

INCREMENTAL PROCESSING EFFECTS IN NOMINAL COMPOUNDS

By

Alicia Parrish

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Linguistics – Master of Arts

2017

ABSTRACT

INCREMENTAL PROCESSING EFFECTS IN NOMINAL COMPOUNDS

By

Alicia Parrish

Sentence processing shows the effects of a series of continual building, repairing, predicting, accessing, and remembering operations that may be the output of one underlying process or many. Within even smaller phrases such as nominal compounds, we see all of these operations having an effect. What is relatively unknown, though, is how these processes interact with each other in real time as the phrase is build up incrementally. This ERP study, through the use of an Icelandic triple noun compound paradigm that manipulates agreement features on the first and second constituents of the compound, investigates the nature of commitments to a structure and the processes that predict more structure or revise an interpreted structure. The findings are generally in line with models that have a parser make commitments to a structure as soon as possible, and the findings expand on that to say that syntactic mismatch is sufficient to trigger a structural prediction and that a revision of that prediction is identical to the revision of a structure built from incoming lexical items. This study further uses the paradigm to assess the predictions of Gibson's (1998) model of sentence processing that makes use of working memory costs. The study finds that, when incremental commitments are taken into account, we see the effects of syntactic agreement cues modulating working memory effects.

Copyright by
ALICIA PARRISH
2017

ACKNOWLEDGMENTS

I would first like to thank my supervisor, Alan Beretta, and our colleague at the University of Iceland, Matthew Whelpton, for their enthusiasm at the earliest stages of this project's design. Alan's consistent encouragement inspired me to continually look at the data in new ways and to push further, even when I was tired of looking at what seemed like the exact same graph 100 different ways. The experiment itself couldn't have been carried out without Matthew who, never without a smile, did all the work of organizing logistics and managing an emerging lab group in Iceland who created all of my stimuli.

Along with Alan and Matthew, I am also indebted to the many lab members at Michigan State who gave substantial input in the design and analysis phases of this project: Karthik Durvasula, Joseph Jalbert, Kaylin Smith, Drew Trotter, Patrick Kelley, and Brian Pinsky. In addition to the helpful input, Kaylin was also instrumental in carrying out the experiment with me that is the main focus of this study. I'd also like to thank my entire thesis committee, Alan, Karthik, and Suzanne Wagner, for their comments on my proposal for this project.

I'm also very thankful for the countless hours of help in creating the Icelandic sentences through many revisions, and the patience of the lab members at the University of Iceland as they time and again explained the nuances of Icelandic agreement and compounds to me. Those people are, again, Matthew Whelpton, Þórhalla Guðmundsdóttir, and Bjarni Barkarson who helped with the sentences, and additional help in managing the lab work from Lilja Bjork, Alec Shaw and Tatiana Kantorovich.

This is certainly a project I couldn't have done on my own, so I'd also like to express a general thank you to everyone that I discussed this work with in passing, or in depth. This includes the audience of GLEAMS 2017 and, in particular, Ellen Lau for the helpful conversation afterwards that provided guidance in how to tell a much more focused story from the data.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER 1 INTRODUCTION	1
1.1 Basic info about nominal compounds in Germanic languages	2
1.2 Existing studies on the processing of compounds and complex phrases	3
CHAPTER 2 AN AGREEMENT MANIPULATION EXPERIMENT OF ICELANDIC TRIPLE NOUN COMPOUNDS	7
2.1 The motivation for this study	7
2.2 The paradigm	8
2.3 Methods	14
2.3.1 Participants	14
2.3.2 Materials	14
2.3.3 Procedure	17
2.3.4 Data analysis and recording	18
2.4 Pretests	19
CHAPTER 3 RESULTS AND DISCUSSION OF THE TRIPLE NOUN STUDY AS THEY RELATE TO COMMITMENT AND REVISION	21
3.1 Replication of previous findings	21
3.2 Extension of previous findings	25
CHAPTER 4 RESULTS AND DISCUSSION OF THE TRIPLE NOUN STUDY AS THEY RELATE TO WORKING MEMORY	29
4.1 The sustained anterior negativity (SAN) and working memory	31
4.2 The importance of working memory’s interaction with models of commit- ment and revision	33
4.3 Results across constituent places	33
4.3.1 The agree-agree condition	36
4.3.2 The nonagree-agree condition	39
4.3.3 The agree-nonagree condition	41
4.3.4 The nonagree-nonagree condition	43
4.4 Comparisons of the patterns seen in the different conditions	45
CHAPTER 5 NEXT STEPS AND CONCLUSION	49
5.1 What’s next	49
5.2 Conclusion	51
APPENDICES	52
APPENDIX A SAMPLE EXPERIMENTAL ITEMS WITH GLOSSES	53

APPENDIX B	SET OF SAMPLE FILLER ITEMS	55
APPENDIX C	SET OF SAMPLE LLICIT PRE-TEST ITEMS	56
APPENDIX D	TABLE OF <i>F</i> -VALUES FOR ALL COMPARISONS	57
APPENDIX E	LONG WINDOWS	58
BIBLIOGRAPHY	60

LIST OF TABLES

Table 1:	Diagram of four conditions used in this study	9
Table 2:	Predictions of processing steps for all conditions	9
Table 3:	Counts for the inflectional features on nouns used for experimental items .	16
Table 4:	Mean scores given on the acceptability judgment task for selected stimuli	20
Table 5:	Predictions from Gibson (1998) for the I(n) cost in the Agree-Agree condition	36
Table 6:	Predictions from Gibson (1998) for the I(n) in the Nonagree-Agree condition	39
Table 7:	Predictions from Gibson (1998) for the I(n) in the Agree-Nonagree condition	41
Table 8:	Predictions from Gibson (1998) for the I(n) in the Nonagree-Nonagree condition	43
Table 9:	Table of significant values (F values shown) for each time window, com- parison, and condition	46
Table 10:	Summary of the total integration cost at each constituent in each condition	47
Table 11:	F-values for all comparisons	57

LIST OF FIGURES

Figure 1:	Diagram of assumed structure of the compound in sentences in (2a) . . .	16
Figure 2:	Electrode groupings for the five regions of interest	19
Figure 3:	Effect of agreement of N1 measured on N1	22
Figure 4:	Effect of Agree-Agree versus Nonagree-Agree measured on N2	23
Figure 5:	Effect of N2 agreement measured on N3	24
Figure 6:	Comparison of Agree-Nonagree and Nonagree-Nonagree conditions at N2	25
Figure 7:	Comparison of Agree-Agree and Nonagree-Agree condition on N3	27
Figure 8:	Differences across noun position with all conditions averaged together . .	34
Figure 9:	Differences across noun position in the Frontocentral-Left region	35
Figure 10:	Waveforms of N1, N2, and N3 within the Agree-Agree condition	38
Figure 11:	Waveforms of N1, N2, and N3 within the Nonagree-Agree condition . . .	40
Figure 12:	Waveforms of N1, N2, and N3 within the Agree-Nonagree condition . . .	42
Figure 13:	Waveforms of N1, N2, and N3 within the Nonagree-Nonagree condition .	44
Figure 14:	Long window showing each electrode for each condition	58
Figure 15:	Long window showing the five ROIs for each condition	59

CHAPTER 1

INTRODUCTION

Sentence processing is a complex action that must dynamically integrate lexical retrieval, syntactic parsing, and attentional resources. Sentence comprehension, which must be built up incrementally as sounds that compose words are encountered one after another (as opposed to all at once), is a process that, even in healthy adults, remains imperfect. From the more phonological Freud and Brill's (1938) 'slips of the tongue' and standard speech errors (Fromkin, 1971) to the more syntactic errors found in garden path (Osterhout, Holcomb, & Swinney, 1994) and Escher sentences (Montalbetti, 1984), the need for repair and revision can affect both production and comprehension. Questions of at what point in the sentence a given word or phrase gets interpreted and when and how that interpretation might need to be revised are fundamental to psycholinguistic studies. And yet relatively little is known about the way sentences are processed in the brain, and littler still about how the building and repairing processes interact and function in real time.

In addressing questions about sentence processing, this study exploits nominal compounds as a way of investigating processes that work across a sentence (and possibly across discourse) to manipulate features necessary in processing on a smaller, more controlled scale. The following sections describe some background information of compounds in Germanic languages, as well as provide an overview of the relevant literature on compound processing. Chapter 2 describes the experiment that is the main focus of this paper. Chapters 3 and 4 describe the results of two very different analyses of that study as they relate to (i) models of commitment and revision in processing and (ii) models of the importance of working memory in sentence processing, respectively. Finally Chapter 5 speculates on some future studies that may add relevant information and nuance to the conclusions drawn in the previous chapters.

1.1 Basic info about nominal compounds in Germanic languages

To attempt to answer some of the open questions in psycholinguistics, several researchers have focused on the processing of noun-noun compounds because such studies can allow for an investigation of modification, heads, and gender/number agreement (among other important aspects) while controlling for possible word category effects. Questions about the nature of sentence comprehension can be posed on a smaller, phrase-level scale which allows researchers a great deal of control over the stimuli used.

In considering the terminology for discussing noun-noun compounds, this study uses the terms *modifier* and *head* to describe the roles of each noun in a noun-noun compound. For example, in a compound such as *table lamp*, we would say that *lamp* is the compound's head (what the item *is*) and *table* is the modifier (it tells you what kind of lamp).

Overall, Germanic languages tend to have a very productive process of creating nominal compounds. That is, such compounds can be created as novel forms and are easily understood within a discourse. Additionally, such compounds are potentially infinite (but in the same way that sentence embedding is infinite, where processing still constrains how many embeddings we can actually follow and still understand). For example, “college student council committee report” consists of five nouns, and is perfectly sensible. One could even add “press conference interview” after it, creating an eight noun compound that, while becoming a bit more difficult, is still sensible, and likely understood to be an interview at a press conference about the committee report put out by the student council at a college.

In Germanic languages that have gender and number agreement, a determiner would agree with only the head noun, the right-most constituent in these language. Given the previous example of the five-noun compound in English, in the German phrase, a determiner such as “the” would have to agree in gender and number with “report”, regardless of the gender and number features on the other nouns. In the eight noun compound, “the” would agree with “interview”, though.

Furthermore, the case of the determiner and the head must also match. Modifiers often

take either genitive or bare forms for case. However, due to the high degree of syncretism in the Icelandic inflectional system, there are many cases where a root form appears the same as, for example, the accusative form. It can also be the case that the determiner has the same form in the genitive and the accusative, for example. The importance of this syncretism for agreement, and how it was used in this study will be discussed in more detail in the following chapters.

1.2 Existing studies on the processing of compounds and complex phrases

Past theoretical literature has focused on the role of the compound's head in determining its interpretation. We see that in languages that have gender and number agreement (e.g., German, Icelandic), the determiner's agreement features come from the head noun rather than the modifier noun. Scalise and Guevara (2006) argue that the difference between transparent and opaque compounds also lies in the head: when there is alignment of the syntactic and semantic head, the compound is transparent; the modifier's contribution to meaning does not necessarily play a role in the compound's transparency.

This study focuses exclusively on transparent compounds, as there is evidence that additional or different processing mechanisms may be engaged for the processing of opaque ones. It is, however, important to keep in mind that all nominal compounds have "inherent ambiguity" (Jackendoff, 2009), and can be assigned effectively infinite meanings once we consider the role of that context can play. Downing (1977) provides a classic example of this openness to context in her discussion of *apple juice seat* to refer to a seat on an airplane that happens to have an apple-flavored juice box in front of it. Additionally, this study will avoid, where possible, highly frequent compounds, as there is evidence that those may be lexicalized, and thus an additional or different lexical retrieval mechanism may be at play (Sandra, 1990; Zwitserlood, 1994).

The existing neurolinguistic research on nominal compound processing has found evi-

dence for automatic morphological decomposition in transparent compounds (Fiorentino & Poeppel, 2007) (results for opaque compounds have been mixed). This automatic decomposition may be affected by semantic composition at different steps as well, though. Koester, Gunter, Wagner, and Friederici (2004) found that gender mismatches on constituents in noun-noun compounds in German elicited a left anterior negativity (LAN), often associated with morphosyntactic processing or errors. This was found on both the modifier and the head constituent. When the mismatch was on the head, though, there was an additional positivity that occurred at the offset of that head.

In a test of morphosyntactic decomposition, (Koester, Gunter, & Wagner, 2007) used an auditory task to look at the effects of gender incongruencies in German compounds. In German, the head noun is the rightmost constituent in a compound and what the determiner or adjective that comes before it must agree with in gender and number, as it would with a single noun. They modulated the gender agreement of the first and second nouns in compounds presented with a determiner during an ERP study with auditory presentation of stimuli and found that a gender mismatch in elicited a LAN. This effect was relatively small when there was a mismatch on the first constituent, and larger when the mismatch occurred on the second constituent. Because gender information on the first constituent is irrelevant to the grammaticality of the phrase in German nominal compounds, Koester et al. (2007) argue that the presence of a LAN indicates that participants accessed the gender information on the non-head constituent (i.e., there was a process of automatic morphosyntactic decomposition of the constituents).

Later studies, though, have found that this effect can be modulated with the added input of non-ambiguous prosodic cues (Koester, 2014; Isel, Gunter, & Friederici, 2003). There is also recent evidence from the language production literature to support the claim for full lexical decomposition of compounds in German, as a recent production task by Lorenz, Madebach, and Jescheniak (2017) demonstrated facilitation effects for gender congruency of a determiner with both modifier and head constituents for novel compounds. These exper-

iments in German, however, were primarily auditory. When prosodic cues are unavailable, as in a reading task, the parser would only have access to syntactic or semantic cues to determine if a noun encountered following the determiner is a head or a modifier.

One such reading study was done in German by Jalbert, Roberts, and Beretta (2016). Varying gender agreement between the determiner and the modifier and head nouns, they found broad left negativities on the head noun for cases where the modifier had agreed in gender with the determiner. They interpreted these finding in line with past studies that also found higher processing effects for cases where the modifier noun allowed for a grammatically licit parse if the sentence were to end there: the parser makes a commitment as early as possible and, when later information requires such a parse to be reanalyzed, there is a processing cost.

Similar findings had previously been found in Icelandic in full sentence contexts (Whelpton et al., 2014). While parsing a sentence, there is an intermediate point in the compound where only the modifier has been encountered, and so the parser does not yet know whether they have a single noun, or a modifier to a nominal compound. Whelpton et al. (2014) used a condition where, because of the gender mismatch on an initial noun, the parser had to build in the syntax for a second noun on that first constituent. In this “extra syntax” condition, they found that mismatched conditions elicited a large P600, consistent with the need to do extra syntactic work at that point.

The findings of Whelpton et al. (2014) and Jalbert et al. (2016) suggest that a parser commits to its syntactically possible head (even when it turns out later to be a modifier), and then must revise that parse once the true head is arrived at. This is in line with many other studies, and is what I will refer to more generally as the “commitment and revision” style model.

