UNIVERSITY OF CALIFORNIA
Los Angeles

The Role of Prosodic Phrasing on the Acceptability of Agreement Attraction in English

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Linguistics

by

Adam J. Royer

2021

ABSTRACT OF THE DISSERTATION

The Role of Prosodic Phrasing on the Acceptability of Agreement Attraction in English

by

Adam J. Royer
Doctor of Philosophy in Linguistics
University of California, Los Angeles, 2021
Professor Jesse Harris, Co-Chair
Professor Sun-Ah Jun, Co-Chair

When a subject NP has a singular head noun and a plural noun in some lower syntactic phrase (i.e. local noun), occasionally a plural verb will be produced in a sentence (i.e., agreement attraction) (Bock 1991,Bock et al. 2001). Evidence from production (Eberhard 2005) and comprehension (Badecker 2007, Wagers 2009) studies have conflicting accounts for the mechanisms at play in these agreement attraction sentences (i.e. Marking and Morphing and cues-based retrieval). As of yet, however, neither account has incorporated prosody into our understanding of agreement despite what is known about prosody's role in sentence processing (Frazier 2006). This study bridges these areas of processing by investigating the role of phrasing in the processing of subject-verb agreement. Results of the current study show that while prosodic phrasing has little to no direct effect on the agreement mechanism's accuracy, phrasing an intervening plural local noun and a plural verb into separate intonation phrases does speed up how quickly participants judge these agreement attraction sentences as unacceptable.

The dissertation of Adam J. Royer is approved.

Shari Speer

Tim Hunter

Sun-Ah Jun, Committee Co-Chair

Jesse Harris, Committee Co-Chair

University of California, Los Angeles

2021

To my mother Joan Royer . . .

who—when telling people I study linguistics—would say,

"Adam's studying linguistics so he can figure out what the hell I'm trying to say."

TABLE OF CONTENTS

1 Introduction **1**

1.1 What is this dissertation about? 1

1.2 Agreement Attraction 3

1.2.1 Representation vs. Retrieval Accounts 7

1.2.2 Where does Prosody Fit In? 17

1.3 Prosodic Hierarchy and Syntax 18

1.3.1 Prosodic Phrasing's Role in Processing 20

1.3.2 Pitch Accents and Processing 23

1.4 How does it all connect? 25

1.5 Goals and scope of this dissertation 25

1.6 Research Questions 26

1.7 Hypotheses 27

1.7.1 Hypothesis 1: The Chunk-Internal Availability Hypothesis 27

1.7.2 Hypothesis 2: The Accentual Processing Boost Hypothesis 30

1.8 Overview of Thesis 33

2 Experiment 1: Agreement Attraction Production **34**

2.1 Methods 34

2.1.1 Stimuli Design 36

2.1.2 Procedure 43

2.1.3 Data Processing and Transcription 43

2.2 Results 44

2.2.1 Intonational analysis 45

2.3 Discussion 49

2.3.1 Similar rate of attraction as Bock and Miller (1991) 49

2.3.2 Implications of Explicit Prosody on Reading Studies' Results . . . 50

3 Experiment 2: Speeded Acceptability 55

3.1 Review of Hypotheses 55

3.2 Methods 56

3.2.1 Stimuli Design 56

3.2.2 Procedure 66

3.2.3 Participants 66

3.2.4 Data Analysis and Statistical Methods 68

3.3 Results 70

3.3.1 Review of Main Questions and Hypotheses 70

3.3.2 Ratings 72

3.3.3 Time-to-decision RTs 77

3.4 Discussion 88

3.4.1 Ratings support grammaticality asymmetry 88

3.4.2 No detectable effect of phrasing on ratings 88

3.4.3 Grammatical symmetry in RTs: Slowdown for mismatched subject NPs 90

3.4.4 Phrasing facilitates time-to-decision for acceptability 90

4 General Discussion 94

4.1 Overview of findings 94

4.1.1 What patterns of phrasing and accenting are typical for agreement attraction sentences? Does the intonation of agreement attraction sentences differ from non-attraction sentences? 94

4.1.2 How does phrasing influence the acceptability of agreement attraction sentences? 95

4.1.3 How does phrasing affect the speed with which participants determine the source of agreement? 96

4.2 The Bigger Picture 96

4.2.1 Representation matters 98

4.3 Future directions 99

4.3.1 Incorporation of prosody into computational models 99

4.3.2 Variety of syntactic configurations 99

4.4 Conclusions 101

4.4.1 What have we seen, and what does it all mean? 101

A Appendix: Stimuli 111

LIST OF FIGURES

1.1 A screenshot of a tweet by Carissa Byrne Hessick that shows a sign that says "Any Use Of Tobacco Products Is Prohibited In Chapel Hill Parks" with a "no smoking" logo on the sign. The word "is" has been crossed out and written over with "ARE" in an example of agreement attraction of the local noun "tobacco products". 4

1.2 Diagram of the components of number marking in production. Taken from Eberhard et al. (2005) 8

1.3 Two syntactic trees showing a morphologically plural word closer (left) and farther (right) from the root node of the DP. Highlighted in blue is the head noun of the subject NP and highlighted in yellow are the words with plural morphology. 10

1.4 Syntactic tree diagram labeled with the values for *S(n), S(m),* and *w* which go into calculation of *S(r)* for "The key to the cabinets". Taken from Eberhard et al. (2005). 11

1.5 Diagram taken from Lewis & Vasishth (2005) showing the declarative chunks and their associated cues for the sentence "The writer surprised the editors." . 15

1.6 Examples pitch tracks from (Frazier et al., 2014) showing the early boundary (top) and late boundary conditions (bottom). 22

2.1 Two examples of stimuli showing the difference between the SAME IP condition (2.1a) and the DIFFERENT IPS condition (2.1b) 40

2.2 Two examples of production stimuli showing the durational and f0 between the "same IP" condition (2.2b) and the "different IP" condition (2.2a). 42

2.3 Comparison of the break sizes after the local noun. The number in the middle of each bar is the raw counts of that break size. 47

2.4 Comparison of the break size and tonal markings after the local noun. The number in the middle of each bar is the raw counts of that tonal type for that break size. 48

2.5 Comparisons of accent types on head and local nouns in attraction and non-attraction productions. 52

2.6 Counts of the different combinations of accent types that appeared on the head noun and local noun. 53

2.7 Counts of the accent and boundary size combinations used on the plural local noun and singular/plural verb. For clarity, the counts for nuclear pitch accents are in red and prenuclear accents are in blue. 54

3.1 The spectrograms show the final syllable of the subject NP "The reason for the fort(s)". The top row shows stimuli from the SAME IP condition and the bottom row shows the DIFFERENT IPS condition. The left column shows the plural forms and the right column shows the singular forms. 60

3.2 Duration of the Local Noun across prosody and plurality conditions. 63

3.3 89% credible intervals for the parameter estimates produced by the model. This includes main effects of local noun number, verb number/grammaticality, and phrasing condition. The result of a Test of Practical Equivalence. 64

3.4 A screenshot of the two instruction pages that preceded the practice trials. . . 67

3.5 A diagram showing the sequence of actions in a given trial. 67

3.6 The *brms* code used to run the Bayesian hierarchical model for participant acceptability judgements. 70

3.7 The *brms* code used to run the Bayesian hierarchical model for participant RTs. 71

3.8 Posterior Predictive Check comparing the actual data, y, with 100 simulated data based on the model predictions, y_{rep}. 73

3.9 Average acceptability rating for each condition. '1' is an acceptable response and '0' is an unacceptable response. 75

3.10 89% credibility intervals for estimates of the Ratings parameters. 76

3.11 Posterior Predictive Check comparing the actual RT data, y, with 100 simulated datasets based on the model predictions, y_{rep}. 79

3.12 Samples from the posterior predictive distribution showing for participant RT by condition. 80

3.13 89% Credible Intervals for estimates of the RT parameters. 83

3.14 Estimated marginal means pairwise comparisons for Mismatch x Grammaticality x Response Type 84

3.15 Estimated marginal means pairwise comparisons for Mismatch x Phrasing x Response Type 85

3.16 89% Credible Intervals for estimates of the RT parameters. 86

3.17 Samples from the posterior predictive distribution just for Mismatch conditions. 92

LIST OF TABLES

1.1 Schematic representation of the agreement process in the phrasing conditions. The purple boxes surrounding the words represent IPs. The em dashes refer to word boundaries, and the size of a word reflects its accessibility or level of encoding. Larger words are more accessible than normal size words, which in turn are more accessible than smaller words. 29

1.2 Schematic representation of the agreement process in the phrasing conditions. The purple boxes surrounding the words represent IPs. The em dashes refer to word boundaries, and the size of a word reflects its accessibility or level of encoding. Larger words are more accessible than normal size words, which in turn are more accessible than smaller words. 32

2.1 2x2 table showing how the sentences varied across local noun number and the verb number conditions. Red boxes with a drop-shadow above indicate singular word forms and blue boxes with a drop-shadow below indicate plural word forms. The head noun and verb are in square boxes to indicate that the head noun is the the expected source of agreement for the verb. The intervening local noun is in a box with rounded edges to indicate is it not the expected source of agreement. 38

2.2 Schematic representation of the phrasing conditions using the agreement attraction condition (MISMATCH X UNGRAMMATICAL) as an example. The purple boxes surrounding the words represent IPs. 38

2.3 Example of stimuli differences 41

2.4 Counts of head noun, local noun, and verb number usage. The row highlighted in yellow represents the agreement attraction cases. 45

3.1 Predictions about the effect of phrasing on the acceptability of agreement attraction sentences for both prosodic hypotheses. 56

3.2 Predictions 57

3.3 2x2 table showing how the sentences varied across local noun number and the verb number conditions. Red boxes with a drop-shadow above indicate singular word forms and blue boxes with a drop-shadow below indicate plural word forms. The head noun and verb are in square boxes to indicate that the head noun is the the expected source of agreement for the verb. The intervening local noun is in a box with rounded edges to indicate is it not the expected source of agreement. 60

3.4 Schematic representation of the phrasing conditions using the agreement attraction condition (SP X UNGRAMMATICAL) as an example. The purple boxes surrounding the words represent IPs. 61

3.5 Examples of the filler items used. 1-3 are grammatical, 4-5 are semantically anomalous, 6 is marginally grammatical, and 7-8 are ungrammatical by Mainstream US English standards. 62

3.6 Coefficients of stimuli duration model 65

3.7 Examples of the levels of each categorical factor in the model. 70

3.8 Counts of acceptable and unacceptable responses broken down by condition. . 74

3.9 Parameter estimates from the ratings model. 77

3.10 Mean winsorized RTs in milliseconds and their standard errors in parentheses by condition. 81

3.11 Pairwise comparison of the estimated marginal means of the interaction between local noun number and response type. 82

3.12 Table of median parameter estimates with 89% Credible Intervals from the RT model. 87

ACKNOWLEDGMENTS

What a wild ride it was to write a dissertation during a global pandemic! I definitely would not recommend it. I'd give it a -6/10. I start with this to preface how there is no way I could have finished this dissertation without the love, care, and support of so many people.

