paced reading evidence from Modern Standard Arabic. *URL https://matthew-tucker.github.io/files/papers/gender-attraction-msa-comprehension.pdf*.

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., & 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. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247-263.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a Perceptual Process* (pp. 621-648).

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.

Van Valin, R. D. Jr. (1996). Toward a functionalist account of so-called 'extraction constraints'. In B. Devriendt (Ed.), *Complex Structures: A Functionalist Perspective* (pp. 29-60).

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

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

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*(11), 968-982.

Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*(4), 455-478.

Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, *45*(3), 368-390.

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

Villata, S., Sprouse, J., & Tabor, W. (2019). Modeling Ungrammaticality: A Self-Organizing Model of Islands. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1178-1184).

Wagers, M. W., Borja, M. F., & Chung, S. (2018). Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition*, *178*, 207-221.

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.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality*, pp. 119–161.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, *26*(6), 714-738.

Wechsler, S., & Zlatić, L. (2000). A theory of agreement and its application to Serbo-Croatian. *Language*, 799-832.

Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes East. *Journal of Psycholinguistic Research*, *22*, 339—364

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience letters*, *346*(3), 165-168.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356-366.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, *4*, 708.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245-272.

Appendices

Appendix A: Additional analysis for Experiment 3

In the main analysis of Experiment 3, I observed a pattern where the timing of reanalysis costs was modulated by dependency type. On the critical region, reading times in wh-questions where significantly higher than in *regarding*-dependencies, but the I did not observe a difference between the *regarding*-phrase and *with*-phrase conditions. In contrast, on the spillover region, no reliable difference was detected between wh-questions and the *regarding*-phrase condition, yet a contrast arose between the latter and the *with*-condition (with a processing cost for *regarding*-dependencies). With this additional analysis I aimed to test whether the contrast in timing is also observable as an interaction of condition and region. Thus, I included data from both regions and added the region factor (critical vs. spillover, sum coded) to the original model.

The analysis revealed a reliable interaction between condition (wh- vs. regarding-phrase) and region, such that the contrast between the wh- and regarding-phrase conditions was more pronounced in the critical word. This was observed on both frequentist and Bayesian analyses. However, the other crucial interaction term involving the contrast between the regarding- and with-phrase conditions was reliable only on the Bayesian analysis.

On pairwise comparisons (Bonferroni-corrected for six comparisons) the original pattern was retained. Both Bayesian and frequentist analyses detected a reliable increase in reading times for wh-questions relative to *regarding*-dependencies, on the critical region. At the spillover both analyses detected an increase in reading times for *regarding*-dependencies relative to the *with*-phrase condition. No other effects were detected by either analysis.

	Estimate	SE	*t*	*p*
Main effects:				
Position	0.020	0.006	3.55	< .001
With-phrase vs. if-baseline	-0.026	0.018	1.45	.15
Regarding-phrase vs. with-phrase	0.034	0.018	1.87	.06
Wh-question vs. regarding-phrase	0.069	0.018	3.80	< .001
Interaction terms:				
With-phrase vs. if-baseline	0.004	0.016	0.25	.80
Regarding-phrase vs. with-phrase	-0.030	0.016	1.88	.06
Wh-question vs. regarding-phrase	0.051	0.016	3.16	.002
Pairwise comparisons:				
With vs. if-baseline, critical word	-0.022	0.024	0.92	> .99
Regarding vs. with, critical word	0.003	0.024	0.15	> .99
Wh-question vs. regarding, critical word	0.119	0.024	4.94	< .001
With vs. if-baseline, spillover	-0.030	0.024	1.25	> .99
Regarding vs. with, spillover	0.064	0.024	2.64	.049

Wh-question vs. regarding, spillover	0.018	0.024	0.75	> .99

Table A1. Results of the frequentist region interaction analysis: Estimate, standard-error, t-value and (Bonferroni-corrected) p-value of the tested contrasts.

	Posterior mean and CrI (ms)	Probability of the posterior beyond zero
Main effects:		
Position	10 [2, 18]	99%
With-phrase vs. if-baseline	-13 [-33,6]	90%
Regarding-phrase vs. with-phrase	17 [-5, 38]	93%
Wh-question vs. regarding-phrase	34 [11, 56]	> 99%
Interaction terms:		
With-phrase vs. if-baseline	2 [-16, 19]	59%
Regarding-phrase vs. with-phrase	-15 [-31, 1]	96%
Wh-question vs. regarding-phrase	25 [7,44]	> 99%
Pairwise comparisons:		
With vs. if-baseline, critical word	-11 [-38, 15]	80%
Regarding vs. with, critical word	2 [-25, 28]	55%
Wh-question vs. regarding, critical word	59 [31, 88]	> 99%
With vs. if-baseline, spillover	-15 [-41,11]	87%
Regarding vs. with, spillover	31 [5, 59]	99%
Wh-question vs. regarding, spillover	9 [-20, 38]	73%

Table A4. Results of the Bayesian Surprisal analysis: Posterior means of the tested contrasts (with 95% Bayesian credible intervals) and the probability of the parameter being beyond zero.