In addition to syntactic cues, semantic plausibility may also play a role in this model of commitment and revision. A sentence that is semantically implausible will also show an effect of this violation at the modifier noun in ERPs (Koester, Holle, & Gunter, 2009;

Kutas & Hillyard, 1980) and eye tracking (Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007). Parrish, Jalbert, and Beretta (2015) also found an effect of revision at the head noun in an English compound study that varied the semantic plausibility of the first constituent. Though some have suggested this effect can be attenuated in compounds when the implausible noun is a commonly occurring or predictable modifier (Kennison, 2005), even in these cases others have measured a violation (E Pratarelli, 1995). One explanation for these findings is that the parser commits to any syntactically possible head as soon as it can, leading to the anomalous interpretations. Another interpretation is that the parser uses this semantic information to decide whether what lexical items have been encountered can be committed to, with semantically illicit structures never eliciting commitment. These ERP studies have consistently measured an N400 for this effect of semantic implausibility, which is likely explained by the semantically implausible noun being a highly unpredicted noun in that context (Lau, Almeida, Hines, & Poeppel, 2009) due to the verb's selectional features or other contextual cues.

The implications of this model of automatic commitment to a head and the process of revision are central to the questions posed in this study. Furthermore, these processes and their interaction with cognitive processes more generally will be analyzed, as committing to a structure, predicting further structure, and revising something that had been committed to all have effects on the working memory resources needed in comprehending that sentence.

CHAPTER 2

AN AGREEMENT MANIPULATION EXPERIMENT OF ICELANDIC TRIPLE NOUN COMPOUNDS

2.1 The motivation for this study

The model of head commitment and revision that is suggested from the results of several studies discussed in the previous chapter would state that a commitment is made when a parser arrives at a noun that *can* be interpreted as the head of its phrase. Such an interpretation is possible whenever such a pairing is semantically and syntactically licit. This theory can be thought of as an application of early attachment models on a smaller scale, and it is generally in line with other models that make use of the need to revise a structure that had been committed to.

The present study and many others that came before it have relied on nominal compounds as a tool for measuring the effects of different syntactic operations as they occur incrementally. Specifically with nominal compounds, when the first noun is encountered, it is interpreted as the head of the noun phrase, as the parser does not know another noun is coming (leaving aside cases with strong prosodic cues). This is because, given the sentence *John set down the table lamp*, there is a point in hearing or reading the sentence that only *John set down the table* has been presented. Such a sentence is both semantically plausible and syntactically possible, so the parser would commit to *table* as being the head of a noun phrase and to being the argument of the verb. Additionally, the parser would integrate the noun at that point with the determiner, and then integrate that DP with the verb phrase. Thus when the second noun is presented, the parser must reanalyze its earlier commitments.

However, earlier studies that used noun-noun compounds could only look at a partial paradigm. In order to measure differences between grammatical phrases, the N2 of any experimental item had to agree with its determiner, leaving only a variation on N1 as a possible

experimental manipulation. If they had varied N2 to be mismatched with a determiner, for example, then the results would have been potentially confounded by the ungrammaticality of a mismatching head noun, an effect that we already know will have processing costs due simply to the unacceptability of the sentence. The question of what happens when a predicted structure is rejected and of how continued effects of revision or prediction affect parsing remained unable to be addressed. Using triple noun compounds solves some of these issues because it allows for us to use a full 2 x 2 design of gender agreement on the first two nominal constituents, while still using only grammatical sentences as experimental items.

Thus the main strength of a study with triple noun compounds is that it allows us to extend previous findings while replication is built in to the study, and it also allows us to ask new questions about the nature of prediction in parsing.

2.2 The paradigm

This study makes use of triple noun compounds in Icelandic to create a 2 x 2 paradigm of agreement on the first and second nouns of the compound (N1 and N2). The explanation for each of the conditions is shown below (1).

- (1) Four conditions used within a quad
 - a. Agree-Agree: The first and second constituent could each be heads of their phrase with the given determiner
 - b. Nonagree-Agree: The second but not the first constituent can be the head of its phrase with the given determiner
 - c. Agree-Nonagree: The first but not the second constituent can be the head of its phrase with the given determiner
 - d. Nonagree-Nonagree: Neither the first nor the second constituent can be the head of its phrase with the given determiner.

As a visual way to understand where agreement is happening, in an actual triple noun compound, Table 1 provides a sketch of the type of pattern being used. For ease of comprehension, lexical items agreeing with the adjective or determiner are colored green, and those that mismatch are colored red.

Table 1: Diagram of four conditions used in this study

nýrra	N1	N2	N3	condition
"	hafra oat	grjóna grain	seyða broths	Agree-Agree
"	bygg barley	grjóna grain	seyða broths	Nonagree-Agree
"	hafra oat	mjöls flour	seyða broths	Agree-Nonagree
"	bygg barley	mjöls flour	seyða broths	Nonagree-Nonagree

Based on the previous studies of compound processing, we expect to see effects of prediction and revision. While certainly an oversimplification, using just three possible operations to define what processes we expect to happen at each constituent in each condition is useful in laying out the predictions for this particular study. The specific predictions that we are making with regards to these three possible steps (commitment, prediction, and revision) are diagrammed in Table 2.

Table 2: Predictions of processing steps for all conditions

Agree-Agree		N1	N2	N3
	Steps	commit	revise & commit	revise & commit
Nonagree-Agree		N1	N2	N3
	Steps	predict	commit	revise & commit
Agree-Nonagree		N1	N2	N3
	Steps	commit	revise & predict	commit
Nonagree-Nonagree		N1	N2	N3
	Steps	predict	revise & predict	commit

For the Agree-Agree condition, we expect to see initial commitment on the first noun as it is a licit head noun for the DP. Once N2 is encountered, the parser must revise that first commitment to N1 and make a new commitment to the [N1 + N2] compound, as N2 is also a syntactically licity head. Finally, once N3 is encountered by the parser, there is again revision followed by commitment for the same reasons that these steps were needed when N2 was encountered.

In the Nonagree-Agree condition, however, we see the need for prediction. When N1 does not match in case, gender, and number with its determiner, the only way for the parser to maintain a grammatical structure is to conclude that this constituent is actually a modifier noun in a nominal compound. Thus N2 is predicted from the grammatical mismatch present on N1 in this condition. Once N2 arrives, the parser has all the features needed to create a syntactically licit structure, so it commits to N2 as the head of the phrase. Once N3 arrives, it again needs to revise its commitment to N2 as a head and re-commit to N3 as the head of the phrase.

In the Agree-Nonagree condition, the processing at N1 is identical to that of the Agree-Agree condition, where the parser simply commits to N1 as the head of the NP. Once the parser arrives at a mismatched N2, though, it must revise its commitment to N1 as the head. However, in this case the parser is not able to commit to N2, as its features are not a match for those of the determiner. Therefore, a third noun is predicted in order to have a syntactically licit structure. Once N3 appears, the work of building it in to the structure has already been done at N2, and thus the only step needed here is for the parser to commit to N3 as the head of the NP and integrate it with the determiner.

Finally, for the Nonagree-Nonagree condition, the processing step on the first constituent will be to predict another noun, just as it was in the Nonagree-Agree condition. At N2, the prediction of a noun that can be integrated is not met, so the parser needs to revise the predicted structure and again predict another noun. Once N3 is encountered, all mismatches are resolved and the parser can commit to N3 as the head of its phrase.

One particularly important note here is that this model of prediction takes the very strong form, that structure with node labels and features are all able to be predicted. Much of the existing literature on prediction focuses on lexical prediction (Lau et al., 2009; Kutas & Federmeier, 2011, 2000; DeLong, Urbach, & Kutas, 2005, among others) rather than structural prediction. Indeed, even if structure is predicted, that is no guarantee that features on the next category are predicted. The implications of these assumptions and some possible future directions are discussed in more detail in Chapter 5.

Consistent with what Whelpton et al. (2014) found for effects of prediction of upcoming structure, the constituents that have the processing step of “predict” may show a P600 effect when compared to constituents where there is only commitment. They additionally found a brief negative peak in the anterior region for the condition that predicted extra syntactic structure. In the Whelpton et al. (2014) study, once the predicted noun appeared and was congruent in featural information to the determiner, (analogous to our Nonagree-Agree condition), there was a diminished N400 response compared to the neutral case where the first noun had also agreed in gender with the determiner (analogous to our Agree-Agree condition). They explained this diminished effect as being due to an ease of integration, as the structure for the constituent had already been built on the previous noun.

In a study with a similar paradigm that presented compound phrases in isolation rather than in sentential contexts, Jalbert et al. (2016) failed to find an effect for prediction on an N1, but did find effects related to revision once the parser reached N2. There was a left negativity between 480 and 550ms at the point where it is clear that the structure must be reanalyzed. These findings, however, occurred when they used non-compound fillers. When these fillers were removed, such that the only experimental stimuli were compounds and thus a compound was highly predictable, they found a posterior negativity for the effect of revision in the 325 to 470ms time window. And so given the similarity of this experiment’s paradigm with that of Jalbert et al. (2016), we can expect revision effects to show negativities in either the left hemisphere or the posterior regions. Either option is potentially viable because

for the present experiment, although the fillers did all contain compounds, they were all in sentential contexts, which may have made the pattern of compounds where a participant would know what word category to expect at each lexical item less predictable.

A much more canonical finding associated with revision costs, though, would be a positivity in the Posterior region, peaking around 600ms, known as a P600. Many early studies of garden path sentences have noted that structural revision is associated with a clear P600 (Osterhout et al., 1994; Friederici, 1995; Hagoort & Brown, 2000; Kaan & Swaab, 2003, among others). Thus we also have a strong expectation to find a P600 associated with the costs of revision to a syntactic structure.

As another important note, we assume here that commitment may not have a measurable cost, at least not with regards to a comparison between committing and revising. Certainly, the effect of integration of a new lexical item is important in processing, but for this part of the analysis we assume that integration is a sort of “default” and that, at the very least, every constituent encountered by the parser must be processed, even if not immediately integrated into the structure built. Integration will certainly be a main feature of the secondary analysis of this study that directly assess integration as it relates to working memory, and this issue will be discussed in Chapter 4.

Going beyond just these predictions for the study’s results, the triple noun paradigm allows us to possibly expand in some additional directions. The rest of this section discusses reasonable directions that findings may go in, based on other, similar studies.

A study by Koester et al. (2009) found continued semantic implausibility resulted in progressively greater amplitude of response on each constituent where a mismatch existed. If his findings for semantic manipulations hold for syntactic ones as well, then we expect the second constituent of a second occurrence of a costly processing step, such as “predict” to show a greater amplitude of response on its second occurrence. This would be the case in comparing N2s of the Agree-Nonagree and Nonagree-Nonagree conditions. If we do find an increased amplitude, it could be explained by increased processing load at successive

constituents.

As the need to revise a previous interpretation continues, we may also expect to see increasing amplitude of this revision response in the Agree-Agree condition compared to Nonagree-Agree condition, as measured on the third noun. This finding would be expected because the parser must, on N3 in the Agree-Agree condition, revise a structure that is already more costly, having already been revised. This prediction, though, is based on the assumption that the parser is using more than just the previous word encountered to determine whether revision is necessary. While this appears to be a reasonable assumption to make, it is also the case that the parser may, at that point, be treating [N1 + N2] already as a single unit due to the semantic plausibility of them forming a sub-compound together. The cost of revising a larger, more specific, or more costly structure may also place a higher burden on other, non-linguistic cognitive faculties, such as attention or working memory.

In fact, it may be the case that the effects of revision are stronger on N3, since the revision of what the parser interpreted as a compound to a sub-compound structure would require that a larger unit be reanalyzed. This possibility would still be expected based on an increased processing load and possible increases to the working memory load as the larger unit must be moved to a modifier role. It would also be reasonable to consider this as possibly within the realm of coercion, as the modifier meaning may be a coerced form of the head noun. However, there is a very broad literature on coercion effects, and they can differ quite a bit based on what it being coerced, so an explicit tie-in of coercing something from a head to a modifier with the existing literature is far beyond the scope of the present work.

As a more exploratory kind of analysis, this paradigm additionally allows us to investigate whether there is a difference between reanalyzing a structure that was committed-to based on presentation of a syntactically licit nominal constituent versus one that was only predicted due to a mismatch of grammatical features. By comparing the processing effects of N2 in the Agree-Nonagree (committed-to structure needs revision) and Nonagree-Nonagree (predicted structure needs revision) conditions, we may be able to investigate whether there

is a measurable difference in the “strength” of a structure that is build from lexical items available compared to one built from knowledge of grammaticality. The lack of an effect in this comparison, while not conclusive, would be consistent with a strong form of prediction theory, where a prediction can literally build the structure needed for the next element(s).

2.3 Methods

2.3.1 Participants

Forty-seven right-handed, neurotypical, native Icelandic speakers participated in this experiment. They were compensated for their time with a gift certificate for the campus bookstore/cafeteria worth approximately \$10. All participants had normal or corrected-to-normal vision. Male and female participants were equally represented, with about 45% of the participants being female. Participant ages ranged from 19 to 37, with the mean age at 25.5.

2.3.2 Materials

Experimental items were created in 35 quads. The sentence frame was completely consistent within a given quad, and only the nouns making up the first two constituents of the compound were manipulated. The third constituent always remained the same, and it was always syntactically licit with the determiner.

Given the four conditions within the quad described in (1), an example of what these four conditions looks like with actual Icelandic compounds is shown in (2), and a partial list of other experimental items can be found in Appenix A:

- (2) Konan útbjó fjolda nýrra ... í eldhúsinu
 The-woman prepared a-number-of new ... in the-kitchen

a. **Agree-Agree**

hafra grjóna seyða ...
 oat.m.pl.gen grain.n.pl.gen broth.n.pl.gen

b. **Agree-Nonagree**

bygg grjóna seyða ...
barley.m.root grain.n.pl.gen broth.n.pl.gen

c. **Nonagree-Agree**

hafra mjöls seyða ...
oat.m.pl.gen flour.n.sg.gen broth.n.pl.gen

d. **Nongree-Nonagree**

bygg mjöls seyða ...
barley.m.root flour.n.sg.gen broth.n.pl.gen

Note that there are changes to some grammatical features even within the Agree-Agree condition. For example, both *hafra*, a masculine noun, and *grjóna*, a neuter noun, are compatible with the form of the adjective given (*nýrra*). This is due to the high degree of syncretism in the Icelandic inflectional paradigm. Therefore, both *hafra* and *grjóna* can be said to agree with *nýrra*, but *mjöls* cannot because the *nýrra* is incompatible with a noun inflected for neuter singular genitive. We assume here that the cost of structural revision effects on N2 and N3 of the compounds will overshadow any potential effect of featural revision (which, to be fair, may also be structural, though on a smaller scale). Some implications and possibilities for future studies to investigate this assumption are discussed in more detail in Chapter 5.

As an added control, all items used were structurally identical following presentation of N3. That is, each item had the same left-branching structure where the first and second constituents form a modifier and minor-head relationship, as shown in (1), assuming a structure of multi-noun compounds as discussed in (Berg, 2011).

With regard to the different possibilities for gender, case, and number features on each noun, Table 3 shows the occurrence of each feature on each constituent.