I would first like to thank my committee for their unending fountain of patience with me through my struggles. To Tim, I really appreciate the feedback you've given me on this project and also waking up at an ungodly hour for my dissertation defense. To Jesse, you were an immense help in this process and I appreciate how you helped me to see how I could deepen and expand my work. To Shari, I will never forget walking into your office (before even the first day of my freshman year at OSU!) and how accepting you of me and how willing you were to take a chance on me. To Sun-Ah, you shepherded me through this Ph.D. with unconditional support and helped me to realize my potential as a linguist and intonationist. To you all I will forever be grateful.

In the following I would like to call out some other individuals who helped me get to where I am. If I omitted you, I apologize.

Kiwa, it has been such a fun trip working with you and going from your research assistant to a collaborator of yours. I have always enjoyed thinking about my time in Pomerene Hall and all the great eye tracking we did together.

Scott! The Systema practices, our Buddhism discussions, our Peet's hangouts, and your friendly, listening ear helped me to keep my sanity when everything felt like it was falling to pieces. "SERENITY NOW! SERENITY NOW!"

Lavena, my bestest bestie. My PLP. WE DID IT! As a whiny rapper once said, "Started from the bottom, now we're here / Started from the bottom, now my whole team fuckin' here".

Ann, it's difficult to find the right words that match how grateful I am for our friendship. From the get-go we got each other in a way that few else had. You made me feel at home in a place that felt so alien. You are the best and baddest swamp bitch I know.

Iara, I will never forget the day in the TA office hours room in Rolfe when you didn't hesitate to say that I was "R1 material" and the confidence boost that that gave me. I have always respected the work you do, so that meant a great deal to me, especially when I was feeling so inadequate.

Isabelle, without art parties to bring ourselves back to human form, what would we have done? I thank you so much for your kindness and all the wonderful museum trips we've made.

Beth, you know just what this last year was like for me, and your encouragement has meant the world. To quote Michael Scott, "Sometimes I'll start a sentence and I don't even know where it's going. I just hope I find it along the way." With your help, I finished those sentences.

Estephanie, you and I started our journey into linguistics at the same time and its been great getting see how far we've both come. You made OSU just a lovely place to be, even when time were hard.

Margit, whenever I had a random semantics question, you'd indulge me which I loved so much. In the words of Icy Spicy Leoncie in "Come On, Viktor", "Your tender loving arms mean so much".

To all my RAs throughout the years, you all are the REAL MVPs! You all are the oft unseen heroes of countless research projects and I want y'all's contributions to be celebrated! You all not only helped me to complete by projects, but you helped me to grow as a mentor and I hope that I was able to help you all grow as academics and people as well. Don't forget that you can always reach out to whenever for literally anything no matter how long it has been since we've last spoken.

To my brother, Keith. You have always believed in me and my potential.

Finally, I would like to thank my mom, who passed away the month before I started at UCLA. It has been rough with you physically here, but I know you've been with me all along this wild journey I've been on, and you will continue to accompany me in my new chapter of life.

CHAPTER 1

Introduction

1.1 What is this dissertation about?

This dissertation focuses on the role of prosodic phrasing in the processing of subject-verb agreement. Agreement attraction is a broad term for grammatical mismatches between two words in a syntactic dependency relationship. While this can refer to a variety of syntactic dependencies, the classic example used is English subject-verb agreement. Sentences like Example (1) present an intriguing puzzle because the verb "are" appears to agree in number with the noun in the prepositional phrase, "cabinets", as opposed to the head of the subject noun phrase, "key" and yet still are judged as grammatical by some speakers. English speakers may agree that Example (2) is grammatical and Example (3) is ungrammatical. However, they may disagree about whether or not Example (1) is grammatical or not despite sharing those other grammaticality intuitions.

(1) ?The key to the cabinets are on the table.

(2) The key is on the table.

(3) *The key are on the table.

A large body of psycholinguistic work on the nature of agreement attraction has been developed over the last 30 years. The primary focus of this research has been on the nature of the linguistic representations involved in agreement and the mechanism that results in agreement attraction. Many factors have been found to contribute to agreement attraction, including, but not limited to, the presence of plural morphology, the syntactic structure of the agreeing phrase, and the plurality inherent in a word's meaning. One under-

investigated area, however, has been prosody's role in agreement. Work on intonation in sentence processing has revealed that prosodic phrasing can guide syntactic parsing online and delimit domains of interpretation. Subject-verb agreement is a phenomenon heavily influenced by syntactic and semantic constraints, which is why this gap in the agreement literature is worth investigating further.

This dissertation uses the Autosegmental-metrical model of English intonation to examine the role of intonation in the processing of agreement attraction. The aim of this study is to shed light on questions left unaddressed by accounts of agreement attraction and the agreement mechanism more generally. The questions this dissertation addresses are as follows:

(1) What patterns of phrasing and accenting are typical for agreement attraction sentences? Does the intonation of agreement attraction sentences differ from non-attraction sentences?

(2) How does phrasing influence the acceptability of agreement attraction sentences?

(3) How does phrasing affect the speed with which participants determine the source of agreement?

The research questions of this dissertation focus on the production and comprehension of attraction. Firstly, Research Question (1) focuses on the prosodic realization of agreement attraction sentences. Experiment 1 was designed to answer this question through an elicitation task. A small dataset of attraction and non-attraction sentences was collected and analyzed to determine common intonational patterns from this elicitation task.

Secondly, when thinking about the comprehension of spoken, as opposed to written, agreement attraction sentences, the question of how differing phrasing patterns influence acceptability arises. Research questions (2) and (3) address this open question through a speeded acceptability rating task of auditory stimuli. One possibility is that particular phrasings could delimit the domain of interpretation and guide the process toward one interpretation over the other. Finally, we also would like to know how phrasing inhibits or

speeds up the processing of agreement, and ultimately the timing of these decisions about acceptability.

These questions are addressed in light of the current literature on the nature of agreement attraction and what it tells us about sentence processing and the mind. Two influential camps in the study of agreement attraction have differing views on how attraction occurs: representational and retrieval accounts. The representational family of accounts explains attraction through the errorful encoding of the source of agreement due to interference of various meaning-level and grammatical influences on subject NP plurality. However, the retrieval account of agreement attraction situates the problem in the agreement process itself, particularly in the competition between nouns for retrieval as the source of agreement. In the following sections, I will discuss both of these accounts in more detail and show how the current study addresses the gaps left in the current literature.

1.2 Agreement Attraction

A modern example agreement attraction is shown in Figure 1.1, an anti-tobacco sign found on UNC-Chapel Hill's campus (Hessick, 2021). The sign reads "Any Use Of Tobacco Products Is Prohibited In Chapel Hill Parks" with the verb "is" crossed out and replaced with "ARE", presumably to agree with "tobacco products" despite the head noun being "use". What this example illustrates is the powerful psychological effect of agreement attraction. The robustness of this effect has made it an excellent test case for studying both the production and the processing of agreement.

Several linguistic factors, including morphology, syntax, semantics, and pragmatics, play an essential role in determining grammatical agreement in production and comprehension. Morpheme selection, syntactic role assignment, message level constraints, and more all interact with one another in an instance in real-time language production, which is why agreement poses such interesting questions about language and the mind. Some of these questions include: are all of these factors equally weighted when determining agreement? Can the strength of one factor modulate another? What conditions lead to more or

Figure 1.1: A screenshot of a tweet by Carissa Byrne Hessick that shows a sign that says "Any Use Of Tobacco Products Is Prohibited In Chapel Hill Parks" with a "no smoking" logo on the sign. The word "is" has been crossed out and written over with "ARE" in an example of agreement attraction of the local noun "tobacco products".

less agreement attraction? These early questions shaped the direction of agreement attraction studies and were the start of the representational and retrieval families of accounts we find today.

Bock and Miller (1991) provides some of the earliest experimental work looking for answers to the questions raised previously. Bock and Miller were interested in how these factors for agreement modulate the production of agreement attraction errors. In a production study, participants listened to the beginning of a sentence like "The key to the cabinet…" and produced an ending to that sentence as fast as possible. The results pointed to a few critical findings that have been replicated in recent studies. Firstly, agreement attraction is sensitive to the grammatical number of the subject head noun. Secondly, the syntactic position of the interfering local noun is significant. The farther the local noun was syntactically from the head noun, the lower the rate of attraction. Lastly, the semantic properties of the head and local nouns are not a substantial factor in attraction.

For subject NPs with mismatched grammatical number, singular head nouns with plural local nouns (e.g., "the key to the cabinet") resulted in many more cases of agreement attraction than when the grammatical number of the nouns was flipped (e.g., "The keys to the cabinet"). Additionally, these subject NPs with singular heads and plural local nouns resulted in a much greater number of attraction errors when the local noun was in a prepositional phrase like "the key *to the cabinets*" than when it was in a relative clause like "the key *that opens the cabinets*". Thus, these grammatical number and syntactic configuration effects were considered two of the most influential factors in the production of agreement attraction.

Other factors of interest, such as memory constraints and the lexical meaning of the head and local nouns, proved less influential. A manipulation of post modifier length (e.g. "The key to the cabinet" vs. "The key to the ornate Victorian cabinet") was conducted to test whether putting more time between the head noun and the verb would increase the likelihood of attraction errors. The reasoning behind this manipulation was that participants might forget the head noun and thus choose the local noun as the source of agreement. This manipulation, however, resulted in no effect on the rate of agreement attraction. Se-

mantic properties of the nouns, namely animacy and concreteness, were also manipulated to test how the head and local nouns' meaning guided attention. No evidence was found supporting animacy or concreteness as substantial factors in the rate of errors. Due to this weak/non-existent evidence and the strength of the morphosyntactic effects, it was claimed that these agreement errors were driven primarily by syntactic relations of the nouns.