Appendix B: Results of the Experiment 4 when offsetting for a timing contrast

After conducting the analysis which appears on the main text (section 3.6), I wanted to take into account the possible contrast in timing of reanalysis costs, as suggested by Experiment 3. It is possible that reanalysis is observable one region earlier for wh-questions. Therefore, I conducted an additional analysis where the filled-gap verb was considered the critical region of wh-question conditions, but for *regarding*-dependencies the critical region was the word following that verb. Spillover results were adjusted accordingly, using the first word after the verb for wh-questions, and the second word after the verb for regarding-dependencies. The effects on these (non-matching) regions were analysed in two additional models (with the same contrast coding and analysis protocol).

This analysis might also be more reliable since in this case, the spillover region, where more effects were observed, has the same lexical material in both *regarding*-phrase conditions and wh-questions. Since the regarding dependency requires a pronoun while filler-gap dependencies do not, immediately following the critical verb, the conditions diverged for one word – a pronoun appeared in regarding-phrase conditions, and the first word of the final PP in wh-questions, Thus, shifting the analysis of the *regarding*-condition's spillover by one word also realigns the lexical content at that region. See Figure B1.

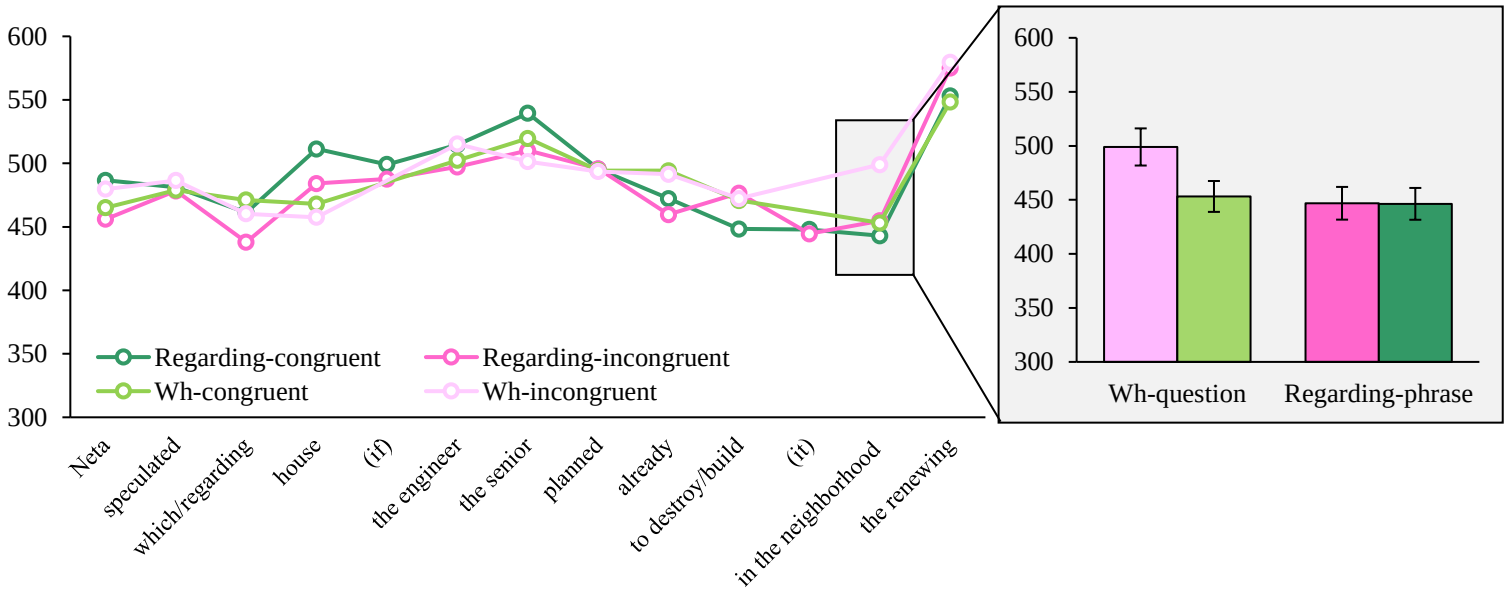

Figure B1. Word-by-word RT means (ms) by condition, Experiment 4. Error bars represent +/-SE.