Looking at the differences in the gender of the items, there are roughly equal numbers of masculine, feminine, and neuter used throughout the paradigm. A post-hoc Chi-Square test of independence did show that gender varied with noun position, $\chi^2(4) = 15.47$, $p < 0.01$. The values for case features also varied with noun position, $\chi^2(6) = 76.72$, $p < 0.01$. As all

Figure 1: Diagram of assumed structure of the compound in sentences in (2a)

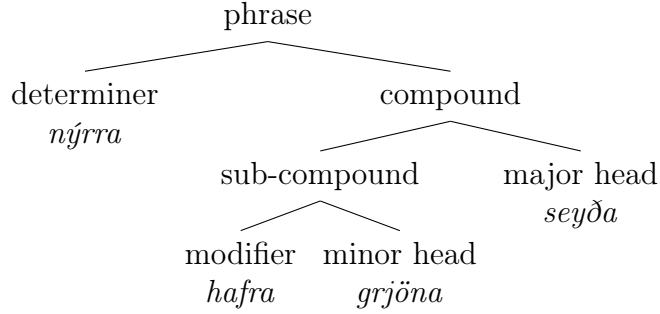


Table 3: Counts for the inflectional features on nouns used for experimental items

		N1	N2	N3	Total
Gender	Masc.	54	36	32	122
	Fem.	38	66	52	156
	Neut.	48	38	56	142
Case	Acc.	48	46	64	158
	Nom.	8	2	12	22
	Gen.	56	92	64	212
	Root	28	0	0	28
Number	Sing.	60	100	80	240
	Pl.	52	40	60	152
	Root	28	0	0	28

of the stimuli were in non-subject positions, it makes sense that accusative and genitive case occur more frequently than the nominative. The same is true for the number feature, which also varied with noun position $\chi^2(4) = 70$, $p < 0.01$. We assume that neither singular nor plural is particularly “marked” in any way that would be relevant to this study.

The χ^2 values for case and number were both extremely high, but this is likely due in large part to a small number of root lemma items occurring only ever on N1. Additionally, it is not clear that these significant test values tell us anything meaningful about a difference in the stimuli items. While it is certainly true that with root lemmas, they can never appear as heads, and will thus never appear at the final noun position, all other combinations could conceivably occur in any of the three positions.

The fillers also contained compounds. The items used as fillers in this experiment were

part of a separate experimental paradigm run concurrently with this study of triple nouns. The compounds within the filler items all consisted of noun-noun compounds, all of which were grammatically licit, though one quarter of the items were semantically implausible. For reference, a partial set of filler items is provided in Appendix B.

2.3.3 Procedure

Participants began the experiment with a set of eight practice sentences to get used to the task to be used in the study. All example sentences were presented in random order via rapid serial visual presentation (RSVP) with a 350ms presentation time and a 350ms inter-stimulus interval. Each sentence started with a fixation cross that will be displayed for 1000ms at the center of the screen. Each sentence was followed by a yes/no comprehension question about the sentence they had just read that the participant responded to via keypress before moving on to the next sentence.

We set a threshold of 90% accuracy on the comprehension questions to assure that the participants were appropriately attending to the task. Based on this threshold, we did not have to exclude any participants from analysis.

For the presentation of the nominal constituents of the compounds, we presented each as a separate word, as was done in previous similar studies, such as Whelpton et al. (2014); Jalbert et al. (2016); and Parrish, Kelley, and Beretta (2016). Though this is not the standard means of orthographically representing compounds in Icelandic, such a step was absolutely necessary for the purposes of this experiment, as we had to take separate measurements at each noun in the compound. Participants noted verbally after the experiment that they did not find this mode of presentation especially difficult.

The entire experiment took approximately 45 minutes once participants started the practice questions, including the short breaks that were offered to participants approximately every five minutes.

2.3.4 Data analysis and recording

The electroencephalogram (EEG) used a 32 Ag/AgCl electrode elastic cap (GND WaveGuard 32 Electrode cap; Advanced Neuro Technology BV., Enschede, The Netherlands). The amplifier was a full-band EEG DC Amplifier (Advanced Neuro Technology), with a 256 Hz sampling rate. Offline post-processing used a 0.01 to 30 Hz bandpass filter. Impedance at each electrode were kept below 5 k Ω in order to assure a clean signal could be read. The signal consisted of a continuous recording with a whole-head average reference applied. Stimulus presentation was done via PsychoPy software (Pieirce, 2007), post-processing (including rejection of artifacts, baselining and averaging) was done via MATLAB (MATLAB, 2010) with the EEGLAB plugin (Delome & Makeig, 2004), and all statistical analyses and visualizations were done with R (R Core Team, 2013).

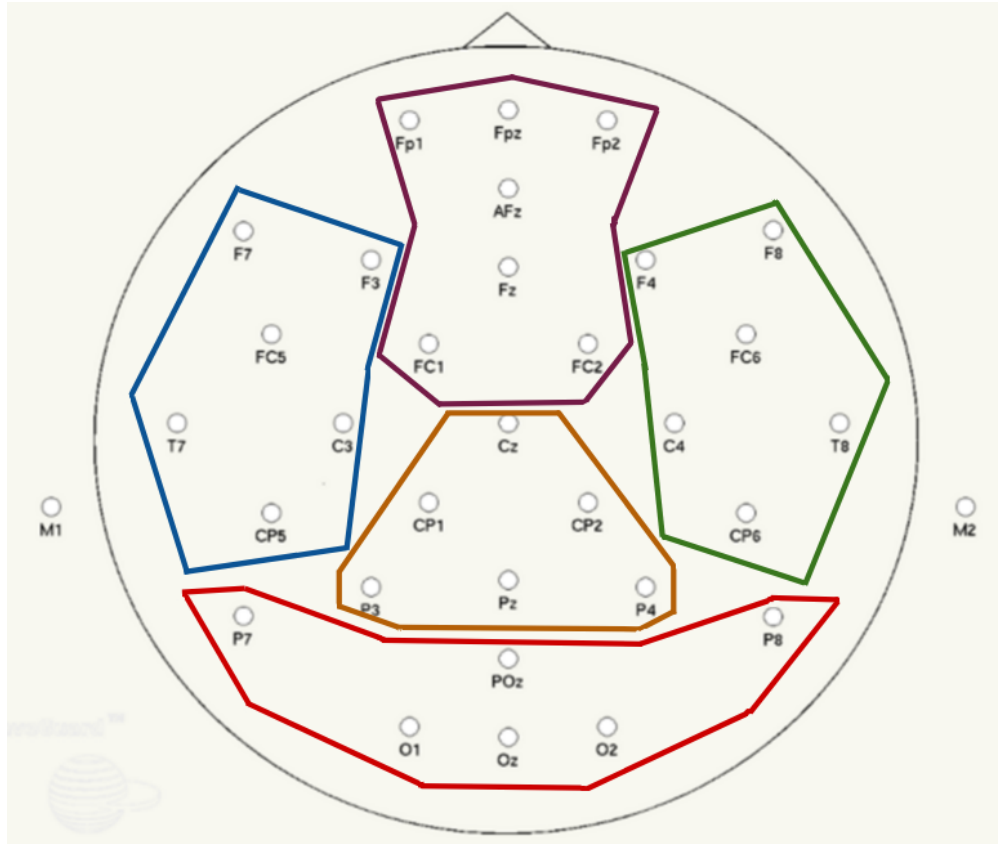
The artifact rejection rate for this experiment was set at $\pm 50\mu V$ with a peak-to-peak moving window. Six participants were excluded from analysis because their overall artifact rejection rates were above 15%. The remaining 41 participants had an average rejection rate of 3.2% across all trials, with no participant having an individual condition with greater than a 20% rejection rate.

The electrodes were grouped into five different regions of interest (ROIs). The grouping used is identical to that of Parrish et al. (2015), another study of commitment and revision within nominal compounds. The five regions defined were as follows:

- Anterior (Fp1, Fpz, Fp2, Fz, FC1, FC2);
- Frontocentral-Right (F4, F8, FC6, C4, CP6, T8);
- Frontocentral-Left (F3, F7, FC5, C3, CP5, T7);
- Centroparietal (Cz, CP1, CP2, P3, Pz, P4); and
- Posterior (P7, POz, P8, O1, Oz, O2).

This ROI configuration for all of these electrodes shown in Figure 2.

Figure 2: Electrode groupings for the five regions of interest



2.4 Pretests

In order to make sure that all the items presented in this study were sensible, transparent compounds, we conducted a pre-test. This pre-test was an acceptability judgment task that used a 1-5 likert rating scale. A total of 100 triple noun compound items created by the experimenters and intended to be good items and 50 items intended to be nonsensical were presented to participants without any sentence frames. The items were split into two version so that each participant only saw half the experimental items from any given quad in order to mitigate any possible effects of participants noticing a pattern in the data. A sample of the “bad” items is provided in Appendix C.

Quads of the experimental items were selected such that no individual item within a given quad received an average acceptability judgment score below a 3.0. The average scores in each condition for the 35 selected experimental quads are shown below in Table 4:

Table 4: Mean scores given on the acceptability judgment task for selected stimuli

Condition	Mean score
Agree-Agree	4.43
Nonagree-Agree	4.46
Agree-Nonagree	4.51
Nonagree-Nonagree	4.35

There was no significant effect of either N1 agreement or N2 agreement. Additionally, there was no significant interaction between N1 and N2 agreement in these judgment scores. We thus conclude that the triple noun compounds used in the experiment, while novel, were semantically plausible to an equal (and fairly high) degree across all four conditions.

CHAPTER 3

RESULTS AND DISCUSSION OF THE TRIPLE NOUN STUDY AS THEY RELATE TO COMMITMENT AND REVISION

For this particular analysis, we are comparing different conditions at different noun positions. None of the comparisons are done across noun positions in this chapter, as such comparisons are confounded by additional processing costs that are not taken into consideration with an analysis that directly compares the four conditions in the experimental paradigm. The focus of this chapter is assessing (i) to what degree can we replicate previous findings; (ii) to what degree can we expand on the findings of others; and (iii) to what degree can we come to novel conclusions about questions that have not been addressed in similar studies. The following sections address each of these aims.

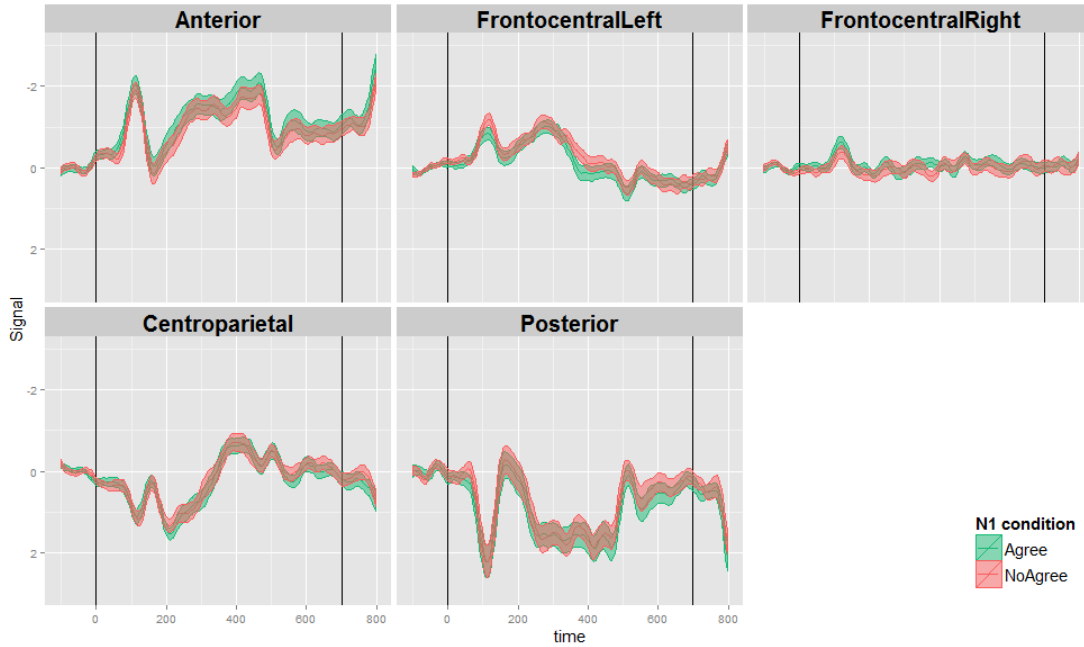
3.1 Replication of previous findings

The first comparison is a direct replication of a comparison made in Whelpton et al. (2014) between their neutral condition and the condition where extra structure was predicted. Figure 3 is a comparison of the two conditions where N1 agrees with the determiner with the two conditions where N2 does not agree with the determiner. The results show a brief negativity for the Nonagree condition at N1 in the Frontocentral-Left region, very similar to the brief anterior negativity measured by Whelpton et al. (2014). Here we see a negativity of Nonagree in the 350-450ms window, $F(1,40) = 4.6$, $MSE = 0.39$, $p = 0.038$.

With regards to the possibility of expecting a P600 effect, there was no significant effect of condition anywhere past 500ms in either the Centroparietal or Posterior regions, all $ps > 0.2$.

Note that our distribution of ROIs is not identical to that used by the other two main studies that we are using for comparison. Because of this, some effects may be obscured. This is one possible explanation for the lack of a P600 effect in Figure 3, even though there

Figure 3: Effect of agreement of N1 measured on N1



was a very clear effect in the other study. Additionally, Whelpton et al. (2014) were looking at coercion as well as syntactic predictions and revisions, thus their N1s also differed along the lines of different types of noun, which may have interacted with the syntactic effects in their study, but would not in the present study.

When looking at revision effects, we can compare the effects at N2 between the Agree-Disagree condition and the Nonagree-Disagree condition, as shown in Figure 4.

We do not see the same effects that we would expect from Jalbert et al. (2016), nor did we find the diminished N400 response that would be expected from Whelpton et al. (2014). Based on visual analysis, there appears to be an effect in the Anterior region after 300ms. Comparing Anterior to the two ROIs next to it, Frontocentral-Left and Frontocentral-Right, there is an interaction of ROI x Condition, $F(2,80) = 7.22$, $MSE = 1.21$, $p = 0.001$.

Looking within the Anterior region, we see that is significant for the long 300-700ms time window, $F(1,40) = 6.607$, $MSE = 2.12$, $p = 0.018$. This same effect was marginally significant in the Frontocentral-Right region, $F(1,40) = 3.24$, $MSE = 0.74$, $p = 0.08$, and also for the Frontocentral-Left region, $F(1,40) = 3.54$, $MSE = 0.63$, $p = 0.067$.