However, later work found that the meaning of the subject NP itself is indeed relevant to agreement attraction. One semantic factor that has been found to influence attraction, if only subtly, is the plausibility of a notionally plural distributive reading of an NP (Vigliocco, 1996). Notional plurality refers to the inherent plurality of a noun based on its referent(s) in the world. For example, the word "group" is notionally plural because a "group" refers to a set of individuals. Nonetheless, "group" is morphologically unmarked for number and thus grammatically singular.

In distributive NPs, the head noun is syntactically singular, but notionally could refer to a plurality of that noun. An example of a sentence where a distributive reading is highly plausible is the sentence in Example (4).

(4) *The name on the billboards are potentially visible from the highway.

In this example, the name that appears is presumably on every billboard, such that in reality, there are multiple instances of the name across multiple billboards. Therefore, while "name" is grammatically singular, it can be interpreted as notionally plural. This type of notional plurality is less context-dependent for nouns like "team", "crowd", or "committee", which are inherent made up of a plurality of things. Crucially, the distributive reading and notional plurality of "name" arises from the plurality of the local noun and the pragmatics of the situation described. This distributive reading is nonetheless highly improbable in sentences like 5. While the possibility of a distributive reading of an NP is known to somewhat influence attraction comprehension, this was not explicitly controlled for in the current study.

(5) *The city with the billboards are visible from the highway.

1.2.1 Representation vs. Retrieval Accounts

In light of early work on agreement attraction, two families of accounts emerged as the main frameworks for understanding the causes and mechanisms behind agreement attraction. Representational accounts focus on the encoding of the subject NP and retrieval accounts focus on the selection of a noun as the source of agreement. In this section, I discuss the basis for the two camps and details about their frameworks.

1.2.1.1 *Marking and Morphing*

The *Marking and Morphing* theory of agreement is a representational account that details how the message-level meaning of a subject NP and number information from morphological constituents of said NP contribute to the selection of a singular or plural verb form (Bock et al., 2001). These sources of number information relate to the *marking* of the notional plurality of the subject NP and then the *morphing* of the number representation of the NP based on the presence of unambiguously singular, unambiguously plural, and/or number ambiguous morphemes.

The head of the subject NP is *marked* based on its notional plurality. For example, a noun like "group" is more notionally plural than "member" since the former refers to a plurality of people whereas the latter is singular. *Morphing* refers to the reconciliation of the lexemes/morphemes and their syntactic roles into a representational unit that can then be sent off for phonological encoding. In this process, the plurality of the parts of the structure can influence the plurality of the whole representation. Figure 1.2 shows an illustration of the various levels of representation at play during speech production in the number computation of a sentence.

Concerning agreement attraction sentences, the explanation for the presence of a plural verb after subjects like "the key to the cabinet" is that the number representation of the subject NP was influenced by the presence of plural morphology lower in the structure. During planning for the number for the upcoming verb, the agreement mechanism encounters the subject NP as grammatically plural, and therefore a plural verb like "are"

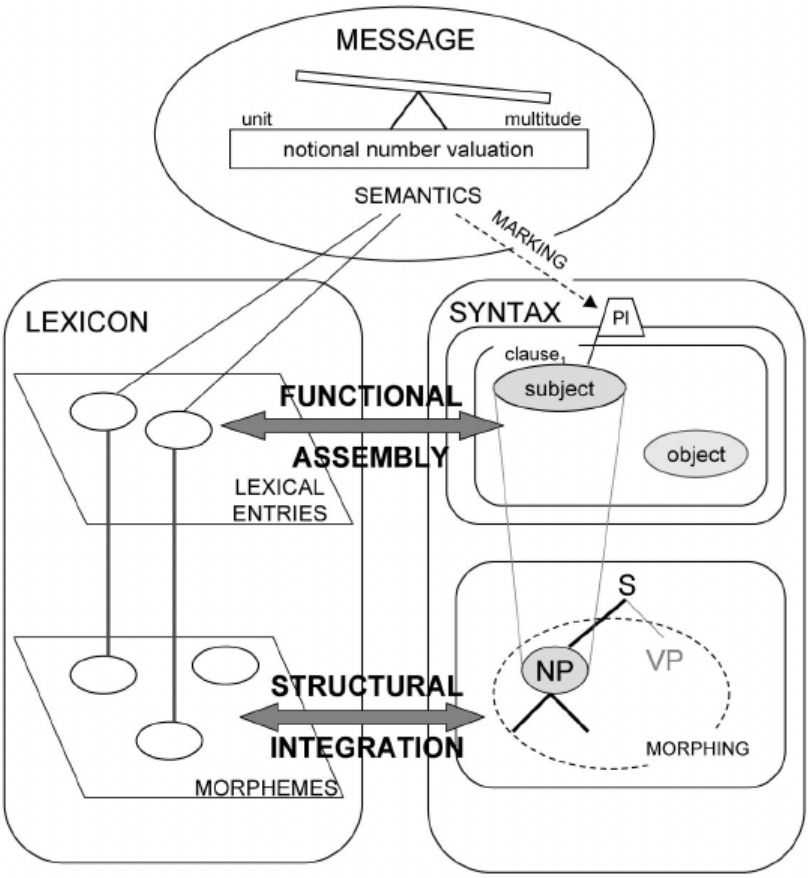


Figure 1.2: Diagram of the components of number marking in production. Taken from Eberhard et al. (2005)

was planned. In this account of attraction, the issue is not that there has been a *functional* error by the agreement mechanism, but rather that the subject NP was encoded incorrectly, which precipitated the plural verb production.

A more detailed computational implementation of this model was done by Eberhard et al. (2005). This model grounds the *Marking and Morphing* theory of agreement in an activation-based cognitive model that relies on corpus study and experimental results as its foundation. In this model, subject NP's plurality, which is a gradient value and not a categorical factor, is activated to a greater or lesser degree depending on the *marking* and *morphing* of the NP. The calculation of the subject NP's number is instantiated computationally using Equation 1.1. This formula takes into consideration notional plurality of the subject (i.e., *S(n)*; ranging from 1 for an unambiguous multitude to 0 for a single unit of some thing to -1 for an unambiguous individual) and lexical number specifications of individual lexemes/morphemes (i.e., *S(m)*; calculated based some set specification of the morpheme's number and a comparison of the frequency of singular and plural forms of the word[1]) in the NP. The contribution of number information from morphemes is weighted to replicate previous findings that structurally lower morphemes are weaker in their influence than those morphemes that are closer to the root node of the subject NP (Bock and Miller, 1991; Franck et al., 2002; Hammerly and Dillon, 2017). The *S(m)* value of a morpheme is weighted by w, where w is smaller the farther the morpheme is from the root node. This weighting is needed to account for the fact that plural morphemes farther from the root node hierarchically produce fewer agreement attraction errors than closer plural morphemes. In Figure 1.3, "cabinets" in the left tree would have a higher *S(m)* value than "windows" in the right tree because of this weighting factor.

Using all of these values, a single value for the root node's plurality, *S(r)*, is calculated. The *S(r)* value can be logistically transformed in order to determine the probability that a verb will be plural when agreeing with that subject NP.

[1]Comparison of the frequency of both forms of the word are used to derive how often the forms contrast with one another. Eberhard et al. (2005) gives the example of "bubble" versus "bubbles" to show a relatively balanced set of forms, compared to "sud" versus "suds", where the singular form is much more infrequent than the plural form.

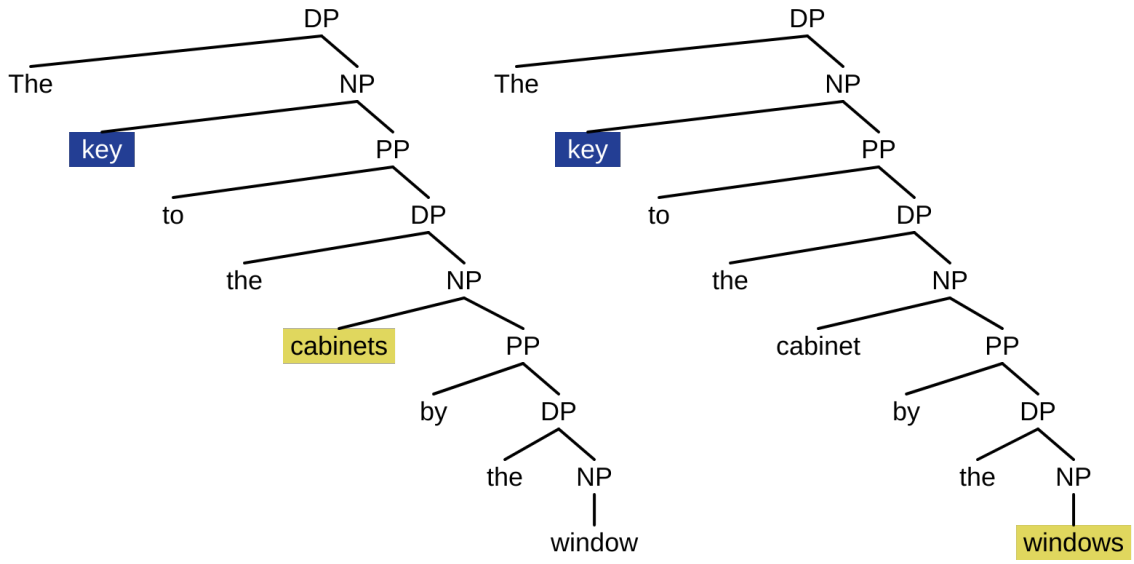


Figure 1.3: Two syntactic trees showing a morphologically plural word closer (left) and farther (right) from the root node of the DP. Highlighted in blue is the head noun of the subject NP and highlighted in yellow are the words with plural morphology.

$$S(r) = S(n) + \sum_j (w_j \times S(m)_j) \tag{1.1}$$

A vital point of this framework is that agreement is probabilistic. Thus, the selection of either a singular or plural verb form is distributional instead of deterministic. Staub (2009) reports results from a study that supports an account of attraction errors being driven by a probabilistic number valuation process like *Marking and Morphing*. In a speeded two-alternative forced-choice (2AFC) task, participants were presented with subject NPs through word-by-word rapid serial visual presentation (RSVP) and asked to follow the subject with a singular or plural verb by pressing one of two keys corresponding to the verbs. The purpose of this task was to test the predictions that Eberhard et al. made by looking at reaction times (RTs) and the verb that was ultimately selected. The results replicated previous findings that plural distractor nouns induced more plural verb responses. However, the RTs for correct and error responses showed no differences in the plural local noun condition compared to the singular local noun condition (where error RTs were longer than correct answers). Only in this probabilistic framework would we expect correct RTs and

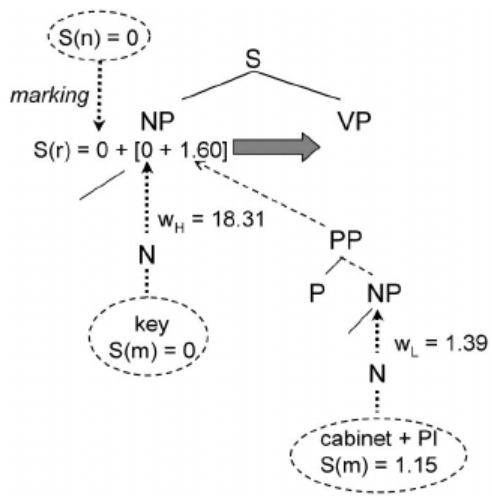


Figure 1.4: Syntactic tree diagram labeled with the values for *S(n), S(m),* and *w* which go into calculation of *S(r)* for "The key to the cabinets". Taken from Eberhard et al. (2005).

error RTs to be roughly equivalent.