Critical Region

As in the initial analysis, at the critical region no effects were reliable. The only effect detected was a main effect of dependency type, such that wh-questions were read slower than *regarding*-dependencies. This was observed on the Bayesian analysis (posterior mean and CrI: -10ms [-20, 0], with 97% of the sampled posterior beyond zero). However, the effect was not significant in the frequentist analysis (Estimate = -0.016, SE = 0.010, t = 1.63, p = .11).

Spillover Region

The analysis of the amended spillover region exhibited the same patterns as those reported in the main text, with only one difference: On this analysis the crucial interaction was reliable also in the Bayesian analysis. The results of the frequentist and Bayesian analyses are available in Tables B1 and B2, correspondingly.

	Estimate	SE	*t*	*p*
Main effect of dependency	-0.045	0.008	5.89	< 0.001
Main effect of congruency	-0.021	0.008	2.75	.006
Interaction	0.019	0.008	2.52	.012
Pairwise semantic persistence comparisons:				
Wh-question	-0.081	0.022	3.73	< 0.001
Regarding-phrase	-0.004	0.022	0.19	> .99

Table B1. Results of the frequentist analysis of Experiment 4, at the spillover region: Estimate, standard-error, *t*-value and (Bonferroni-corrected) *p*-value of the tested contrasts.

	Posterior mean and CrI (ms)	Probability of the posterior beyond zero
Main effect of dependency	-13 [-21, -5]	> 99%
Main effect of congruency	-9 [-17, -1]	99%
Interaction	9 [0, 17]	97%
Pairwise semantic persistence comparisons:		
Wh-question	41 [15, 66]	> 99%
Regarding-phrase	12 [-13, 36]	82%

Table B2. Results of the Bayesian analysis of Experiment 3: Posterior means of the tested contrasts (with 95% Bayesian credible intervals) and the probability of the parameter being beyond zero.

Appendix C: Results of by-region analyses of Experiment 8A

Experiment 8A was initially analyzed with two by-region models (one for the critical word and one for the post-critical one). I subsequently explored the interaction between region and condition and conducted a unified analysis (with data from both regions) as reported in the main text. Here I report the results of the two by-region models, which are broadly in line with those of the unified model.

	Estimate	SE	*t*	*p*
Model 1, critical region:				
Match – Baseline	0.013	0.021	0.62	> .99
Mismatch – Baseline	0.064	0.022	2.91	.010
Model 2, spillover region:				
Match – Baseline	0.097	0.020	4.94	< .001
Mismatch – Baseline	0.046	0.019	2.40	.033

Table C1. Results of by-region frequentist analysis of Experiment 8A: Estimate, standard-error, *t*-value and *p*-values (corrected, within each model, using Bonferroni correction for two comparisons).

	Posterior mean [CrI]	Posterior beyond zero
Model 1, critical region:		
Match – Baseline	-7 [-24, 11]	79%
Mismatch – Baseline	17 [-3, 38]	95%
Model 2, spillover region:		
Match – Baseline	28 [11, 46]	> 99%
Mismatch – Baseline	11 [-5, 27]	91%

Table C2. Results of by-region Bayesian analysis of Experiment 8A: Posterior means of the tested contrasts (with 95% Bayesian credible intervals) and the probability of the parameter being beyond zero.

Appendix D: Coding criteria used in the analysis of Experiment 10

In the coding of production data from Experiment 10, sentences were considered SR completions if they included one of the following:

1. A (definite) NP marked for accusative case (82% of SR completions).
2. An (indefinite) NP which was unmarked for case but mismatched the verb (11% of SR completions).
3. A verbal argument of another category (e.g. a PP) which did not include a resumptive pronoun (obligatory in indirect objects in Hebrew) and was not grammatically compatible with a direct object gap (3% of SR completions).

Sentences were considered as OR completions if:

1. The sentence included a (definite) NP marked for nominative case (22% of OR productions).
2. The sentence was missing an obligatory direct object (24% of OR productions).
3. The sentence included a corresponding resumptive pronoun which was grammatically inconsistent with another gap in subject position (53% of OR productions).

I aimed to refrain from reading into possible typos and/or assuming the likelihood of different readings. Therefore, completions which did not match these criteria were coded as ambiguous productions. This included sentences with an ambiguous structure (e.g. with indefinite noun phrases, unmarked for case, that match the verb; with resumptive pronouns in possessive positions, allowing an additional subject gap), and completions which were partial or contained errors unrelated to the filler's agreement features. For grammatically ambiguous productions, an exception was made when the three coders judged that the production was semantically plausible only in one of the readings (these cases constituted 4% from the SR coded data and less than 1% of the OR coded data).