Figure 4: Effect of Agree-Agree versus Nonagree-Agree measured on N2



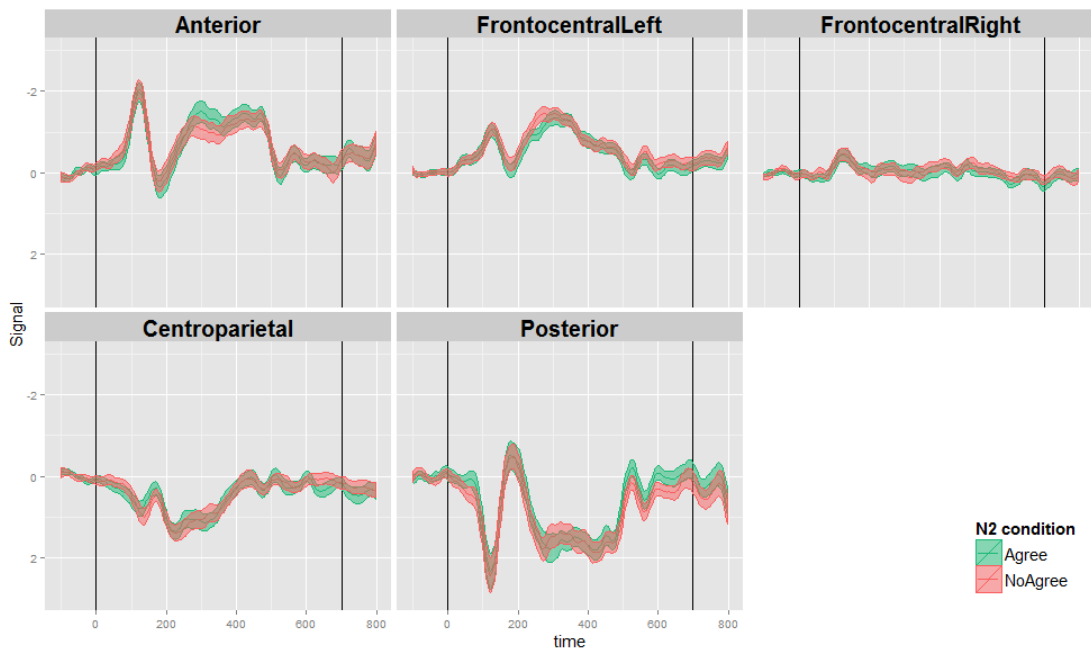
As it is only in the condition where N1 agreed with the determiner that revision is necessary, results from previous studies suggest that this is the condition that will incur the higher processing cost. And this is precisely what we have found. It is unclear, however, why this effect is very different from the one noted in Jalbert et al. (2016), as it differs in ROI and latency.

Overall, the effects of revision to a structure, while there is a clear effect, have not clearly lined up with what was found in the past studies most similar to the present study. Another possibility is that this effect is a slightly more middle-distributed left anterior negativity (LAN). Many past studies have found a LAN for syntactic processing costs, including costs related to morphosyntactic processing and decomposition or phrase structure violations (Friederici, 1995; Coulson, King, & Kutas, 1998; Neville, Nicol, Barss, Forster, & Garrett, 1991; Münte & Meinze, 1993, among others). However, most of the studies that report LAN effect are related to clear morphosyntactic violations such as a gender mismatch. The relationship between frontal negativities in the 300-500ms range and costs associated with revision is not established. Furthermore, LANs are consistently found to be in the left hemi-

sphere. Therefore it is very unlikely that this effect is directly related to what has been measured previously as a LAN.

Continuing on to the other area where we expect to see a cost of revision, though, we can look at the costs associated with N3 based on differences on N2. This contrast is shown in Figure 5.

Figure 5: Effect of N2 agreement measured on N3



Here, we see a brief anterior negativity for the condition where the previous noun agreed in gender with the determiner, which mirrors the effect noted on N2, though only for the 275-350ms window, $F(1,40) = 4.44$, $MSE = 0.59$, $p = 0.044$. (The interaction with Frontocentral-Left and Frontocentral-Right here is only marginally significant, $F(2,80) = 2.94$, $MSE = 0.34$, $p = 0.059$). This effect is clearly much smaller than the effect measured on N2 and shown in Figure 4, though the effect is significant in the same ROI and the negativity begins around the same time window.

Overall, these results can be considered a partial replication of previous studies of commitment and revision. Certainly, with regard to prediction, we replicated the brief anterior negativity that was measured by (Whelpton et al., 2014). The effect of revision, though

not what has been found by the two most similar studies to this one, showed a moderately consistent effect in the two areas where we expected to find costs of revision. Thus the results of the findings of revision for this study do not match up precisely to well-documented components, though they are internally consistent.

3.2 Extension of previous findings

An experiment simply for the sake of replication, while valuable, would not add much that is new. Thus this study also aims to test an extension of the findings reported by Koester et al. (2009) in order to see if the effects of continued semantic implausibility can extend to cases of continued prediction or continued revision steps. It may be the case, then, that predicting further structure on N1 and then again on N2 is more costly than simply having made one prediction. If this were the case, we would expect to see a greater amplitude for the negativity measured for the case of prediction in the Frontocentral-Left area (as this is what the measured effect was for prediction in this study in the previous section).

Figure 6: Comparison of Agree-Nonagree and Nonagree-Nonagree conditions at N2

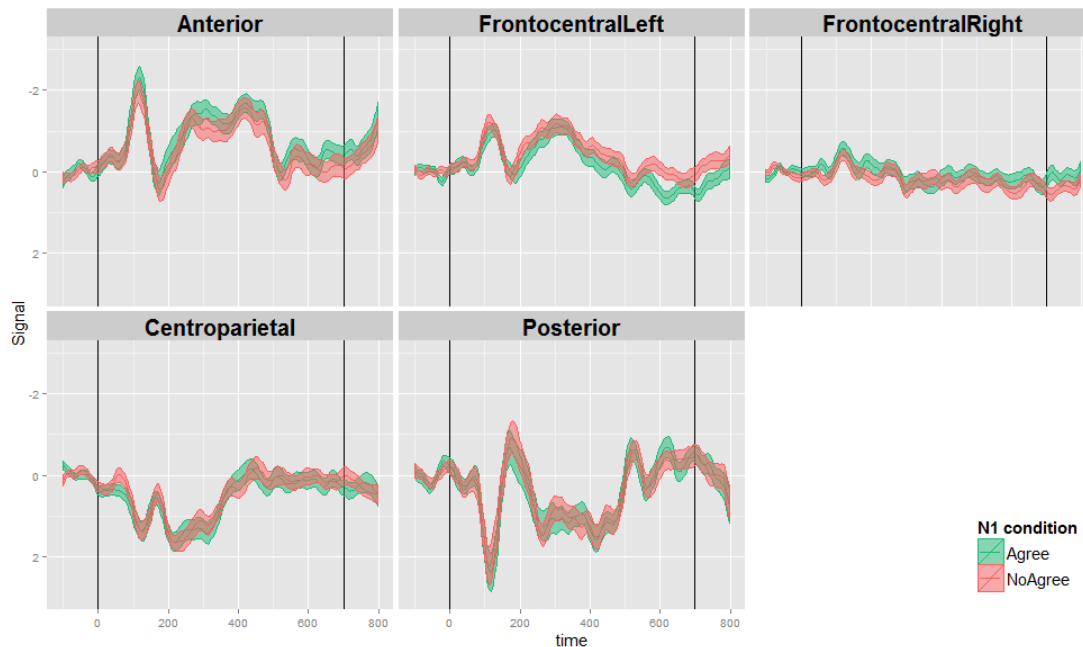


Figure 6 shows results that would be consistent with this. This figure shows the effect

on the N2s that did not agree with the determiner. When N1 also mismatched with the determiner, that is, when there is an effect consistent with prediction twice in a row, we again see a negativity in the Frontocentral-Left area, like in Figure 3. Here it is marginally significant for an extended window, from 400-700ms, $F(1,40) = 3.25$, $MSE = 0.63$, $p = 0.079$, and it becomes much more significant from 550ms on, $F(1,40) = 4.52$, $MSE = 0.66$, $p = 0.039$.

Figure 6 can also be interpreted in terms of addressing the question of whether there's a meaningful distinction between building structure based on constituent integration and building structure based on prediction from cues on a lexical item. In the case of both the Agree-Nonagree and the Nonagree-Nonagree conditions, a previously assumed structure must be rejected. However, in the Agree-Nonagree case, that structure was actually built, whereas in the Nonagree-Nonagree case it was predicted based on the mismatch in gender from N1. The only effect seen in Figure 6, though, is consistent with prediction. The effect of revision, seen in Figures 4 and 5, appeared in the Anterior region, but there is no difference in the revision effects measured here based on whether it was a built or predicted structure that was revised.

While not definitive, this result is consistent with the idea that when structure is predicted, it is *actually built* in the same way that it is when the structure is built from incoming lexical items that must be integrated. Indeed, saying that predictions were made, but that the structure was never actually built, would be a rather vacuous and weak claim to make. The stronger claim, and the one that is consistent with the effect measured here, is that there is no measurable difference in what is built when the parser has cues that come from a lexical item giving cues that it needs to be integrated into the structure versus a lexical item that gives cues structure needs to be built for an upcoming item(s).

For the effect of extended revision, we here look in the Anterior region, as that is where the effect was measured in Figures 4 and 5. The results of the extended revision are shown below in Figure 7, which shows the effect on N3 for both the cases where N2 agreed, and

compares the differences in agreement from N1. This means that the Agree condition here is where revision has now happened twice, and the Nonagree is where it is only happening once.

Figure 7: Comparison of Agree-Agree and Nonagree-Agree condition on N3



In this case, there are no significant effects for any window in the Anterior region, all $ps > 0.2$. This result indicates that we have not measured any compounding effects of continued revision. A null result is never conclusive on its own, but compared to the fact that there was an effect consistent with extended prediction, it is worth considering why there may be a cost to continuing predictions but not revisions. It may be the case that what matters is the size of the element being revised, rather than the number of times in a row that revision must be made. Furthermore, a continued prediction would cause the parser to need to hold more items in working memory as it goes on to the next item, but to revise a second time, it may just need to rebuild one item.

While it seems intuitive that a revision affecting only [Det + N1] to revise to [Det [N1 + N2]] would be more costly than revising something with a third noun, the actual number of changes that need to be made in terms of structural nodes that need to be changed is

identical. Because all items used had a final compound structure where N1 and N2 formed a sub-compound, the third noun only needs to revise what node the determiner attaches at. Thus in both the Agree-Agree and Nonagree-Agree cases (and the cases that mismatch on N2, for that matter), the exact same part of the structure is being revised each time. With this in mind, the lack of an effect at N3 in Figure 7 makes sense: the cost the the parser is identical because the structural revisions at that point are identical. The cost of having revised a structure does not carry on to the next lexical item.

Further studies may want to push this conclusion in comparing triple noun compounds of this left-branching structure to those with a right-branching structure of [N1 [N2 + N3]], as in those cases there would be an additional change that has to happen once N3 is presented (compared to the change at N3 in the present study). Thus by looking at the differences in processing on the third noun in compounds with left- versus right-branching internal structures, the effect of greater revision costs to greater changes to the structure could potentially be probed.

CHAPTER 4

RESULTS AND DISCUSSION OF THE TRIPLE NOUN STUDY AS THEY RELATE TO WORKING MEMORY

In this chapter, we examine the results from the experiment outlined in Chapter 2, but we examine the data in a novel way. Rather than comparing the effects of different conditions directly against each other, this section will analyze the differences that occur as the parser moves from N1 to N2 to N3. The reason for this comparison is that, given many models of working memory to be described here, we would expect to see working memory costs associated with adding additional nouns. What is not known, though, is whether these costs can be affected by syntactic cues related to grammatical features. What follows is a brief overview of the relevant issues in working memory research and how they are related to the questions of the present study.

One of the most influential models of working memory to be developed has been outlined and championed by Alan Baddeley. His model has made significant advances by separating out some functions of working memory between the way people process visual and phonological input and by describing the processes of the central executive to modulate them. Later models included an interface between these processes in the form of the episodic buffer and also interfaces with language, memory, and recognition (Baddeley, 2010).

It has been known for quite some time that working memory in the way Baddeley outlined it must interact with language and linguistic structure in some way. Baddeley (2012) describes one of the first studies to show this back in 1987, where memory span in terms of the number of words remembered increased when those words were given in a sentence as opposed to being unrelated words strung together. He interpreted this as being more than just an additive effect of systems related to phonology and semantics, rather he reported it to reflect an interaction between them. The main point from this, however, is that language, when structured, interacts with the working memory in such a way that it is clear

the structure is maintained in working memory.

Some alternative theories to Baddeley’s model of working memory have tried to recast the entire process as activated long term memory with the need for some attentional resources (Cowan, 2005) or they have added long term working memory as an additional component (Ericsson & Kintsch, 1995). With regards to sentence comprehension and language processing, this does not seem to make any substantially different predictions. All these theories of working memory agree that long term memory plays an important role, and we can assume that linguistic knowledge and the lexicon would be stored in long term memory, though it would become active or be manipulated in working memory during language processing.

Given the importance of working memory in the psychology literature, it is no surprise that one particularly influential model of computing costs in sentence processing deals explicitly with effects coming from working memory. Gibson’s (1998) Syntactic Prediction Locality Theory (SPLT) aimed to explain the range of processing cost effects that had been noted across multiple studies, paying close attention to the effects seen in subject versus object dislocated relative clause structures. His theory was particularly novel in not only the way it separated two different computational costs, but also the importance it placed on discourse referents. Note that for Gibson, tensed verbs counts as discourse referents because they introduce new discourse events, though this does not directly affect anything being measured in this study.

The two crucial components of Gibson’s SPLT model are the integration cost component and the memory cost component. Throughout the discussion of the costs associate with these components, Gibson stressed the theory of locality, highlighting its importance because “syntactic predictions held in memory over longer distances are more expensive, and longer distance head-dependent integrations are more expensive.”

Gibson quantifies his memory cost component in terms of memory units (MUs). These units are calculated by determining where a parser can make a prediction about upcoming structure, and then carrying that prediction on each constituent until it is realized. When

it is finally realized, the prediction is discharged, and there is no additional memory cost to bearing out the prediction. Crucial to his theory, MUs become more costly the more discourse referents are processed between the lexeme where the prediction was made and the point where the prediction was discharged.

The parser is able to perform a needed integration when the MUs occupied with holding the lexical items in memory does not exceed the memory capacity. The integration cost is represented in terms of energy units (EUs) and is calculated based on whether a new discourse referent is created and the distance between the two items being integrated, measured in the number of intervening discourse referents. The important part for this study is that each “commitment” made, according to the model of commitment and revision supported by this study, is, for the time it takes until the next lexical item arrives, a full discourse referent. Thus a later integration of an N2 on which N1 was considered a discourse referent should be more costly than one in which N1 was not (i.e., when it mismatched in gender, case, or number and thus couldn’t integrate to form a new referent). The steps in calculating the cost in terms of $I(n)$ for each nominal constituent in each condition is detailed in the subsections that follow.

Note that Gibson’s model is not too different from Phillips (1996) suggestions of counting nodes as a measure of syntactic complexity, and thus the costs associated with sentence processing. In this analysis, the number of integrations to make on a specific word and the number of discourse referents introduced in between that integration are both related to the distance in terms of total number of new nodes. Therefore it is likely that much of Phillips’s “Parser is the Grammar” model would also be compatible with this processing analysis.