One of the benefits of this framework is that the gradient approach to plurality allows for the accommodation of a wide range of results from various constructions. At the same time, the computational implementation provides precise and empirically driven predictions that can be easily tested. Evidence for this flexibility extends into morphologically rich languages like Mexican Spanish and Dominican Spanish as well (Foote and Bock, 2012), where overt singular morphology has been shown to override effects of notional plurality (i.e., distributive readings of head nouns) in cases of agreement attraction.

While *Marking and Morphing* was created to explain the production of agreement, some comprehension studies have found moderate support for the theory as well. In one such study, Patson and Husband (2016) conducted a self-paced reading experiment that sought to probe the final interpretation that readers come to after agreement attraction sentences. The plurality of the local noun and the verb was manipulated in a 2x2 design, with the head noun's plurality held constant as singular. Sentences like those in Examples (6) - (9) were presented by a moving window display.

(6) The key to the cabinet is on the table.

(7) The key to the cabinets is on the table.

(8) *The key to the cabinet are on the table.

(9) *The key to the cabinets are on the table.

After reading each sentence, participants were asked a comprehension question that probed the plurality of the head noun, such as "Was there more than one key?". Answers to the comprehension questions showed that plural local nouns (e.g., "cabinets") did influence the conceptual plurality of singular head nouns (e.g. "key") such that participants often reported the head noun as being plural. This effect was even stronger when the verb of the sentence was also plural.

However, the results of the reading time data contradict predictions made by *Marking and Morphing*. According to *Marking and Morphing*, ungrammatical sentences with plural local nouns like in Example (9) are predicted to be read faster than grammatical sentences with plural local nouns like Example (7). The slowdown in reading grammatical sentences comes from readers building a plural representation for the subject, encountering a singular verb, and therefore judging the sentence to be ungrammatical. In Patson and Husband's results, only the ungrammatical sentences showed an effect of local noun plurality. While readers read the ungrammatical sentences faster with a plural local noun than a singular one, there was no comparable slowdown of grammatical sentences with plural local nouns compared to singular ones. Thus, while the interpretation results align with the idea that a misrepresentation of the number of the subject NP arises from plural local nouns, the self-paced reading results do not.

Despite the richness of this work focusing on errors in the encoding of NPs, other studies utilizing a more general framework than *Marking and Morphing* have looked at the comprehension of agreement and found an alternate explanation. Cue-based retrieval in a content-addressable memory system has been a leading framework in sentence processing for analyzing many dependency resolution issues (Parker et al., 2018). One of the central ideas of cue-based retrieval is that dependency resolution relies on successfully re-accessing linguistic items in memory. In the case of agreement attraction, agreement errors arise from the misretrieval of the source of agreement. The following section will

discuss this framework in detail and how it builds upon our understanding of agreement.

1.2.1.2 *Content-addressable Memory and Cue-based Retrieval accounts*

Another prominent family of accounts of agreement tackles attraction from a domain-general perspective that is premised on the limitations of memory. In particular, these accounts revolve around a content-addressable cue-based retrieval mechanism. A content-addressable system is one in which information is searchable based on a set of cues instead of a serial search mechanism (McElree, 2001). To demonstrate the mechanism of cue-based retrieval, let us consider an analogy. Imagine that you are trying to find "Prosodic Typology: The Phonology of Intonation and Phrasing" edited by Sun-Ah Jun in the research library. One search method would be to start with the first stack of books on the first floor and look at every title, in every stack, on every floor, until you come across the book. This method is akin to the serial search mechanism where information is organized sequentially and only accessible in a given order. The obvious drawback to this method is that it is incredibly time and energy-intensive to execute. A more efficient method would be to input the key characteristics of the book like, "typology", "phonology", and "intonation" and find the exact location. In content-addressable memory, each piece of information has cues or "keywords" that help identify its characteristics. Cue-based retrieval utilizes these cues to directly access the needed information instead of checking superfluous entries. In the library analogy, our keywords would allow us to get the exact floor, stack, and location of the book to go directly there and pick it up without having to reference any unneeded titles. These keywords act as "cues" that point to the information's precise location.

While there are various ways in which content-addressable memory has been implemented, for the current study, I will detail the type proposed by Lewis and Vasishth (2005) for illustrative purposes. This framework operates under the hypothesis that sentence processing is not a part of some unique language faculty but a specialized skill that uses domain-general faculties. This framing of language processing does not consider the need

for specialized linguistic representations but rather the idea that language processing uses a unique set of tools separate from other non-linguistic processes. Instead, they work off the assumption that sentence processing operates under the same cognitive constraints as other forms of processing that are non-linguistic because it utilizes those same faculties. Lewis and Vasishth use the Adaptive Character of Thought-Rational (ACT-R) theory of cognition (Anderson et al., 1997) to formalize their theory of sentence processing into an entirely computationally specified model. ACT-R is a theory of how the mind is organized and operates and a computational model that allows researchers to implement their theories. The output generated from ACT-R models can then be compared to human behavioral data to see if the assumptions made in the model result in human-like behavior. One of the benefits of having a testable computational model is that it first requires researchers to explicitly state and formalize assumptions in their theory, which leads to more predictive and explanatory theories in the end,(Guest and Martin, 2021).

As for the mechanics of the ACT-R model, it is crucial to understand that everything in the model is either considered declarative knowledge or procedural knowledge. Declarative knowledge in ACT-R is made of "chunks", which are items composed of feature-value pairs and a single symbol to represent the chunk. Procedural knowledge is a set of production rules defined by *conditions* that need to be met for some *action* to occur. These rules dictate how chunks can be combined to form novel relationships with one another. Procedural knowledge is used to execute all the processes necessary for skilled sentence processing, such as syntax building and dependency resolution. Figure 1.5 shows an example of the chunks for the sentence "The writer surprised the editors". The features of a chunk are made up of syntactic node information about the chunk, such as its category, number value, specifier, complement, head, and more, depending on the particular node. While not explicitly shown in the diagram itself, the result of applications of production rules is evidenced by the chunks representing the combination of smaller chunks or terminal nodes on the tree. For example, chunk IP3 is represents chunk DP3 in the specifier position and chunk VP7 in the complement position of the sentence. This chunk made of chunks represents the highest syntactic node of the tree.

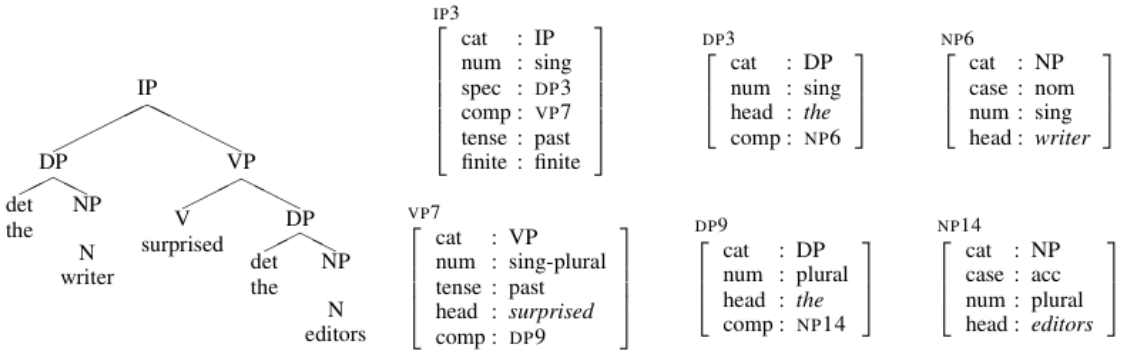


Figure 1.5: Diagram taken from Lewis & Vasishth (2005) showing the declarative chunks and their associated cues for the sentence "The writer surprised the editors."

In order to better understand how ACT-R works, let us walk through a simplified example of how an agreement attraction sentence like Example (1) is parsed. Parsing begins from the left and works rightwards. The first word, "the", is encountered and a declarative knowledge chunk representing that determiner is built. That chunk may contain a key-value pairing such as "cat : det" since its category is "determiner" and "art : definite" because it is a definite article. Next., the parser would encounter "key" and generate a chunk with values like "cat : N" and "num : sing". A procedural rule would then apply to make our N into an NP and combine our other two chunks, the det and NP, into a new chunk, the DP. This third chunk presents the novel relationships between these previous chunks.

Concerning how interference works, the idea is that each chunk has some activation defined by its base level of activation and how many times it has previously been retrieved. When an item is retrieved, its activation level increases.

Support for this family of accounts comes from the lack of predicted symmetry of attraction based on *Marking and Morphing*'s representational account of attraction. Thus, for example, a subject NP like Example (10) has a fixed *S(r)* value which makes it slightly plural, and thus more likely to be followed by a plural verb in production than a subject like Example (11).