The analysis counted ambiguous productions as non-SR productions along with the OR ones, to keep conservative estimates of SR productions and to allow variance in the Match-Verb.SG condition (where OR productions were extremely sparse). Yet, a similar pattern of results can be observed when including the ambiguous responses (see Figure B1).

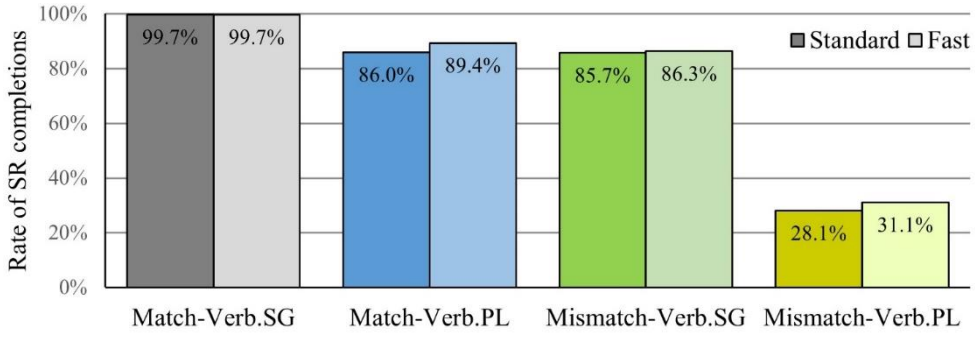

Figure D1. Rate of subject relative completions by condition in Experiment 10, when excluding all ambiguous and partial responses.

תקציר

עבודה זו חוקרת כיצד קוראים מתמודדים עם טעויות לשוניות בפירוש, זיכרון, תפיסה והפקת משפטים. אני בוחנת נושא זה בשלושה מחקרים נפרדים, אשר מתמקדים בסוגיות מרכזיות בספרות על עיבוד משפטים: הקושי של שינוי הפירוש הראשוני ושיקולי העלות-תועלת בחיזוי המשך המשפט; הפגיעות של הקידוד בזיכרון והשליפה ממנו להפרעות ממרכיבים דומים; והמנגנונים הזמינים לצורך פיצוי על קלט רועש או מטעה.

במחקר הראשון, אני בוחנת איך קוראים מאזנים בין היתרונות של חיזוי ובין הסיכון שבטעויות. כשאנחנו מבצעים ניחושים (מושכלים) בכל צעד עיבודי, אנו בהכרח נכשלים לעיתים. פעמים רבות אנחנו "מהמרים" על הפירוש הנכון. אבל פעמים אחרות נצטרך לחזור על עקבותינו ולשנות את הפירוש הראשוני. כיצד קוראים מתגברים על הקושי של פירוש שגוי בשלב מוקדם של העיבוד?

על מנת לבדוק זאת, אני בוחנת את העלות של חיזויים כושלים שהונעו משיקולים שונים: תיקוף המבנה התחבירי, אינטגרציה סמנטית מהירה, או התאמה להעדפות פרגמטיות. אני מוצאת עדות לחיזוי של תלויות ארוכות טווח גם בהעדר לחצים תחביריים. יחד עם זאת, אני מוצאת כי חיזוי שמונע על ידי לחצים תחביריים מוביל לקושי מוקדם יותר ובולט יותר במקרים של כישלון החיזוי. בנוסף, המשמעות הסמנטית של הפירוש המקורי נותרת חזקה יותר לאחר ניתוח מחדש של תלות שנחזתה על בסיס תיקוף תחבירי. אני טוענת שההבדל בעלויות הניתוח המחודש טמון ברמות של מחויבות לחיזוי. אני מציעה שיצירה של תלויות היא מהירה יותר או מחייבת יותר כשמדובר בחיזוי מבוסס-תחביר ביחד לחיזוי מבוסס-פרגמטיקה.

במחקר השני, אני סוקרת הפרעות בזיכרון העבודה במהלך עיבוד משפטים, תוך התמקדות באשליות התאם ובמנגנונים שעשויים לאפשר פיצוי על טעויות מסוג זה. טעויות ביחסי התאם, שנובעות מהפרעה לייצוג או לשליפה מהזיכרון, הן תופעה נפוצה. על כן, ייתכן כי היה יעיל עבור המעבד להשתמש בנקודות ביניים במשפט (רכיבים נוספים שמייצגים את ההתאם של הנושא למשל) על מנת לצמצם את הצורך בגישה לרכיבים מוקדמים. על מנת לבחון את האפשרות הזו, אני בודקת את העיבוד של כינויי גוף רפלקסיביים, כינויי גוף שדורשים זיהוי של נושא המשפט ותכוניות ההתאם שלו. שימוש בתכוניות ההתאם של הפועל (שנמצא קרוב יותר לכינוי הגוף הרפלקסיבי) יכול להוות מעין קיצור דרך או נקודת ביניים לעדכון תכוניות ההתאם, מכיוון שהפועל אמור (דקדוקית) לשאת את אותן תכוניות כמו הנושא.