4.1 The sustained anterior negativity (SAN) and working memory

Anterior negativities in the sentence processing literature usually refer to left-lateralized early effects (the ELAN) or ones that occur between 300-500ms (the LAN, discussed briefly in Chapter 3). While there is some debate over the status of whether the LAN should

be considered an independent component (Molinaro, Barber, & Carreiras, 2011; Molinaro, Barber, Caffarra, & Carreiras, 2014) or an artifact of an N400 and a P600 (Tanner, 2006), it is for the most part thought to reference morpho-syntactic processing and error detection (Sprouse & Lau, 2013).

A relatively less commonly studied variant of components showing an anterior negativity is known as the sustained anterior negativity (SAN). This component differs in that it continues far beyond 500ms post-stimulus, sometimes several words later. Crucially for this study, the SAN has been reported as a separate component from the LAN not only in its morphology, but also in what it indexes.

A study by Fiebach, Schlesewsky, Lohmann, von Cramon, and Friederici (2005) looked at German sentences with embedded object wh-questions. Varying the length of the dependency, they found differences in the ERP signal not only on the moved wh-word, but a sustained effect that lasted until the gap was arrived at by the parser. This sustained effect was noted as a negativity for the wh-dependency condition. Such a difference was predicted by a working-memory view of language processing, as the wh-word must be held in working memory until the gap from which it was moved can be processed. Thus they concluded that the SAN they measured in their study aligns with increased working memory load.

Another study of wh-dependencies conducted by Phillips, Kazanina, and Abada (2005) found a SAN at the wh-word, followed by a P600 at place where the dependency was completed. Like Fiebach et al. (2005), they also varied the length of the dependency; however, they did not find any differences in the amplitude of the response and concluded that both the SAN and the P600 are used in conjunction with syntactic dependencies, but they are not sensitive to length. The effect of a SAN in this study also supports its role as indexing working memory processing costs, as it was measured at the site where a dependency had to be maintained over the course of several words.

This study uses the findings related to SAN effects in order to investigate to what degree the predictions made by Gibson's (1998) are borne out.

4.2 The importance of working memory’s interaction with models of commitment and revision

Working memory and language have been shown convincingly to interact, often in a way where it is reported that language structure facilitates memory (Baddeley, 2012). The primary question to ask here, though, is whether working memory effects are modulated by syntactic information from gender agreement. From what was shown in chapter 3, we can conclude that in cases where the syntax is providing information that a given constituent *cannot* possibly be the phrase’s head (i.e., there was a syntactic mismatch), we see evidence to support the view that this mismatched constituent is not integrated. Assuming that working memory is essential to language processing, the effect of holding a constituent versus integrating two (or more) together should have different effects when we look at components that reflect the processes of working memory.

If we can measure working memory effects through an ERP component like the SAN, then we can measure the way that the syntactic differences in this study show up when we look at the patterns occurring in different conditions. Here, Gibson’s SPLT model is especially useful as it has an index to count integrations (or, as they’ve been labelled earlier in this paper, “commitments”). Also important is that the idea of “revision” as discussed in Chapter 3 has a clear analogue in the SPLT: new discourse referents are introduced in the course of integrations, the need to revise would only come up if an integration was made, and a greater number of discourse references in intermediary steps of two lexical items integrating results in higher processing costs, therefore we expect processing costs to be attenuated for constituents where integration is or was blocked, i.e., constituents on which there was a featural mismatch.

4.3 Results across constituent places

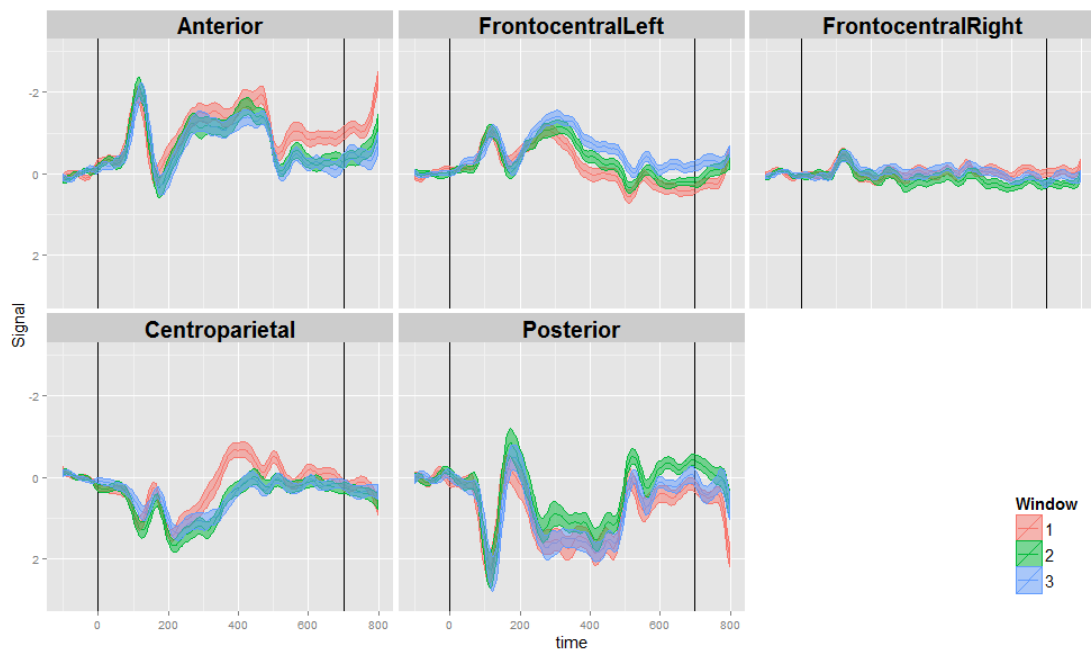
This analysis is fundamentally different from the one discussed in Chapter 3. Rather than comparing, for example, N1 from one condition to N1 from another condition, this part

of the analysis compares N1 to N2 to N3 all within the same condition.

It is important to note here that any model of working memory will predict there to be a processing cost as more and more constituents are added to a compound, as it would increase processing load in some way. However, such a model by itself would not be able to predict anything about differences between conditions. That is, working memory models will simply assume that each consecutive constituent adds to the processing load that working memory must hold or manipulate. However, if the syntax is giving information to the parser to *not* integrate at certain points, we expect a difference in processing costs to the working memory.

Looking across constituent places in the ROIs defined previously, Figure 8 shows that there are, in fact, increasing negativities with each consecutive nominal constituent in the Frontocentral-Left region. Note that this analysis does not yet differ based on condition. All conditions are collapsed together in this graph.

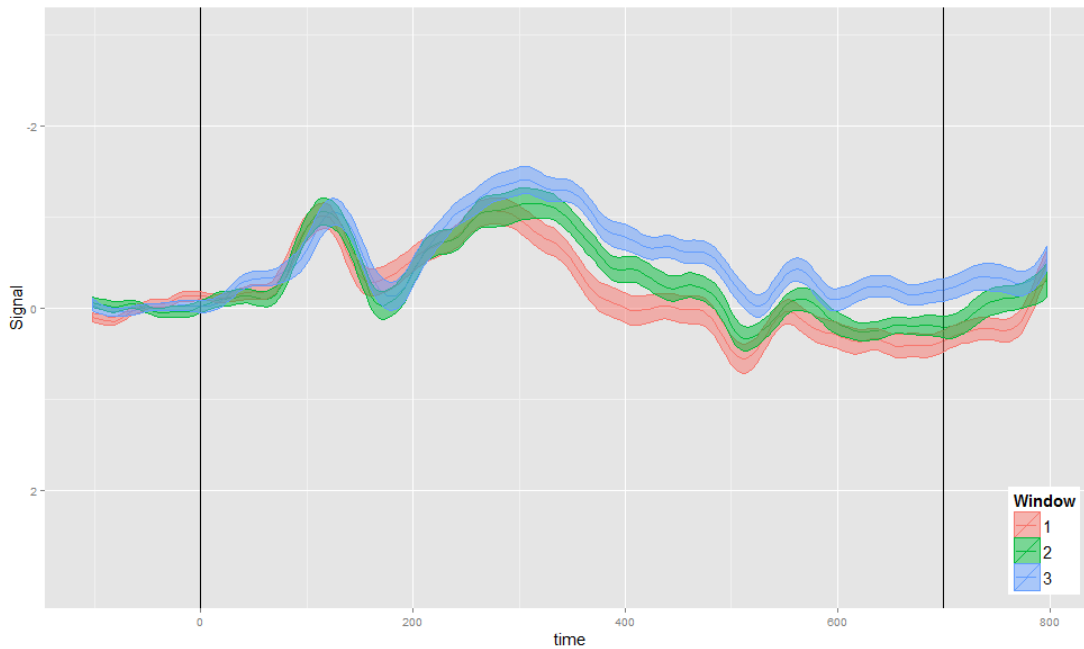
Figure 8: Differences across noun position with all conditions averaged together



The effect occurs quite clearly in the Frontocentral-Left region, highlighted below in Figure 9, with clear distinctions between the first, second, and third noun windows. While

this effect was reported by Phillips et al. (2005) to occur along the anterior midline, it is worth noting that this study differed significantly in the ROIs used. It may be that we did not measure the SAN in the Anterior region because our Anterior region included the frontal poles, which added some noise to the data. So we report this as a slightly left-localizing SAN.

Figure 9: Differences across noun position in the Frontocentral-Left region



Thus while the general trend will unsurprisingly be that we see an increasing sustained negativity with each additional constituent, by looking at how different conditions pattern, we are able to compare the way that syntax interfaces with working memory at each part. Thus the crucial comparisons for this part of the study cannot be made directly between conditions, but rather we are looking at the differences in the ways that the waveforms pattern in each condition. The long-window waveforms for each of these conditions (i.e., a window of 2400ms created without rebaselining at each new constituent) is provided for reference in Appendix E.

4.3.1 The agree-agree condition

Extending the sentence processing theory outlined by Gibson (1998), the values assigned at each consecutive word in the Agree-Agree condition of this experiment are outlined in Table 5 and described below.

Table 5: Predictions from Gibson (1998) for the $I(n)$ cost in the Agree-Agree condition

Det	$N1_{Agree}$	$N2_{Agree}$	$N3_{Agree}$	
	$I(0)$	$I(1)$	$I(2)$	Integrate(referents) w/ det
	-	$+I(0)$	$+I(0)$	Integrate(referents) w/ Ns
Total	-	$I(0)$	$I(1)+I(0)$	$I(2)+I(0)$

First, upon presentation of the determiner, the parser predicts that a noun that can be the head of an NP will come. This prediction is represented by $M(0)$ from the $Det(eterminer)$, and it is assigned a number of 0 because the prediction is made at that point and thus no referents could have been introduced since the prediction was made. This cost is not included in the calculations of the table because it is assumed to be constant throughout the conditions. Furthermore, EUs, and thus the calculation of integration cost in $I(n)$ terms, include MUs in the calculation, thus it would be redundant to calculate MUs at each point.

Once the first noun arrives, there is an integration that takes place, as marked by $I(0)$. The integration is between the noun and the determiner, and no other referents have been introduced between the two elements being integrated, so the number in parentheses is 0 .

At the arrival of the second noun, it is not completely clear what Gibson’s model would say must happen with the MUs . The lack of clarity is because, according to all examples given by Gibson, once a prediction is borne out, the parser does not again need to access that prediction. It is not clear if the parser no longer is able to access that prediction, or if it would re-access it and then count the number of referents introduced between the new place where the prediction is borne out and the original place the prediction is made. There are two distinct possibilities here: (i) the parser re-activates the prediction and must count Det + N1 as a referent, leading to a cost of $M(1)$, or (ii) there is no additional cost here

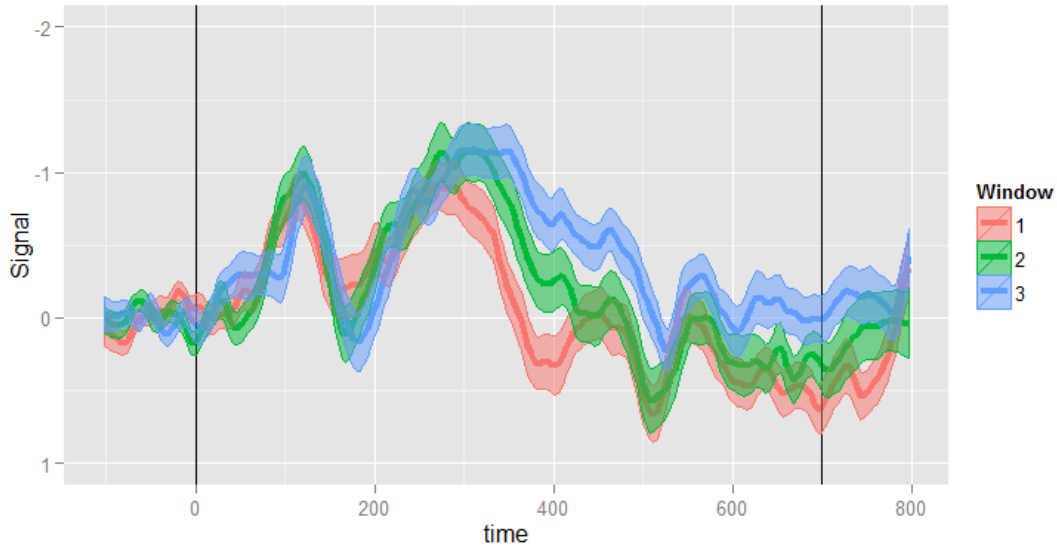
because N2 is still in line with the prediction made on the determiner, leading to a null cost represented by (*). However, in calculating the total cost at each constituent, integrations are actually much more clear here. There are two different integrations that must be made at this point. One integration is between N1 and N2 in order to form the compound. This step is represented by $I(0)$, as it is an integration with no full referents introduced between the two constituents (though the integration itself introduces a new referent to the discourse). The second integration is between that compound and the determiner. This step is represented by $I(1)$ because one referent has been introduced in between the two nodes that are integrating. The importance of treating Det + N1 will be discussed in more detail in Chapter 4.4, but the main point for now is that the parser must incur an additional cost for combining Det + [N1 + N2] because, from the time that the determiner was interpreted, there has been one full referent interpreted. It should also be noted that this interpretation of a referent was, upon presentation of N2, clearly erroneous. However, the parser is incremental, and thus not a magical time-machine that can undo a cost already introduced.

Finally, the parser encounters N3. At this point, it is again not clear what the *MUs* is supposed to read, and may be either $M(2)$, as two new referents have been introduced since the prediction of a noun was first made (N1 and [N1 + N2]), or it could be (*), as this is a location where the prediction of a noun is simply discharged. However, it is clear that there are two integrations that must be made at this point. N3 must integrate with the compound [N1 + N2], and this is represented by $I(0)$. Then there is an integration between the full compound and the determiner, between which two referents have been introduced in the course of incrementally processing the sentence (N1 and [N1 + N2]), and thus this integration is represented by $I(2)$.

Looking across the three nominal constituents, we would expect to see increasingly greater cost associated with each additional noun. Figure 10 shows the waveforms for these three nouns as measured in the Frontocentral-Left region.