(10) The key to the cabinets

(11) The key to the cabinet

In comprehension, this slightly plural representation is the reason why agreement attraction sentences like Example (1) are sometimes rated as acceptable or read as quickly as a grammatical sentence like "The key is on the table". If Example (10)'s representation is plural, then when a plural verb is encountered, the grammatical number requirement for dependency resolution is met, and resolution is successful. In this sentence, you get an illusion of grammaticality. On the other hand, it then follows that when participants are presented with subjects like Example (10) followed by a *singular* verb, some proportion of their responses must be rated unacceptable or exhibit a slowdown due to an illusion of ungrammaticality. If the representation of the subject is grammatically plural, then encountering a singular verb would result in the sentence being judged as ungrammatical.

Wagers et al. (2009) designed a study to test the prediction that *Marking and Morphing* makes and did not find symmetry between grammatical and ungrammatical illusions. Wagers analyzed the online effects of agreement attraction and the more offline acceptability judgments in a self-paced reading task followed by an acceptability judgment task. Judgment data showed that while a plural attractor noun increased the acceptability of ungrammatical sentences with a plural verb, the plural attractor did not decrease the acceptability of grammatical sentences. This "grammaticality asymmetry" was provided as evidence against representational accounts of agreement attraction.

Wagers et al.'s account for agreement attraction was that a content-addressable retrieval process is initiated at the verb. Reading time data showed that attraction effects generally occurred after the verb[2], indicating that the verb itself was crucial in initiating attraction effects. The self-spaced reading results from Patson and Husband (2016) replicated the grammaticality asymmetry observed by Wagers et al.

Thus far, we have encountered conflicting accounts of how agreement attraction arises and at what time it starts influencing processing. On the one hand, we have a repre-

[2]Due to the addition of an adverb before the verb in two of the experiments, effects showed up earlier on the verb itself instead of after. This earlier effect is explained as the adverb signaling to the participant to expect an upcoming verb. Thus the verb relations were processed more quickly.

sentational account of attraction like *Marking and Morphing* which situates attraction in the difficulty of reconciling NPs with various sources of notional, morphological, and grammatical plurality. On the other hand, we have retrieval-based accounts like the one proposed by Wagers et al. (2009) which point to errors in the agreement mechanism's selection of a noun in the subject NP as a source of agreement. While both accounts have corroborated some of the same factors that influence rates of attraction, despite their differing viewpoints on the locus of attraction, neither family of accounts has taken up the influence that prosody might play in attraction. Given that the study of prosody's influence on sentence processing has been underway for about as long as psycholinguistic work on agreement attraction, this gap in the agreement literature is one worth addressing.

1.2.2 Where does Prosody Fit In?

Prosody refers to the suprasegmental elements that accompany and shape the segments they are associated with. The main acoustic correlates of prosody are f0, intensity, and duration, which in turn relate to perceptions of pitch, loudness, and length, respectively. Intonation, those prosodic phenomena that operate above the level of the word, is influential in many domains of sentence processing. In particular, much work has been done considering how the grouping of words into prosodic constituents (i.e., prosodic phrasing) influences parsing, computation of meaning, and ultimately the comprehension of complex syntactic structures.

One of the significant limitations of both the representational and retrieval families of accounts is that it is unclear from either what role intonation plays in the processing of agreement. Work examining intonation's influence has been limited to the study of implicit prosody projected onto a sentence while reading. In a study looking at implicit prosody's effect on agreement attraction, Pratt and Fernández (2019) manipulated the visual presentation of attraction sentences (i.e., word-by-word, phrase-by-phrase, and whole sentence) in an attempt to facilitate or inhibit the projection of an implicit prosodic contour onto the sentence. These sentences, shown in Examples 12 and 13, were broken into phrases after

the head noun and after the relative clause.

(12) The reporter | who called the senators every so often | write(s)...

(13) The reporter | who called the senators that Scott supported | write(s)...

Word-by-word presentation presented the most significant difficulty concerning generating an implicit prosodic contour.

Readers were asked to judge the sentence's grammaticality and then answered a comprehension question. Results showed that agreement processing was facilitated (i.e., attraction as ungrammatical) by segmenting the sentence into phrases. In addition, participants' comprehension of the sentences was affected by presentation. Word-by-word presentation inhibited successful comprehension and phrase-by-phrase facilitated it, as compared to the whole sentence condition. The natural progression of this line of research is to look for evidence of similar effects in comprehension by using auditory materials.

In the following section, I discuss some of the relevant findings concerning what we know about prosodic phrasing and how it helps guide parsing and the construction of a final interpretation. This discussion is particularly relevant to agreement attraction because the (mis)interpretation of the head noun's number is a common phenomenon (Patson and Husband, 2016).

1.3 Prosodic Hierarchy and Syntax

Prosodic hierarchy refers to the grouping of prosodic units into larger structures. Depending on the language, these units can be as small as a mora or as large as an entire utterance. For this study, I will focus on prosodic grouping above the level of the word, defined by intonation in the Autosegmental Metrical (AM) model. In the AM model of English intonational phonology, the intermediate phrase (ip) and the intonational phrase (IP) the two relevant levels (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Ladd, 2008). These phrases are characterized by suprasegmental phonetic cues such as pitch excursions and duration lengthening. Pitch contours are analyzed as a sequence of high (H)

and low (L) tones that mark prominence and the edges of these prosodic domains. Prominent stressed syllables are marked with pitch accents, and the edges of prosodic phrases are marked with phrase accents and boundary tones (e.g., H*, L*, etc.). The intermediate phrase is marked on its right edge by one of three phrase accents (i.e., H-, !H-, and L-) and durational lengthening of the phrase-final syllable. The intonation phrase is marked on its right edge by one of three boundary tones (i.e., H% and L%) and an even greater amount of lengthening than the ip on its phrase-final syllable. Intonational phrases are composed of one or more intermediate phrases, and thus, the right edge of an IP is co-terminus with the right edge of an ip. The reverse is not necessarily true, however.

Just as the syntax of a sentence is parsed online, so too must the prosodic structure of an utterance (Gee and Grosjean, 1983; Selkirk, 1986; Beckman, 1996; Truckenbrodt, 1999; Nespor and Vogel, 2012). This parsing proceeds online and matters for sentence interpretation, even in silent reading (Fodor, 1998; Kreiner and Koriat, 2005; Hirotani et al., 2006; Breen, 2014; Jun and Bishop, 2015). While syntactic structure and prosodic structure are related to one another, they are not isomorphic. Prosodic structure cannot be directly read off given the syntactic structure of a given utterance (Shattuck-Hufnagel and Turk, 1996; Nespor and Vogel, 2007).

Phonological weight is one factor known to influence phrasing aside from the syntax of an utterance(Nespor and Vogel, 2012; Fodor, 1998). For example, two sentences may be identical syntactically but have different phrasing due to the amount of phonological material in one utterance or the other. For example, compare the sentences “Ann’s green farming idea won her the grant” and “Alexandria’s sustainable agriculture proposal won her the grant.” In a review of the English literature on the size of ips, Schafer and Jun (2002) report the size being, on average, five syllables or four content words. Therefore, in the previous sentence, the former utterance could realistically be broken up into two intermediate phrases of roughly equal size, with the subject forming one and the predicate another. In the latter, however, the subject NP has many more syllables than the predicate and is unlikely to form a single phrase by itself. Thus, the sentence is likely to be broken into three phrases “Alexandria’s (%) sustainable agriculture proposal (%) won her the

grant" for example.

These sorts of divergences between syntactic and prosodic structures have led to investigations of the functional role of prosody in sentence processing. Early psycholinguistic work on prosodic structure proposed that phrasing served to delimit processing "chunks" (Gee and Grosjean, 1983). Further work has refined this understanding of prosody's role in processing and connected prosody with domain-general faculties like memory and attention (Frazier et al., 2006; Swets et al., 2007; Simpson, 2017; Bishop, 2020).

1.3.1 Prosodic Phrasing's Role in Processing

The relationship between prosody and syntax is a complex and nuanced one that requires careful consideration in psycholinguistic theory. Our theories of sentence processing must account for concurrent prosodic processing as well (Carlson, 2009; Lee, 2012; do Carmo Lourenço-Gomes, 2017).

Various research programs have undertaken this challenge to connect our understanding of prosodic structure with syntactic structure. Kjelgaard and Speer (1999)'s work tests how phrasing guides early versus late closure in a sentence like "When Roger leaves the house is/it's dark." Kjelgaard and Speer not only found that prosody that aligns with the parse of the sentence facilitates processing, but phrasing that is incongruous with the meaning disrupts parsing. Therefore, this evidence for disruption shows that speakers have clear expectations of prosody based on meaning and project those expectations onto incoming material when listening. Furthermore, this has also been found in a replication study using pupillometry to measure the difficulty of processing disruptive phrasings (Harris and Jun, 2019).

Phrasing's interaction with syntactic structure can even lead to the misparsing of sentences. Frazier et al. (2014) conducted a study that tested how prosody affects participants' interpretations of sentences with embedded clauses like Examples (14) and (15). They hypothesized that the beginning of an IP could be interpreted as the beginning of a stand-alone sentence or 'root' structure, such that embedded clauses could be interpreted

as ‘root‘ clauses. In the case of both example sentences, a substring of the full sentence could be a stand-alone clause, i.e., “her boss is an alien” and “the player tossed a frisbee.” Frazier et al. found that these partial strings of a sentence can indeed be misinterpreted as stand-alone sentences when they appear in their own intonation phrase, separate from any previous material.

(14) Martin says that Louise believes her boss is an alien.

(15) The FBI questioned the congressman mailed a letter.

In a comprehension experiment, participants were played a recording of a sentence like Example (15) (see Figure 1.6) and asked a question probing their interpretation of the sentence they just heard. The location of a prosodic break was manipulated so that it appeared before or after the main verb, thus creating a stand-alone clause in the late break condition and not in the early condition.

In both the early and late break conditions, the rate of stand-alone interpretations of the sentences was above 75%. There was a main effect of break location, though, such that late breaks resulted in even higher rates of misinterpretation. This work tells us that phrasing can facilitate interpretations that are incongruous with the input but that phrasing can deaccentuate portions of input. In particular, I am referring to the material preceding the embedded clause being disregarded. If listeners interpreted the embedded clause as a stand-alone clause, the previous portion of the sentence could not be integrated into the syntactic structure without making an ungrammatical sentence. Therefore, this preceding information must be more difficult to access since it is not in the same IP as the embedded clause.