באמצעות התאם מגדרי על פעלים וכינויי גוף בעברית, אני מראה שהחישוב של תכוניות התאם בכינוי הגוף הרפלקסיבי מסתמך במידה רבה על הזמינות של סמני התאם פעלי. אני מציגה שני ממצאים מרכזיים: (א) כשהפועל לא תואם לנושא, קוראים מעדיפים כינוי גוף רפלקסיבי שתואם לפועל על פני כינוי גוף שתואם לנושא; (ב) אשליות התאם מתרחשות פחות כאשר הפועל נושא תכוניות התאם. אני מפרשת זאת כעדות

לכך שהתאם פעלי משנה את הייצוג של הנושא או מייתר את השליפה של תכוניות הנושא (כינוי הגוף הרפלקסיבי אינו זקוק לגשת לנושא כאשר ההתאם הפעלי זמין, אלא מוודא התאמה עם הפועל בלבד). אסטרטגיות אלה עשויות להשתלם למעבד מכיוון שהן מאפשרות לו להפחית את ההסתמכות על הייצוג בזיכרון של חלקים מוקדמים של המשפט.

במחקר האחרון, אני מציגה עדות חדשה להיסק רציונלי המבוסס על הנחות ערוץ-רועש. אני משתמשת במשפטי זיקה בעברית בהם עמדת הנושא פנויה והפועל לא תואם לראש הזיקה בתכונית המספר שלו. כך, בפועל, המעבד יכול לבנות זיקת נושא, תוך התעלמות מחוסר ההתאמה, או לבחור במבנה פחות נפוץ של זיקת מושא עם סדר מילים נדיר (נושא פוסט-פעלי) או תחת מבנה נושא ריק (א-פרסונאלי).

אני מראה שההשפעה של חוסר ההתאמה תלויה בשכיחות של המבנים האלטרנטיביים (ההסתברות הפריורית). אני מוצאת עדות לכך שקוראים מעדיפים פירוש בדמות זיקת נושא רועשת (עם טעות התאם) על פני פנייה לסדר מילים נדיר. הדבר מראה ש קוראים מוכנים להתעלם ממידע כמו התאם מספרי בין הנושא והפועל. יחד עם זאת, אני מוצאת גם עדות לכך שקוראים מנסים להישאר מהימנים לקלט הלשוני כאשר הניתוח הדקדוקי אינו נדיר במיוחד. כך, קוראים מעדיפים מבנה תקין דקדוקית ונפוץ יחסית (נושא ריק) על פני מבנה נפוץ עוד יותר אך קלוקל (זיקת נושא עם טעות התאם). ממצאים אלה מעידים שבמהלך עיבוד משפטים, אנו משתמשים בזריזות במידע הסתברותי מפורט אודות התפוצה של מבנים תחביריים שונים. כך, קוראים מוכנים לבנות ייצוג לא דקדוקי אך עושים זאת לא בשל עיבוד שטחי אלא כאסטרטגיה רציונלית שמטרתה הימנעות ממבנים בעלי סבירות נמוכה במיוחד.

לסיכום, הממצאים המוצגים בעבודה זו מציגים מערכת מורכבת של איזונים ובלמים שעשויה לסייע בפירוש מוצלח של משפטים, בהינתן האפשרויות של טעויות בעיבוד או בקלט. כך, המחקרים המוצגים כאן תורמים להבנה של מערכת התקשורת האנושים ומציעים דרכים בהן המערכת עשויה לפצות (בזמן אמת) על כשלים שונים בעיבוד שפה.

אוניברסיטת תל אביב
הפקולטה למדעי הרוח על שם לסטר וסאלי אנטין
בית הספר לפילוסופיה, בלשנות ולימודי מדע

ניתוח שגוי של תלויות תחביריות ועיבוד מחדש בתהליך הבנת המשפט

חיבור זה הוגש לשם קבלת התואר "דוקטור לפילוסופיה"

על ידי

מעין קשב

העבודה הוכנה בהדרכת:

פרופ' איה מלצר-אשר

נובמבר 2020