We do, in fact, see increasing negativities for each noun, such that N2 is more negative

Figure 10: Waveforms of N1, N2, and N3 within the Agree-Agree condition



than N1 and N3 is more negative than N2. These negativities peak around 300ms and are sustained through the presentation of the following word. This pattern is consistent with what has been reported for the sustained anterior negativity (SAN), and is thus interpreted in line with component.

Furthermore, the differences in these waveforms are all highly significant. The three conditions differ across the entire relevant window of 300-700ms, $F(2,80) = 8.34$, $MSE = 0.5$, $p < 0.01$.

Comparing the effect of N1 to N2, N2 is more negative within the 300-500ms window, $F(1,40) = 4.84$, $MSE = 0.59$, $p = 0.033$. Between N2 and N3, the effect persists even longer such that N3 is more negative than N2 in the 300-700ms window, $F(1,40) = 6.6$, $MSE = 0.46$, $p = 0.014$.

An incremental time-window-by-time-window table of the areas of significance for this condition and the other three conditions is provided and discussed in Chapter 4.4.

4.3.2 The nonagree-agree condition

The Nonagree-Agree condition differs from the Agree-Agree condition only on the agreement features of N1 with the determiner. Given the findings discussed in Chapter 3, we expect that the parser knows that there is no integration that can take place at this first noun. Therefore, the costs associated with *EUs* via Gibson’s model will be slightly different, and these costs are outlined in below in Table 6.

Table 6: Predictions from Gibson (1998) for the $I(n)$ in the Nonagree-Agree condition

Det	N1 _{Nonagree}	N2 _{Agree}	N3 _{Agree}	
-		I(0)	I(1)	Integrate(referents) w/ det
-		+I(0)	+I(0)	Integrate(referents) w/ Ns
Total	-	-	2 I(0)	I(1)+I(0)

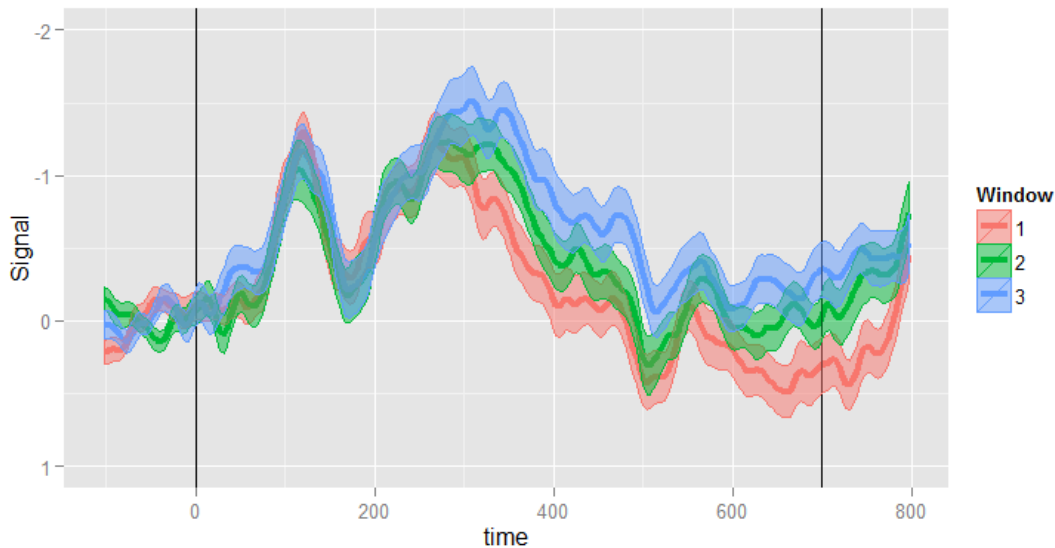
The cost associated with the determiner is identical to the Agree-Agree condition. At the first noun, though, the parser knows there is nothing to integrate, and that integration is impossible given the featural mismatch. Therefore the prediction of an upcoming noun made on the determiner is continued, though no new referents were introduced, which is represented by $M(0)$. There is no integration cost because there is nothing to integrate.

Once N2 arrives, the parser does have something to integrate. And here, there are actually two integrations needed here. N2 must integrate with N1, and the compound [N1 + N2] must integrate with the determiner. No new referents have been introduced in the middle of either of these integrations, so they are both represented by $I(0)$.

At N3, the cost is the same as was seen at N2 in the Agree-Agree condition. Two integrations are needed: (i) N3 must integrate with the compound, and (ii) the triple noun compound must integrate with the determiner. In the case of the integration with the determiner, the [N1 + N2] compound counts as a full referent that was interpreted, and thus the integration cost is $I(1)$ for that integration, while the integration of the triple noun compound with the determiner remains $I(0)$.

Just like in the Agree-Agree condition, this pattern predicts that the cost associated with each nominal constituent will increase as more are added. However, the differences will be less great than seen in the Agree-Agree condition because the cost differences between each constituent is less than it was in Agree-Agree. What was actually measured in the EEG waveform is shown below in Figure 11.

Figure 11: Waveforms of N1, N2, and N3 within the Nonagree-Agree condition



As can be seen, the pattern of results is very similar to what was seen in the Agree-Agree condition in Figure 10, but the increase in the negativity between each constituent following the 300ms point appears slightly smaller than was seen in Figure 10.

The pattern of statistical significance confirms that there is a significant difference between the three conditions in the 300-700ms window, $F(2,80) = 3.98$, $MSE = 0.65$, $p = 0.022$. Though N2 appears more negative than N1, the effect is only marginally significant in the 300-400ms window, $F(1,40) = 3.4$, $MSE = 0.69$, $p = 0.073$, and again in the 600-700ms window, $F(1,40) = 2.98$, $MSE = 1.03$, $p = 0.092$. And for the difference between N2 and N3, N3 is marginally more significant than N2 in the 300-500ms window, $F(1,40) = 2.9$, $MSE = 0.54$, $p = 0.096$.

4.3.3 The agree-nonagree condition

In this condition, it's at N2 that no integration is possible due to the syntactic cues telling the parser that integration is ungrammatical. Table 7 shows the costs associated with each constituent according to Gibson's model.

Table 7: Predictions from Gibson (1998) for the $I(n)$ in the Agree-Nonagree condition

Det	N1 _{Agree}	N2 _{Nonagree}	N3 _{Agree}	
	I(0)	-	I(1)	Integrate(referents) w/ det
	-	+I(0)	+I(0)	Integrate(referents) w/ Ns
Total	-	I(0)	I(1)+I(0)	

The steps at the determiner and at N1 are identical to those of the Agree-Agree condition, and so I will not describe them again here.

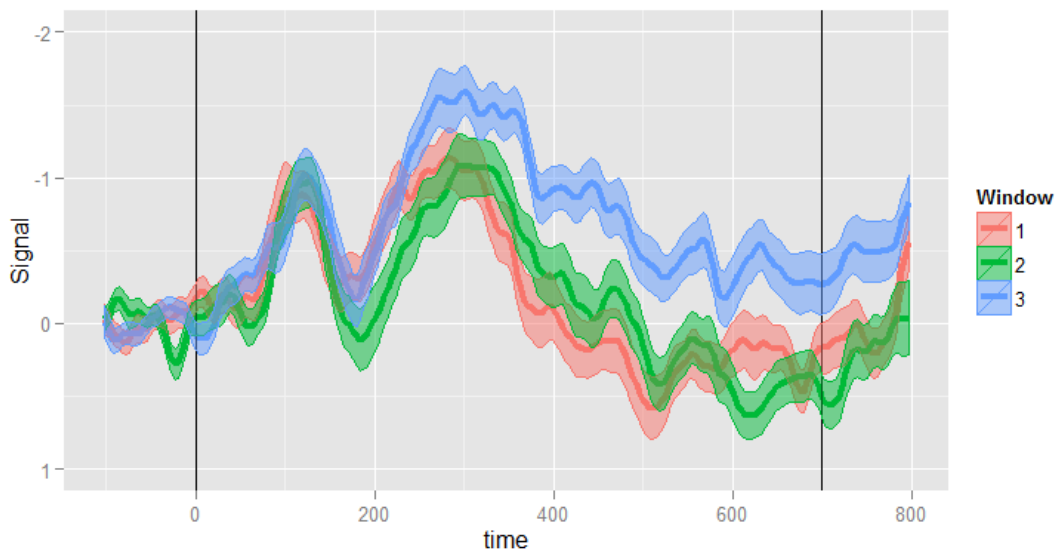
At N2, the parser now knows not to integrate this constituent with the determiner. It can, however, still form a new constituent by integrating with the noun before it. Assuming that the parser has predicted that another noun will come, it still needs to form the correct constituent structure of the triple noun compound. If the frequency of right versus left branching compounds in Icelandic mirrors the pattern seen in English, the perhaps it is reasonable to think that the parser may give preference to the left-branching structure that was used for all of the experimental items. That is, in all cases, [N1 + N2] forms a semantically reasonable combination (see Figure 1). Berg (2011), in a corpus analysis of multi-noun compounds, found that left branching structures in triple noun compounds in English are preferred to right-branching structures by a factor of nearly 3:1. If we assume that, at this step, the parser has sufficient evidence to integrate the two nouns as we know is possible, then this step still incurs a cost of that integration. This cost is represented by $I(0)$ because there are no new referents in between the two nouns being integrated.

At the arrival of N3, We again have two integrations. The first is the integration of this noun with the sub-compound created in the last step. Between these two, there are no new referents introduced into the discourse, and thus the cost is represented as $I(0)$. The second

integration cost is of combining this triple noun compound with the determiner. Since the appearance of the determiner, the parser did, at the first step with N1, interpret a new discourse referent, so the cost at this point is $I(1)$.

Notice, here, that this sets up N1 and N2 as completely identical in terms of integration costs. Thus in the response from participants, we expect the waveform for N1 and N2 to pattern together, and for N3 to separate as the one that is more costly. The waveforms for the Agree-Nonagree condition are shown in Figure 12.

Figure 12: Waveforms of N1, N2, and N3 within the Agree-Nonagree condition



The expected costs from Gibson’s model are exactly what we see borne out in the ERP responses. N1 and N2 follow each other almost exactly from 300ms on, while N3 shows a clearly more negative response. The statistical analyses show that is pattern is also highly significant. Again we see that there is an effect for the difference of all three nouns, $F(2,80) = 9.04$, $MSE = 0.72$, $p < 0.001$.

However, this pattern of differences is entirely driven by N3 being more negative than the other two conditions. There are no areas of significant between N1 and N2. Comparing the effects of N2 and N3, though, we see that N3 is more negative than N2 across the entire 300-700ms time window, $F(1,40) = 12.74$, $MSE = 0.68$, $p < 0.001$.

4.3.4 The nonagree-nonagree condition

Finally, in the Nonagree-Nonagree condition, we have the case where there is information on both N1 and N2 that integration with the determiner cannot yet occur. The expected costs according to Gibson’s model are shown in Table 8 below.

Table 8: Predictions from Gibson (1998) for the $I(n)$ in the Nonagree-Nonagree condition

Det	N1 _{Nonagree}	N2 _{Nonagree}	N3 _{Agree}	
	-	-	$I(0)$	Integrate(referents) w/ det
	-	$+I(0)$	$+I(0)$	Integrate(referents) w/ Ns
Total	-	$I(0)$	$2 I(0)$	

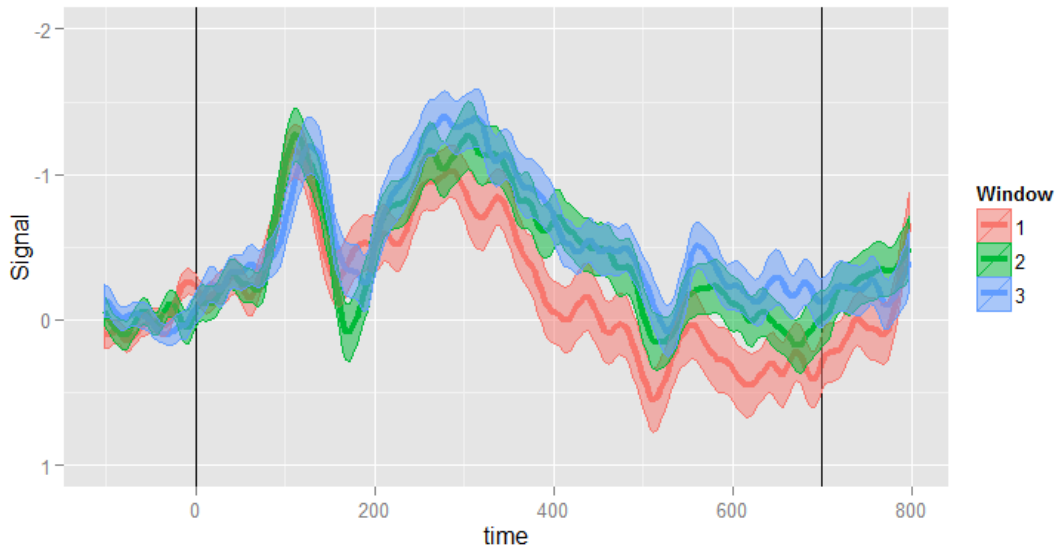
For the determiner and N1, the table is identical to what was described in the Nonagree-Agree condition and thus the reasoning will not be repeated here. At N2, we see the second instance of the inability to integrate the noun with the determiner, and so there is no cost of that process. However, it is still possible to integrate N1 and N2 together to form the sub-compound, just as was the case at N2 in the Agree-Nonagree condition. Thus at N2, a processing cost represented by $I(0)$ is incurred as N1 and N2 integrate, but there are no new discourse referents that have been introduced yet.

At N3, we can now do two integrations. N3 must integrate with the sub-compound created in the last step, and the entire triple noun compound must integrate with the determiner. In neither of these integrations, though, has a new discourse referent been introduced, so the cost incurred by each is $I(0)$.

This would predict that N2 and N3 should both produce a greater amplitude of response than N1. If it’s the case that the number of referents introduced at a given step is what’s driving the increases that have been seen at each successive constituent, then we would expect N2 and N3 to pattern together because in both cases, there have been no full referents introduced into the discourse intermediate to the relevant integrations. If, however, it’s the case that the number of integrations taking place is driving the increasing amplitudes that were seen in the last three conditions, then we would expect a pattern of response similar to

what was seen in the Nonagree-Agree condition where N3 incurs a slightly greater processing cost than N2. The responses that were measured can be seen below in Figure 13.

Figure 13: Waveforms of N1, N2, and N3 within the Nonagree-Nonagree condition



Again we see a significant difference in the comparison between N1, N2 and N3 in the 300-700ms time window, $F(2,80) = 4.31$, $MSE = 0.76$, $p = 0.017$. This difference is now driven entirely by N1 being less negative than N2 and N3. Between N1 and N2, we see that N2 is more negative than N1 in the 400-500ms window, $F(1,40) = 4.33$, $MSE = 1.09$, $p = 0.044$, with the full 300-700ms time window being marginally significant, $F(1,40) = 3.07$, $MSE = 1.0$, $p = 0.087$. There are no significant differences between N2 and N3, though.