With regards to semantics and meaning, Schafer (1998) presented a strong case for prosodic boundaries’ role in delimiting the scope of meaning. In a series of comprehension studies, Schafer showed that phrasing affects not only syntactic resolution but also the domain of focus interpretation and semantic/pragmatic interpretations. The latter was tested by manipulating the presence or absence of an IP boundary after a context-sensitive adjective. A NP like “an inexpensive Porsche”, is ambiguous between an independent *in-*

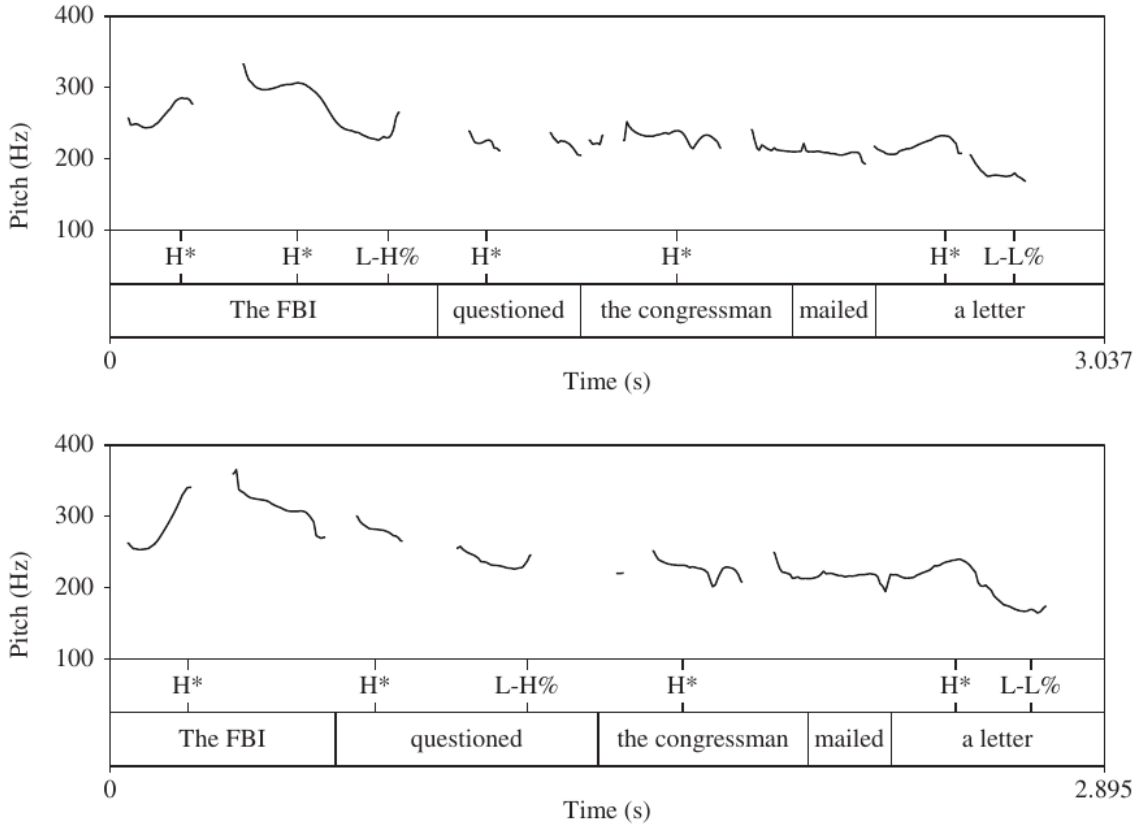


Figure 1.6: Examples pitch tracks from (Frazier et al., 2014) showing the early boundary (top) and late boundary conditions (bottom).

tersective reading of "expensive" or a "head-noun dependent" *subsective* reading. The *intersective* reading is one in which this NP refers to the intersection of all the things that are inexpensive and all things that are a Porsche. Schafer calls this the "independent reading" because the adjective is evaluated with respect to the discourse context of the utterance. In the other reading, "inexpensive" is evaluated with respect to the head-noun "Porsche" such that the NP refers only to the subset of Porsches that are inexpensive. Schafer found that an IP boundary after an adjective led to a greater number of independent readings than when only an ip boundary intervened between the two words. This finding is important because it shows evidence for phrasing's influence on the domain of interpretation of lexical items, which is very much related to higher-level semantic and pragmatic processing. Thus, the IP's influence is not merely limited to determining syntactic groupings but extends into semantic processing by delimiting interpretative domains.

All this work tells us that prosodic phrasing and the IP of English, particularly, delimits syntactic and semantic domains, which can guide interpretation. Given the importance of the interpretation of the subject NP number in agreement attraction, investigation of phrasing's role in affecting that interpretation is a potentially fruitful avenue of research.

1.3.2 Pitch Accents and Processing

In addition to phrasing, pitch accenting and accent type are quite influential in sentence processing as well. In languages that have them, post-lexical pitch accents have been found to guide the resolution of attachment ambiguity, speed up lexical processing, and aid the prediction of upcoming words, among other things (Schafer, 1996; Cutler et al., 1997; Lee and Watson, 2011; Carlson and Tyler, 2018). We also have ample proof from Visual World Paradigm eye-tracking experiments that this accentual information is available to speakers in real-time and that they readily make use of it (Dahan et al., 2002; Ito and Speer, 2006). Moreover, comprehension is aided by the appropriate accenting of new or focal information and the deaccenting of given information, showing that the connection between accenting and representation building is a close one (Bock and Mazzella, 1983).

What ties together many of these studies is that accenting enhances the recall or retrieval of linguistic representations. For example, Fraundorf et al. (2010, 2012) conducted a study that examined the effect of accent type on later recall of events. They were interested in testing the difference between the H* pitch accent and the L+H* pitch accent. An important distinction between these two accents is that the H* is most commonly used to mark new information in a discourse. In contrast, the L+H* is most often used to signal contrastive or corrective focus (Pierrehumbert and Hirschberg, 1990). In the task, participants listened to short discourses, which they later had to recall. The first passage of the discourse (Example (16)) set up a pair of contrasting sets (e.g., Britsh vs. French & Malaysia vs. Indonesia) and the following passage (Example (17)) singled out one member of each set.

(16) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(17) Finally, the (British/French) spotted one of the monkeys in (Malaysia/Indonesia) and planted a radio tag on it.

Fraundorf et al. manipulated the presence accent type on the two words such that either they either both carried the same accent type (i.e., H* and H* or L+H* and L+H*) or they carried different accent types (i.e., H* and L+H* or L+H* and H*). Results indicated that words that carried L+H* were recalled more often than those with H*. Additionally, this facilitation of memory for the words with L+H* persisted for more than a day after the initial task was completed. Thus, the boost to encoding was not short-lived but rather quite robust. Additionally, the effect of the L+H* did not seem to impair memory of the other story detail in the test passage. One explanation for these effects of accenting on memory is that accented words, particularly those in a focused position, gain the benefit of deeper processing and thereby have richer semantic representations (Sanford and Sturt, 2002; Sanford et al., 2009).

1.4 How does it all connect?

Both the agreement and prosodic processing literature overlap in their connection to the encoding and retrieval of linguistic representations. The two primary camps explaining how agreement operates rely on theories of how representations of NPs are derived and how cue-endowed items in memory are retrieved, respectively. Due to the strong influence of phrasing and pitch accenting on representation building and retrieval, our theories of agreement can be developed even further by investigating prosody's particular role in agreement attraction processing.

1.5 Goals and scope of this dissertation

This dissertation aims to test what roles prosodic phrasing and accenting play in guiding the resolution of subject-verb number agreement. The influence of prosody will be evaluated in light of the *Marking and Morphing* theory of attraction and the cue-based retrieval account put forth by Wagers et al. (2009).

To investigate prosody's influence on comprehension, a speeded acceptability judgment task using auditory stimuli was designed. In this experiment, the prosodic phrasing of these agreement attraction sentences was manipulated to see if it affected the perceived grammaticality of these sentences. An essential first step in understanding how phrasing influences agreement attraction comprehension is to know the typical intonational patterns used when speakers produce these sentences themselves. Since earlier, and even more recent, production studies did not analyze or report the intonational patterns of their data, this leaves an open question as to whether attraction sentences are produced with a particular intonational pattern. The elicitation and prosodic analysis of agreement attraction sentences in this dissertation stand to build off of those same studies and help bridge the gap between production and comprehension accounts. In this dissertation, I use the results of the previously mentioned experiments to address the following research questions.

1.6 Research Questions

(1) What patterns of phrasing and accenting are typical for agreement attraction sentences? Does the intonation of agreement attraction sentences differ from non-attraction sentences?

Whether or not agreement attraction sentences differ from non-agreement sentences in their intonational contours and phrasing is an empirical question. The relationship between a word's accentual prominence and its ease of retrieval for the speaker is inversely correlated, telling us something about the representation built by the speaker (Watson et al., 2006; Wagner and Watson, 2010). Therefore, an intonational analysis of speakers' productions could tell us something about the linguistic representation of subjects in agreement attraction sentences. The production experiment in Chapter 2 answers this question through analysis of a small corpus of semi-spontaneous speech containing attraction and non-attraction sentences.

(2) How does phrasing influence the acceptability of agreement attraction sentences?

We have seen that IPs and ips act as "chunking" units that can facilitate the processing of an utterance. Frazier et al. (2014)'s work, in particular, suggests that the processor is sensitive to whether a previously encountered input is in the same IP or not. If the material is no longer in the same phrase as the input currently being processed, this preceding material may be harder to access and utilize. Therefore, whether the main verb and local noun are in the same IP or not could mediate the rate of agreement attraction comprehenders experience. In Chapter 3, the comprehension experiment addresses this question by asking participants to judge auditorily presented sentences as acceptable or unacceptable under time pressure.

(3) How does phrasing affect the speed with which participants determine the source of agreement?

There is much support for a timing difference between the processing of attraction and non-attraction sentences in online measures. We saw from Wagers et al. (2009) that plural local nouns facilitated the processing of ungrammatical sentences with plural verbs, for example. Therefore, differences in the judgments of an attraction sentence are presumably due to some facilitation or inhibition of the agreement process. Prosody is likely to result in a slowdown or speed-up of agreement processes depending on whether it boosts or minimizes the interference of a plural local noun. Chapter 3 presents an analysis of the RTs from participants' acceptability judgments.

1.7 Hypotheses

I propose two primary hypotheses relating to the effect that intonation may have on the processing of agreement attraction. The Chunk-Internal Availability Hypothesis states that prosodic phrases act as processing chunks and mark off the edges of units that should be processed together. The Accentual Processing Boost Hypothesis, on the other hand, focuses primarily on the pitch accenting of the local noun resulting in the deeper encoding of this intervening element. Since the purpose of the production experiment is more exploratory than explanatory, these hypotheses do not necessarily bear on how speakers will choose to phrase and accent agreement attraction sentences. Nevertheless, both of these hypotheses approach intonation from the Autosegmental Metrical framework of intonational phonology.

1.7.1 Hypothesis 1: The Chunk-Internal Availability Hypothesis

The idea behind this hypothesis is that linguistic elements within the same IP are more accessible in memory than elements in a preceding IP. The accessibility of chunk-internal elements facilitates syntactic parsing and meaning construction with those IP-internal elements. Elements in a preceding IP are still accessible, but their retrieval is degraded relative to the retrieval of an element within the same IP. In other words, it is more difficult

to utilize those elements in a different IP than those in the same IP.

1.7.1.1 Linking Hypotheses

Judgment of the sentence's acceptability relates to the encoding of the subject NP. Chunk-Internal elements have a higher resting activation than elements in a preceding chunk and therefore are more easily activated for syntactic and semantic processing. As a result, elements within the same chunk are processed more deeply, resulting in a richer semantic representation (which can be less perturbed by interference from something like a local noun). This is why listeners may arrive at an interpretation that is not faithful to the input they experienced.

Table 1.1 shows a step-by-step schematic of how phrasing influences the representation of the head and local nouns as the parser encounters each critical word. In Step 1, there are no differences in the processing of the two sentences because no prosodic differences have been encountered yet. In Step 2, the plural local noun has been encountered, and this is where we see the first prosodic differences. In the case of the SAME IP condition, the prosodic juncture size at the right edge of "cabinets" is a word boundary. On the other hand, the last syllable "cabinets" in the DIFFERENT IPS condition is lengthened because it is at the right edge of an intonation phrase boundary, and it is carrying a L-H% phrase accent and boundary tone. This IP juncture signals to the processor to wrap up any ongoing syntactic or semantic processes, which results in the representation of the subject NP being encoded more deeply. I propose that this deeper encoding of the subject NP results in the head noun's grammatical singularness exerting a larger influence in the agreement process. Finally, in Step 3, the verb is encountered, which presumably triggers subject-verb dependency resolution. Acting as a sort of baseline, no prosodic interference in the SAME IP condition means that the interference caused by the local noun is not strengthened or weakened by prosodic factors in that condition. However, since the entire subject NP is in its own separate IP in the DIFFERENT IPS condition, resolution of the subject-verb dependency should be more difficult. One way that this difficulty could be

operationalized is in the time-to-decision for listeners' acceptability judgments. As for the accuracy of their judgments, the influence of the head noun's singular number is already enhanced in Step 2 relative to the local noun, meaning that the local noun's interference would be weakened in the DIFFERENT IPS condition relative to the SAME IP CONDITION.

	SAME IP	DIFFERENT IPS
Step 1	The key —	The key —
Step 2	The key to the cabinets —	The key to the cabinets <- L-H%
Step 3	The key to the cabinets are —	The key to the cabinets | are —

Table 1.1: Schematic representation of the agreement process in the phrasing conditions. The purple boxes surrounding the words represent IPs. The em dashes refer to word boundaries, and the size of a word reflects its accessibility or level of encoding. Larger words are more accessible than normal size words, which in turn are more accessible than smaller words.

From the design of these experiments, it is not clear whether differences in response times between conditions can help distinguish between effects of encoding versus effects of retrieval. For example, it could be the case that longer times-to-decision are the result of encoding of the subject NP taking longer or retrieval that takes longer as a result of some interference, or both processes are contributed to RT differences.

1.7.1.2 Predictions

Suppose a plural local noun and plural verb are in the same IP. In that case, they should be easier to process together. Therefore, we expect to see a greater rate of acceptability for those sentences than when the local noun and verb are in separate IPs. If we extend the previously discussed results of self-paced reading to the time-to-decision measure for this study, we would also predict that acceptable responses to ungrammatical attraction sentences would have a similar response time as grammatical sentences. If attraction sentences are read without any disruption to the agreement process, we may also expect to

see the lack of disruption for participants' time-to-decision.

Taking into consideration the two primary agreement attraction theories, let us discuss their contrasting predictions. A representational account of attraction like *Marking and Morphing* would predict that the slight plurality of a subject NP like "The key to the cabinets" would cause ungrammatical attraction sentences (i.e., plural verb) to be rated as acceptable and grammatical attraction sentences (i.e., singular verb) to be rated as unacceptable. This kind of symmetry in attraction is derived from the fact that the subject NP's plurality value is fixed and the same no matter the verb number. Therefore, the subject NP will be interpreted as plural the same proportion of times in both conditions. However, as we have seen, this prediction does not carry with it much empirical support. However, a cue-based retrieval account like that put forth by Wagers et al. would predict that attraction only takes place in ungrammatical sentences. The RT predictions from the comprehension literature are focused on online reading measures, but this task does not give us moment-by-moment information about the processing of the sentence. Therefore it is hard to say if slowdowns at the plural local noun will be apparent in the response times made at the end of the sentence. Nonetheless, it is still possible that the interference of the plural local noun could cause a processing slowdown and inflate response times.

1.7.2 Hypothesis 2: The Accentual Processing Boost Hypothesis

A noun that carries a prosodically prominent pitch accent may be processed more deeply and, therefore, more easily identified as a source of agreement. In particular, then the local noun precedes a phrase edge. In this case, it carries a nuclear pitch accent, which is the most phonologically, if not acoustically, prominent word in an ip. Thus, while both nouns are pitch accented, the local noun gets a nuclear pitch accent, and the head noun receives a prenuclear pitch accent. In English, the nuclear pitch accent is the last pitch accent of an intermediate phrase and is phonologically more prominent than any preceding (i.e., prenuclear) pitch accents (Beckman, 1996; Ladd, 2008) Because the local noun is phonologically more prominent, this may result in deeper processing of the local

noun, thus making it easier to retrieve as the source of agreement.

1.7.2.1 Linking Hypotheses

The judgment of a sentence's acceptability relates to the quality of the encoding of the local noun and its subsequent retrieval. For example, suppose an interfering word, such as a plural local noun, carries the last pitch accent of an intermediate phrase (i.e., NPA). In that case, the listener processes that word more deeply, which translates to a greater degree of interference in the agreement process. This is because it is more easily reactivated later when the verb is encountered, and the agreement process begins. Thus, an interfering word that carries an NPA is more likely to influence the agreement process than words that do not carry an NPA.

Table 1.2 shows a schematic similar to Table 1.1, but focused on the mechanics of the Accentual Prominence Boosting Hypothesis. Again, in Step 1 there is no difference between conditions in the head noun. In Step 2, the primary prosodic feature at play is the local noun carrying a Nuclear Pitch Accent in the DIFFERENT IPS conditions. This NPA results in deeper processing of the local noun. Additionally, as was true in Table 1.1, the whole subject NP is processed more deeply as a result of being in its own IP. The difference, in this case, is that the processing advantage of the NPA'd word results in the local noun exerting more of an influence in the agreement process than the head noun. Therefore, whether or not the whole subject NP is processed more deeply, the local noun still gets processed even deeper than the head noun, which only carries a prenuclear pitch accent.

1.7.2.2 Predictions

A plural local noun carrying a nuclear pitch accent may be more likely to cause interference in the agreement process. This predicts that when the noun and verb are phrased separately, we expect to see a greater rate of acceptability for those sentences than when the local noun and verb are in the same IP.

	SAME IP	DIFFERENT IPS
Step 1	|The key —|	|The key —|
Step 2	|The key to the cabinets —|	|The key to the cabinets| <- L-H%
Step 3	|The key to the cabinets are —|	|The key to the cabinets| |are —|

Table 1.2: Schematic representation of the agreement process in the phrasing conditions. The purple boxes surrounding the words represent IPs. The em dashes refer to word boundaries, and the size of a word reflects its accessibility or level of encoding. Larger words are more accessible than normal size words, which in turn are more accessible than smaller words.

From the point of view of *Marking and Morphing*, if the local noun is encoded more deeply as a result of the carrying a nuclear pitch accent, perhaps this counteracts the weights (*w*) of the lexical item's contribution (*S(m)*) to the overall plurality of the subject NP (**S(m)**). This attenuation of the weighting would mean that we could expect a greater acceptability rate when the local noun carries a nuclear pitch accent (i.e., the noun and verb are phrased in separate IPs) compared to when it does not. As for RTs predictions, since the process of checking agreement is not disrupted, acceptable responses to agreement attraction sentences should match the RTs of the grammatical singular local noun singular verb condition. Again, it is also predicted that when the local noun is plural, and the verb is singular, the proportion of unacceptable answers match the proportion of acceptable answers to the ungrammatical attraction sentences.

The Accentual Processing Boost Hypothesis fits nicely into a content-addressable memory system because of the inclusion of activation of chunks. Prosodic prominence may directly affect the base activation of a chunk, thus making it faster to be retrieved. In that case, acceptable response times for attraction sentences would be faster than unacceptable response times for ungrammatical attraction sentences.

1.8 Overview of Thesis

The remainder of this dissertation is structured as follows: In Chapter 2, I present a production experiment designed to elicit attraction errors in a replication of Bock and Miller (1991). Next, in Chapter 3 I present a speeded acceptability judgment experiment in which participants listen to and judge attraction and non-attraction sentences. The phrasing of these sentences is such that either the local noun and verb are in the same IP, or they are in different IPs. Finally, Chapter 4 concludes with a summary of the results of the two experiments, a discussion of where our models of agreement attraction and processing need to be revised, and future directions for this work.

CHAPTER 2

Experiment 1: Agreement Attraction Production

In this chapter, I present details about the methods of a production experiment designed to elicit agreement attraction sentences. The purpose of this experiment was to answer RQ1: "What patterns of phrasing and accenting are typical for agreement attraction sentences? Does the intonation of agreement attraction sentences differ from non-attraction sentences?". Previous production studies of agreement attraction have not reported the type of phrasing and accenting used in speakers' productions, so this task was designed to fill that gap in the literature through an exploratory analysis.