It seems that now N2 and N3 seem to be patterning together such that they have a greater negativity than N1. This finding suggests two things: (i) integrating constituents is more costly in this measurement than not integrating anything, and (ii) increasing amplitudes of integrations reflect the number of new referents that were introduced into the discourse between the two things being integrated rather than the number of integrations taking place on that lexical item.

4.4 Comparisons of the patterns seen in the different conditions

Putting together the data shown for the results of the triple noun study across constituent places with the predictions of Gibson's SPLT model, we see a surprising amount of overlap. Table 9 provides a summary of the areas of significance for each comparison. The results represent a series of ANOVAs run at each relevant 50ms timestep. Only values that are significant below an alpha level of 0.05 are shown for ease of interpretation, but Appendix D shows the F -values and significance level of every timestep as a more complete representation of the findings.

Table 9: Table of significant values (F values shown) for each time window, comparison, and condition

Comparison	200-250ms	250-300ms	300-350ms	350-400ms	400-450ms	450-500ms	500-550ms	550-600ms	600-650ms	650-700ms
Agree-Agree condition										
1x2			5.719*	6.420*						
2x3				8.127†	21.543‡	12.708‡	6.291*		5.950*	4.714*
1x3			16.992‡	37.559‡	34.847‡	13.558‡	7.619†	7.688†	16.043‡	8.768†
Nonagree-Agree condition										
1x2										
2x3										
1x3			12.168†	12.45‡	6.641*	6.446*			6.445*	5.223*
Agree-Nonagree condition										
1x2	6.509*									
2x3	10.845†	16.343‡	10.221†	14.144‡	15.639‡	8.772†	10.777†	7.405†	17.595‡	7.444†
1x3		8.034†	12.045†	20.979‡	21.613‡	19.469‡	13.579‡	7.033*	5.315*	7.902†
Nonagree-Nonagree condition										
1x2				4.483*	5.145*	5.708*				
2x3		4.626*								5.130*
1x3	4.368*	7.622†	9.683†	12.681‡	9.518†	12.015†	8.522†	11.375†	10.958†	8.581†

Key: $p < 0.001$: ‡; $p < 0.01$: †; $p < 0.05$: *

The important point to note here is that the pattern of results differs between experimental conditions. In the Agree-Agree condition, all comparisons are significant, at least for a short duration. This means that the processing cost rises sharply from N1 to N2 and from N2 to N3. In the Nonagree-Agree condition, there is only a significant difference between N1 and N3, with N2 in between and not significantly different from either on its own. This result means that processing increases from N1 to N2 and from N2 to N3, but not enough to reach a level of significance needed. The Agree-Nonagree condition shows that N2 is patterning completely with N1, and that both differ from N3 significantly and across the entire time window. In the Nonagree-Nonagree condition, though, N2 is here patterning with N3, and together N2 and N3 are more negative than N1.

Now, compare Table 9 to a summary table of total integration costs measured at each constituent in each condition, shown in Table 10.

Table 10: Summary of the total integration cost at each constituent in each condition

Condition	N1	N2	N3
Agree-Agree	I(0)	I(1)+I(0)	I(2)+I(0)
Nonagree-Agree	-	2 I(0)	I(1)+I(0)
Agree-Nonagree	I(0)	I(0)	I(1)+I(0)
Nonagree-Nonagree	-	I(0)	2 I(0)

Remarkably, the predictions made by Gibson’s model with regard to integration costs closely predicts the pattern seen across the data summarized in Table 9. This finding is strong support for both Gibson’s SPLT model and for the commitment and revision style models.

When it comes to the process of revision, this pattern of data suggests that we should take very seriously the importance that Gibson places on introduced referents and the need, in this study, to count those incremental (but not true) referents as if they were really created. This is roughly analogous to the importance that the commitment-and-revision-style approaches place on revising licit commitments. Whereas Gibson’s theory does not explicitly state how the process of revising structure would happen in terms of memory and integration costs (he

focuses instead on some processing questions of maintaining ambiguities), this study adds nuance to his theory in an area that was left relatively unexplored. Crucially, the parser *must* be keeping track of these syntactic commitments and not only integrate them, but integrate them *as new discourse referents*.

Even more importantly, these findings do not necessarily speak to the importance of one model over another. Rather, both models are informing each other, and both models are supported by this finding. Whether it makes sense to see if other costs of revision can be cached out via predictions from the SPLT remains to be seen, though it is certainly a path worth investigating further.

Additionally, this finding opens the door to looking at processing costs across constituent places when looking at nominal compounds as another way of testing theories. Traditional analyses use comparisons like those used here in Chapter 3. These comparisons have the benefit that inferential statistics can be run on the values in a direct comparison from one condition to the next. It is not completely obvious how one would go about comparing the patterns noted here across conditions, and is something that can be developed further. One possibility could be to use difference waves to show that, for example, the differences between N1 and N2 in one condition is different from the differences between N1 and N2 in another condition.

CHAPTER 5

NEXT STEPS AND CONCLUSION

5.1 What's next

A number of open questions were noted throughout this study, and all of these questions potentially lend themselves to further experiments. This section returns to many of those questions and outlines the ways that we could investigate them and what could be gained by those investigations.

The first open question addresses the nature of the stimuli themselves. It was often the case, due to the high degree of syncretism in Icelandic, that an agreeing form at N1 or N2 would not match with the final gender, number, and case information on N3. The question is whether the parser is actively keeping those specific agreement features active (or whether they have been set) after making a commitment. If its the case that the parser commits not just to a structure, but to the featural information on different constituents, then it may be the case that some kind of higher processing cost can be measured in conditions such as Agree-Agree when there is no change in case, gender, and number features between agreeing constituents and when at least one of those features changes from one noun to the next, while still satisfying the agreement requirements. If there is no effect, that would be consistent with the parser only combining when it can and not combining when it has a mismatch. If there is an effect, though, it would mean that the information activated by the parser includes features and that these are, perhaps, an important part of the computation. Thus such a study could shed light on the questions related to the internal representation of grammatical features and how the parser handles those features.

Along this same line of thought, there were some cases where a root form of the noun was used. Upon more careful consideration, these conditions should, perhaps, have been left out. Root forms have at least one clear difference from all the other forms: they can never head

a phrase. Once a root is encountered, just the fact that it is a root form is enough to let the parser know that there will be a mismatch and that, for a grammatical sentence to result, another noun must follow it. For items that did not have root forms of the nouns, though, the parser would need to check if the features are a match in order to know whether or not the constituent fits into the structure. These two are, then, different in that the parser may know from just the form of the noun that it is not the end of the phrase. The same would not be true of a noun bearing the full case, gender, and number features, which must actually be computed in order to know if they fit with the determiner. For the root cases, it would also be possible to say that statistical knowledge alone could let the parser know that a root form will not be the head, as this should have a frequency in the language of 0. Thus a study investigating the effects of roots separate from case markings may prove fruitful again in giving information about the nature of featural representations in the brain.

Additional question open of *what* is being predicted when predictions are made also warrants a closer look. We consistently haven't found effects consistent with reanalysis on N1 when it mismatches from the determiner/adjective. However, if a noun's features are one of the things analyzed, then perhaps we can measure some effect of revision on the first noun, as a determiner should be sufficient to predict a noun will come. Based on the information on the determiner, the case, number, and gender features can be narrowed down. If there is no effect for revision, though, this would imply that the nature of the predictive mechanism may not act on featural information itself. Perhaps only the word category is what is predicted. If it is word category and not featural information causing causing a need to revise, then we would not expect to find revision costs on any N1.

Looking at the analyses conducted in Chapter 4 for the predictions of Gibson's SPLT model, another possible study testing the processing costs of a different number of integrations came up. This study only used left-branching compound structures, and because of this only two integrations were needed on each N3. If we had used right-branching compound structures, though, there would have been three integrations needed at N3: (i) integrating

N3 with N2, (ii) integrating [N2 + N3] with N1, and (iii) integrating the whole triple noun compound with the determiner.

Additionally, as part of this study focused on effects of working memory, it is worth noting that a great deal of progress in working memory research has been in linking and finding correlations in performance among many working memory tasks. Looking for individual differences in working memory capacity, similar to Vos, Gunter, Schriefers, and Friederici (2001), the question is whether we expect ERP responses to correlate with those differences. It is certainly worth considering that looking at individual participants and getting measures of general working memory capacity, when possible, would expand and extend the current study.

And the final open question I will address here is with regarding the statistical tests available to compare across conditions. At present, the study was only able to look at the pattern of significant results across studies by running analyses within conditions. This strategy does give a holistic picture of the patterns that are measured. However, an actual way to use inferential statistics and test the differences across conditions would be preferable and potentially carried forward through future studies.

5.2 Conclusion

Taken all together, the results of this study strongly support both the models of commitment and revision and Gibson's (1998) SPLT model of sentence processing, as both were predictive of where costs would be measured. This study extended previous work done with Germanic compounds and introduced a new analytical method for looking at ERP results by looking across constituent places. Furthermore, this study shows that the processing costs related to commitments to a structure remain so long as a new discourse referent was introduced as a result of that commitment, even when that commitment needs to be revised.

APPENDICES

APPENDIX A

SAMPLE EXPERIMENTAL ITEMS WITH GLOSSES

- (1) Ég keypti fleiri ... í-dag
I bought more ... today

- a. lófa síma hulstur
palm.m.pl.acc phone.m.pl.acc case.m.pl.acc
- b. spjald síma hulstur
panel.n.sg.acc phone.m.pl.acc case.m.pl.acc
- c. lófa tölvu hulstur
palm.m.pl.acc computer.f.sg.acc case.m.pl.acc
- d. spjald tölvu hulstur
panel.n.sg.acc computer.f.sg.acc case.m.pl.acc

- (2) Yfirmaðurinn réð aðra ... í-gær
The-supervisor hired another ... yesterday

- a. bíla geymslu verði
car.m.pl.acc storage.f.sg.acc guard.m.pl.acc
- b. bóka geymslu verði
book.f.pl.gen storage.f.sg.acc guard.m.pl.acc
- c. bíla safns verði
car.m.pl.acc museum.n.sg.gen guard.m.pl.acc
- d. bóka safns verði
book.f.pl.gen museum.n.sg.gen guard.m.pl.acc

- (3) Þau mættu-í alla ... í-síðasta mánuði
They attended all ... last month

- a. sýkla fræði tíma
germ.m.pl.gen studies.f.sg.acc class.m.pl.acc
- b. mál fræði tíma
language.n.root studies.f.sg.acc class.m.pl.acc

- c. sýkla skoðunar tíma
germ.m.pl.gen inspection.f.sg.gen class.m.pl.acc
- d. mál skoðunar tíma
language.n.root inspection.f.sg.gen class.m.pl.acc

(4) mér leiddist þessi ... í morgun
I was-bored-by this ... this morning

- a. sál fræði kennsla
soul.f.sg.nom studies.f.sg.nom teaching.f.sg.nom
- b. tón fræði kennsla
tone.m.root studies.f.sg.nom teaching.f.sg.nom
- c. sál greiningar kennsla
soul.f.sg.nom analysis.f.sg.gen teaching.f.sg.nom
- d. tón fræði kennsla
tone.m.root analysis.f.sg.gen teaching.f.sg.nom

(5) Stúlkan borðaði aðra ... með gafflinum.
The-girl ate another ... with the-fork.

- a. peru bita köku
pear.f.sg.acc bit.m.pl.acc cake.f.sg.acc
- b. súkkulaði bita köku
chocolate.n.sg.acc bit.m.pl.acc cake.f.sg.acc
- c. peru mauks köku
pear.f.sg.acc compote.n.sg.gen cake.f.sg.acc
- d. súkkulaði mauks köku
chocolate.n.sg.acc compote.n.sg.gen cake.f.sg.acc

APPENDIX B

SET OF SAMPLE FILLER ITEMS

- (1) a. Ég hitaði þrjár kaffi könnur í pásunni
I heated three coffee mugs in the-break
- b. Ég þvoði þrjár kaffi könnur í pásunni
I washed three coffee mugs in the-break
- c. Ég hitaði þrjár kaffi kökur í pásunni
I heated three coffee cakes in the-break
- d. Ég þvoði þrjár kaffi kökur í pásunni
I washed three coffee cakes in the-break
- (2) a. Stelpan útbjó ávaxta box fyrir ferðina
The-girl prepared a fruit box for the-trip
- b. Stelpan tæmdi ávaxta box fyrir ferðina
The-girl washed a fruit box for the-trip
- c. Stelpan útbjó ávaxta tertu fyrir ferðina
The-girl prepared a fruit cake for the-trip
- d. Stelpan tæmdi ávaxta tertu fyrir ferðina
The-girl washed a fruit cake for the-trip
- (3) a. Frændi minn bruggaði fimmtíu bjór flöskur í finnunni
Uncle my brewed fifty beer bottles at his-workplace
- b. Frændi minn sótthreinsaði fimmtíu bjór flöskur í finnunni
Uncle my sanitized fifty beer bottles at his-workplace
- c. Frændi minn bruggaði fimmtíu bjór drykki í finnunni
Uncle my brewed fifty beer drinks at his-workplace
- d. Frændi minn sótthreinsaði fimmtíu bjór drykki í finnunni
Uncle my sanitized fifty beer drinks at his-workplace

APPENDIX C

SET OF SAMPLE LLICIT PRE-TEST ITEMS

- | | |
|--|---|
| (1) aðra gælu línu hugson
another petting line thought | (13) annarra vél stúlku hunda
other machine girl dogs |
| (2) aðrar ofna efna tær
other oven material toes | (14) fleiri asna belju snigill
more donkey cow snails |
| (3) allar bruna brellu olboga
all burning trick elbows | (15) fleiri bryggju ugsana ráðum
more pier thought advices |
| (4) allar pils skapara lestir
all skirt creator trains | (16) fleiri gæsa kettlinga gráðum
more goose kitten degrees |
| (5) allt hana ferða vax
all rooster trip waxes | (17) gleiri nefndar tjalds nöfn
more committee tent names |
| (6) annað atriðis merkja eitur
other item tag poisons | (18) fleiri peru rúms svín
more pear bed pigs |
| (7) annað áuga fót meti
other interest foot foods | (19) fleiri refa frosks krukkur
more fox frog jars |
| (8) annað gagn bolta mál
another use ball case | (20) fleirum landa skaps vöðvum
more land mood muscles |
| (9) annað ullar hænsna ljóð
another wool chicken poem | (21) meiri dýra vals súpa
more animal choice soup |
| (10) annar blað tækis vegur
another blade tool road | (22) meiri efna svína hiti
more material pig heat |
| (11) annar leður lána gíraffi
another leather loan giraffe | (23) meiri flutninga rottu keppni
more moval rat competition |
| (12) annar trukka miða markaður
another truck ticket market | (24) hella skýja sjór
more cave cloud oceans |

APPENDIX D

TABLE OF *F*-VALUES FOR ALL COMPARISONS

Key: $p < 0.001$: ‡; $p < 0.01$: †; $p < 0.05$: *

Table 11: F-values for all comparisons

Agree-Agree condition															
Comp- arison	0- 50ms	50- 100ms	100- 150ms	150- 200ms	200- 250ms	250- 300ms	300- 350ms	350- 400ms	400- 450ms	450- 500ms	500- 550ms	550- 600ms	600- 650ms	650- 700ms	700- 750ms
1x2	0.024	1.604	0.276	1.261	0.113	0.531	5.719*	6.420*	1.549	0.003	0.002	0.037	0.075	0.415	0.047
2x3	2.327	0.888	0.609	0.143	0.004	0.415	2.227	8.127†	21.543‡	12.708‡	6.291*	3.907	5.950*	4.714*	5.989*
1x3	2.604	0.135	1.485	0.497	0.140	2.210	16.992‡	37.559‡	34.847‡	13.558‡	7.619†	7.688†	16.043‡	8.768†	11.051†
Nonagree-Agree condition															
Comp- arison	0- 50ms	50- 100ms	100- 150ms	150- 200ms	200- 250ms	250- 300ms	300- 350ms	350- 400ms	400- 450ms	450- 500ms	500- 550ms	550- 600ms	600- 650ms	650- 700ms	700- 750ms
1x2	1.048	0.065	2.335	0.378	0.030	0.004	1.862	3.220	0.621	0.664	0.277	1.347	0.815	2.641	2.517
2x3	0.397	0.372	2.300	0.007	0.148	0.750	4.006	2.788	3.449	2.165	1.616	0.035	0.791	0.408	1.008
1x3	0.111	0.155	0.035	0.659	0.043	0.757	12.168†	12.45‡	6.641*	6.446*	3.356	2.349	6.445*	5.223*	7.218*
Agree-Nonagree condition															
Comp- arison	0- 50ms	50- 100ms	100- 150ms	150- 200ms	200- 250ms	250- 300ms	300- 350ms	350- 400ms	400- 450ms	450- 500ms	500- 550ms	550- 600ms	600- 650ms	650- 700ms	700- 750ms
1x2	0.120	0.904	0.019	2.271	6.509*	1.311	0.376	1.889	0.024	1.526	0.005	0.001	2.867	0.000	1.168
2x3	0.350	2.267	0.366	2.560	10.845†	16.343‡	10.221†	14.144‡	15.639‡	8.772†	10.777†	7.405†	17.595‡	7.444†	9.190†
1x3	0.015	0.013	0.559	0.094	0.299	8.034†	12.045†	20.979‡	21.613‡	19.469‡	13.579‡	7.033*	5.315*	7.902†	3.224
Nonagree-Nonagree condition															
Comp- arison	0- 50ms	50- 100ms	100- 150ms	150- 200ms	200- 250ms	250- 300ms	300- 350ms	350- 400ms	400- 450ms	450- 500ms	500- 550ms	550- 600ms	600- 650ms	650- 700ms	700- 750ms
1x2	0.517	0.304	2.124	3.943	1.906	1.270	2.608	4.483*	5.145*	5.708*	2.280	3.931	4.023	1.458	3.236
2x3	0.352	0.018	0.016	5.853*	0.950	4.626*	2.918	1.505	0.737	1.582	2.960	3.868	3.224	5.130*	0.845
1x3	0.070	0.164	2.121	0.112	4.368*	7.622†	9.683†	12.681‡	9.518†	12.015†	8.522†	11.375†	10.958†	8.581†	6.760*

APPENDIX E

LONG WINDOWS

Figure 14: Long window showing each electrode for each condition

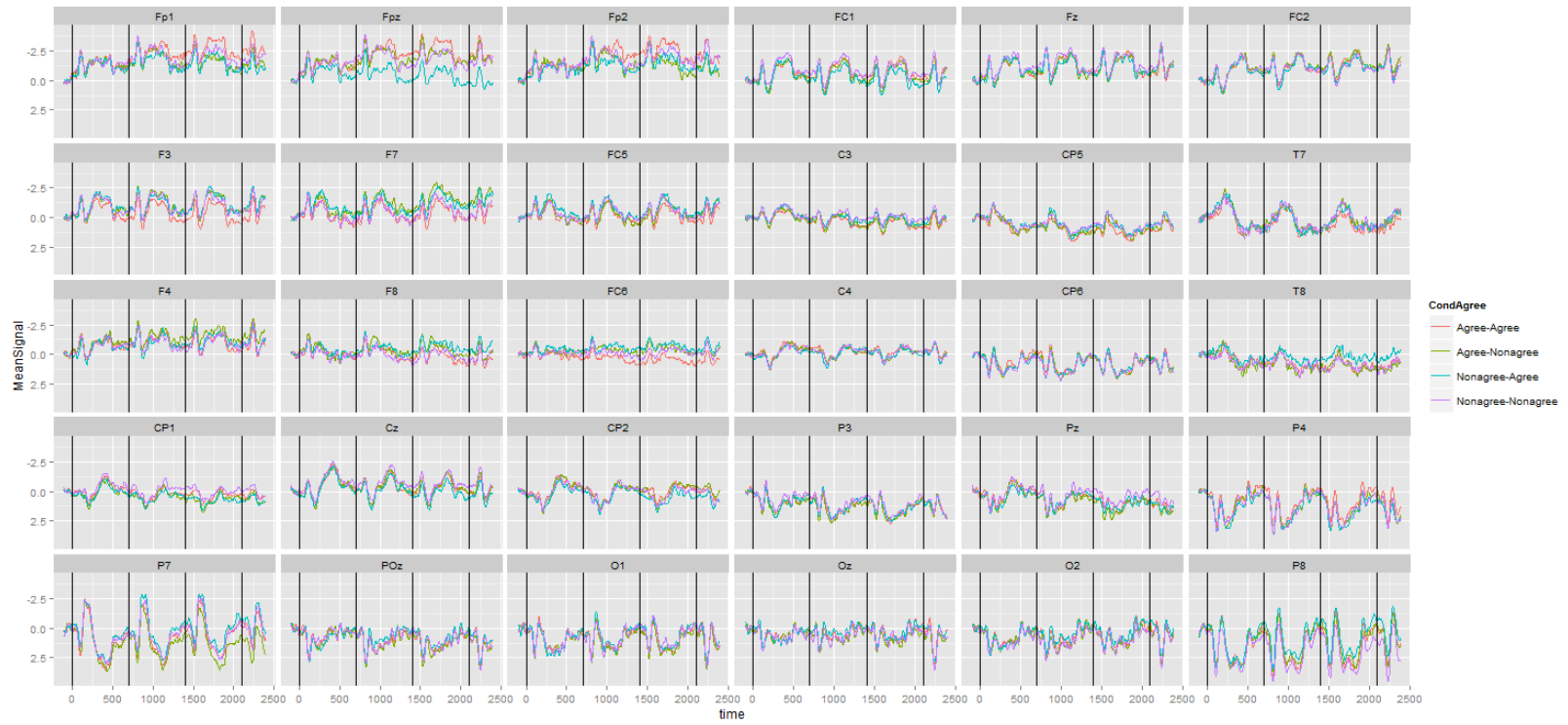
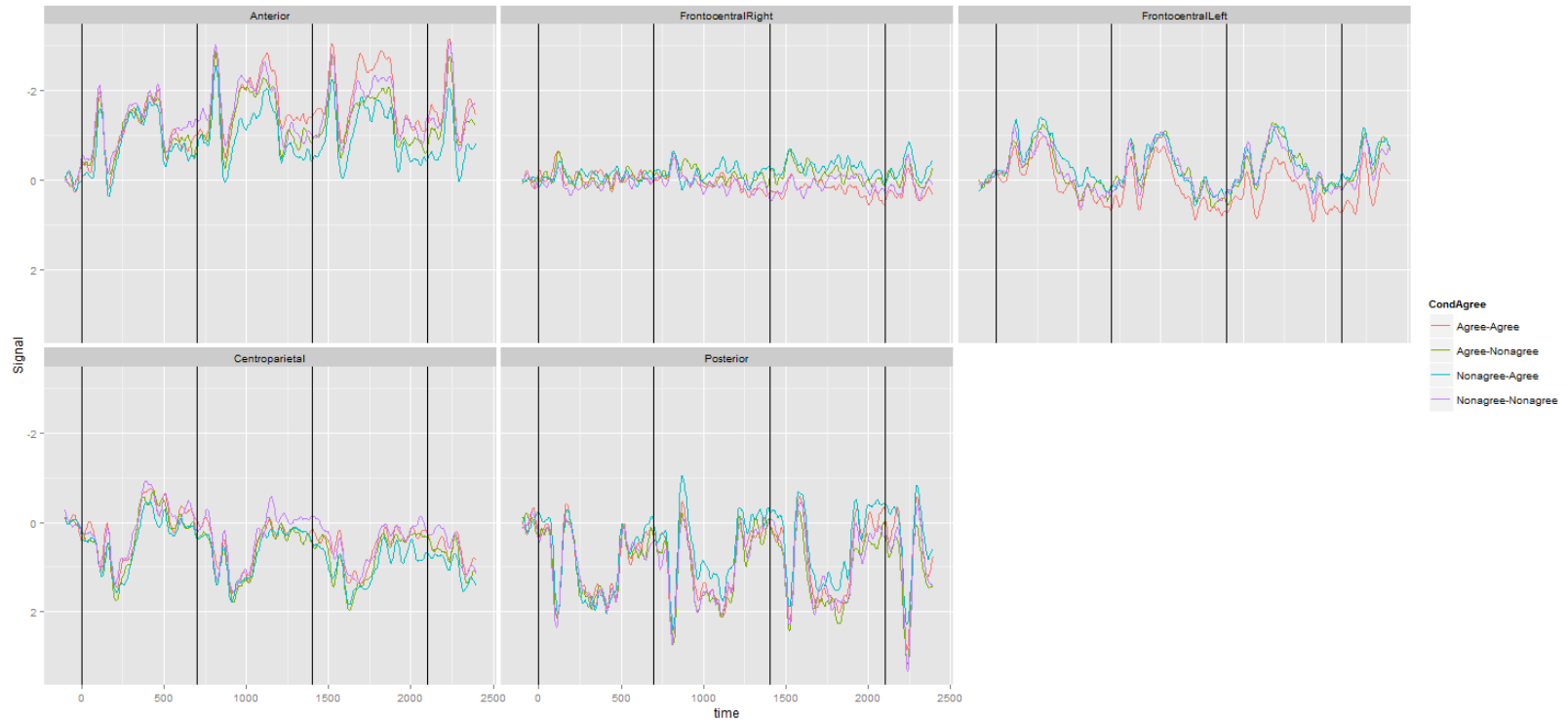


Figure 15: Long window showing the five ROIs for each condition



BIBLIOGRAPHY

BIBLIOGRAPHY

- Baddeley, A. (2010). Working memory. *Current Biology*, 20(4).
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1-29.
- Berg, T. (2011). The modification of compounds by attributive adjectives. *Language Sciences*, 33(5), 725–737.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain responses to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21-58.
- Cowan, N. (2005). *Working memory capacity*. Psychol. Press.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-21.
- 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.
- Downing, P. (1977). On the creation and use of english compound nouns. *Language*, 810–842.
- E Pratarelli, M. (1995). Modulation of semantic processing using word length and complexity: An erp study. *International journal of psychophysiology*, 19(3), 233–246.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, 102(2), 211-245.
- Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's Area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping*, 24, 79-91.
- Fiorentino, R., & Poeppel, D. (2007). Processing of compound words: An meg study. *Brain and Language*, 103(1), 18–19.
- Freud, S., & Brill, A. A. (1938). Psychopathology of everyday life.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and language*, 50(3), 259–281.

- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 27–52.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38(11), 1531-1549.
- Isel, F., Gunter, T. C., & Friederici, A. D. (2003). Prosody-assisted head-driven access to spoken German compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(2), 277.
- Jackendoff, R. (2009). *Compounding in the parallel architecture and conceptual semantics*. na.
- Jalbert, J., Roberts, T., & Beretta, A. (2016). Neurophysiological effects of prediction on head reassignment in German compounds. *Neuroreport*, 27, 186-191.
- Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1), 98-110.
- Kennison, S. M. (2005). Different time courses of integrative semantic processing for plural and singular nouns: implications for theories of sentence processing. *Cognition*, 97(3), 269–294.
- Koester, D. (2014). Prosody in parsing morphologically complex words: Neurophysiological evidence. *Cognitive neuropsychology*, 31(1-2), 147–163.
- Koester, D., Gunter, T. C., & Wagner, S. (2007). The morphosyntactic decomposition and semantic composition of German compound words investigated by ERPs. *Brain and Language*, 102(1), 64–79.
- Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, 16(9), 1647–1668.
- Koester, D., Holle, H., & Gunter, T. C. (2009). Electrophysiological evidence for incremental lexical-semantic integration in auditory compound comprehension. *Neuropsychologia*, 47(8), 1854–1864.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4, 463-470.

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(14), 1-14.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.
- Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009, December). A lexical basis for N400 context effects: Evidence from MEG. *Brain & Language*, *111*(3), 161-172.
- Lorenz, A., Madebach, A., & Jescheniak, J. D. (2017). Grammatical-gender effects in noun-noun compound production: Evidence from German. *The Quarterly Journal of Experimental Psychology*.
- MATLAB. (2010). *version 7.10.0 (r2010a)*. Natick, Massachusetts: The MathWorks Inc.
- Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2014). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex*, 1-4.
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*, 908-930.
- Montalbetti, M. M. (1984). *After binding: On the interpretation of pronouns* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Münte, T. F., & Meinze, H.-J. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, *5*(3), 335-344.
- Neville, H., Nicol, J., Barsz, A., Forster, K., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Cognitive Neuroscience, Journal of*, *3*(2), 151–165.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786.
- Parrish, A., Jalbert, J., & Beretta, A. (2015). *Head commitment and plausibility in English noun-noun compounds*. (Poster presented at The 7th Annual Meeting of the Society for the Neurobiology of Language (SNL), October 2015, Chicago, IL)
- Parrish, A., Kelley, P., & Beretta, A. (2016). *Plausibility and agreement effects of adjectives on noun-noun compounds in Icelandic: An ERP study*. (Poster presented at The 8th Annual Meeting of the Society for the Neurobiology of Language (SNL), August 2016, London, England)

- Phillips, C. (1996). *Order and structure* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407-428.
- Pierce, J. W. (2007). Psychopy - psychophysics software in python. *Neurosci Methods*(162).
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology*, *42*(3), 529–567.
- Scalise, S., & Guevara, E. (2006). Exocentric compounding in a typological framework. *Lingue e linguaggio*(2), 185–206.
- Sprouse, J., & Lau, E. F. (2013). Syntax and the brain. In M. den Dikken (Ed.), *The cambridge handbook of generative syntax*. Cambridge University Press.
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1162.
- Tanner, D. (2006). On the left anterior negativity (lan) in electrophysiological studies of morphosyntactic agreement: A commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Corex*, *66*, 149-155.
- Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes*, *16*(1), 65-103.
- Whelpton, M., Trotter, D., Beck, P. G., Anderson, C., Maling, J., Durvasula, K., & Beretta, A. (2014). Portions and sorts in icelandic: An erp study. *Brain and language*, *136*, 44–57.
- Zwitserslood, P. (1994). The role of semantic transparency in the processing and representation of dutch compounds. *Language and cognitive processes*, *9*(3), 341–368.