The elicitation task was designed as a replication of Bock and Miller (1991)'s task with slight adjustments so that it could be conducted online[1] at the participants' own pace. The recordings of agreement attraction sentences were transcribed orthographically and then subjected to MAE_ToBI labeling of the pitch accents and boundary tones relevant to RQ1. Analysis of the rate of attraction and intonation patterns follows with a brief discussion of the implications of the results.

2.1 Methods

In the elicitation task, participants were played subject NPs like "the key to the cabinets" and asked to think of an appropriate continuation to make a complete sentence. They then produced their entire utterance out loud, and it was recorded.

The experiment was designed using Labvanced (Finger et al., 2017), a cloud-based ex-

[1]Due to the COVID-19 pandemic, both experiments were conducted online to minimize the risk of exposure to the virus for the researcher and participants.

perimental design and hosting website. Participants were recruited using Prolific (Prolific, 2021), an easy-to-use experimental participant recruitment website. Prolific allows for pre-screening of potential participants based on researcher-defined criteria such that only those who meet the study's demographic inclusion criteria could participate. One of the other benefits of using Prolific as opposed to the other popular recruitment tool, Amazon Mechanical Turk (MTurk), is that Prolific users tend to produce better quality data (e.g., fewer instances of participants not following instructions or quitting before completing the experiment) than MTurk (Palan and Schitter, 2018)[2].

Prolific's pre-screening process works through participants first filling out an extensive questionnaire about their background when they first create their account for the website. From that point on, the information from this questionnaire is available to researchers depending on the inclusion requirements of their experiment. The requirements for participation in this experiment were as follows: participants had to be at least 18 years old, a self-identified native English speaker, and have no hearing impairments. Information about other languages spoken, gender identity, race, and ethnicity was collected as well. Additionally, geographic restrictions on the US state of birth and current state of residence were used to recruit speakers of various US regional dialects. Participants were born in one of the following states: Ohio, Kentucky, West Virginia, or Tennessee. The South Midlands and Appalachian regional dialects of English are known to be associated with these states (Labov et al., 2008; Greene, 2010; Hazen, 2014).

Speakers were recruited from these areas to ensure that the sample population included speakers with non-standardized subject-verb agreement. In particular, default singular agreement was the grammatical feature of interest. Default singular agreement is the leveling of the distinction between singular and plural verb forms such that speakers use the

[2]Part of this increase in quality is because Prolific sets a minimum compensation rate of $6.50/hr for participants, which means participants can spend their time on a single experiment as opposed to trying to complete many more experiments for much less compensation. For example, participants in Experiment 1 received USD 3.50 for approximately 20 minutes of their time (USD 10.50/hr), which is "good" according to Prolific's payment rating system (USD 11 or higher is "great!" and USD 9 or lower is "low"). On the other hand, many experiments of the same length on MTurk offer $5/hr or less. Therefore, the prerogative of MTurk workers is to complete tasks as fast as possible to reach a sufficiently high rate of pay per hour of work, which is something Prolific participants are less incentivized to do.

"singular" form regardless of the subject NP's grammatical or morphological number (see Examples (18) and (19)) (Greene, 2010).

(18) Tina's parents is strict about curfew.

(19) We was thinking of going to the mall today.

Previous studies of agreement attraction have excluded speakers with non-standardized agreement systems, but in this study, they were recruited to test for a difference in sensitivity to agreement attraction. In addition, a dialect survey was conducted at the end of the experiment, which elicited acceptability judgments of sentences with default singular. Unfortunately, however, due to a technical error, the results of the dialect survey are uninterpretable.

2.1.1 Stimuli Design

The items were comprised of 48 critical and 48 filler sentences for a 2x2x2 design. This full 2x2x2 was necessary for testing the effect of local noun number, verb number, and phrasing in the comprehension study. However, the production study stimuli only consisted of the subject NP (e.g. "The key to the cabinets"), making for a simpler 2x2 design. The design of the complete 2x2x2 design is given here for clarity. The description of the truncation process in order to prepare stimuli for the 2x2 production task is described later in the methods in more detail.

The critical sentences consisted of a subject NP containing a PP-modifier (e.g. "The consultant for the firm") and some predicate. A majority of the subject NPs used in the design came from previous studies: 32 came directly from Staub (2009), 8 came from Wagers et al. (2009), and 8 were original creations made by the researcher and their RAs. The predicate of each critical sentence was created also original. The researcher and RAs applied two constraints in creating the predicates to control for particular confounds and differences across sentences. Firstly, the predicate had to plausibly apply to both of the nouns in the subject NP to not bias the final interpretation towards only the head noun or

only the local noun. Secondly, the continuation of the sentence after the verb needed to be roughly equal in length and number of words across items.

The critical stimuli were manipulated by local noun plurality, verb plurality, and phrasing of the subject NP. The head noun was always singular. The local noun plurality either matched the head noun or did not (Conditions SS vs. SP). The verb's number was either singular or plural, with only the singular form being grammatical (Conditions GRAMMATICAL vs. UNGRAMMATICAL). Finally, the phrasing of the sentences varied according to whether or not the verb was in the same intonation phrase as the subject NP or a different intonation phrase (Conditions SAME IP vs. DIFFERENT IPS). Examples of these manipulations can be found in Tables 3.3 and 3.4. Each critical sentence was produced such that there was one version for each of the eight condition cells.

Filler sentences differed in that the subject NPs included conjunctions (e.g. "The boots and sneakers"), proper nouns (e.g. "Jane"), genitive constructions (e.g. "The celebrity's neighbors"), nouns with quantifiers (e.g. "All of the sinks"), and simple unmodified nouns (e.g. "The farmer").

Stimuli were designed and recorded as complete sentences and later truncated for this production experiment, a simpler 2x2 design of local noun number and phrasing. The sentences were recorded this way so that they could later be used for the comprehension study, presented in Chapter 3. The truncated stimuli consisted of just the subject NP of the sentence (i.e. "The key to the cabinets") and no predicate.

2.1.1.1 Recording of stimuli

A ToBI-trained linguist read all critical and filler sentences aloud with a specific accenting and phrasing pattern for prosodic consistency across trials (see Figure 2.1). To manipulate phrasing, the speaker was asked to include an intonation phrase (IP) break after the local noun in the recordings for the DIFFERENT IPS condition and avoid placing an IP break after the local noun in the SAME IP condition. Accounting for the phrasing and agreement attraction manipulations resulted in 384 recordings for the critical trials (48

	Nouns MATCH in # **Local N = SG**	Nouns MISMATCH in # **Local N = PL**
GRAMMATICAL	The key to the cabinet is on the table	The key to the cabinets is on the table
UNGRAMMATICAL	The key to the cabinet are on the table	The key to the cabinets are on the table

Table 2.1: 2x2 table showing how the sentences varied across local noun number and the verb number conditions. Red boxes with a drop-shadow above indicate singular word forms and blue boxes with a drop-shadow below indicate plural word forms. The head noun and verb are in square boxes to indicate that the head noun is the the expected source of agreement for the verb. The intervening local noun is in a box with rounded edges to indicate is it not the expected source of agreement.

SAME IP	DIFFERENT IPS
The key to the cabinets are on the table	The key to the cabinets | are on the table

Table 2.2: Schematic representation of the phrasing conditions using the agreement attraction condition (MISMATCH X UNGRAMMATICAL) as an example. The purple boxes surrounding the words represent IPs.

items x 2 levels of phrasing x 2 levels of local noun # x 2 levels of verb #) and 48 filler recordings. The silences after the IP break were spliced out to minimize the inevitable duration differences between prosodic conditions. Nonetheless, due to phrase-final lengthening, local nouns at the right edge of IPs were longer than IP-medial ones. In all conditions, the head noun and the local noun of the subject NP were pitch accented with either a H* or L+H* accent[3]. When there was an intonation phrase break after the local noun, the phrase accent and boundary tone were L-H% to signal that the speaker was not finished speaking. The phrasing of the fillers was also varied, such that half of the items had a prosodic break after the subject NP and half did not.

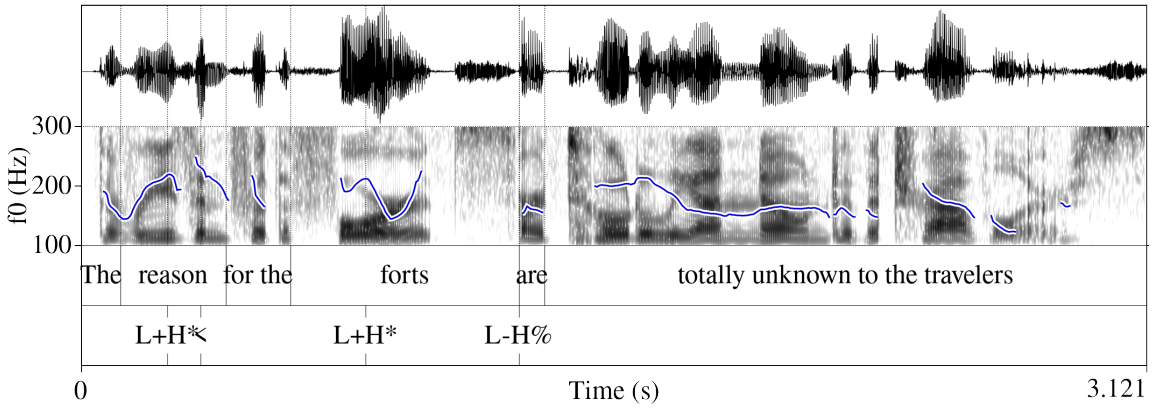
Two recording sessions were needed to fully balance the design of the experiment due to a change in the manipulation of interest. One of these sessions took place before COVID-19 pandemic restrictions and the other while restrictions were in place. The later session was recorded in a different environment with different equipment than the initial recording session.

The first recording session took place in a sound-attenuated booth using a head-mounted microphone. The items were recorded directly into Audacity and saved in a lossless Wav format. The second session during the COVID-19 pandemic was conducted over Zoom. The speaker recorded herself using an iPhone, a laptop running Audacity, and a Zoom H4n Pro Handy Recorder as a fail-safe in case of technical errors. The highest quality lossless recordings from the Zoom recorder were used for the stimuli in this study. After recordings were made, the speaker's speech rate was deemed too fast, so all files from both recording sessions were elongated by 10% to lower the speech rate. In addition, the amplitude of each file was scaled to an average of 65dBs.

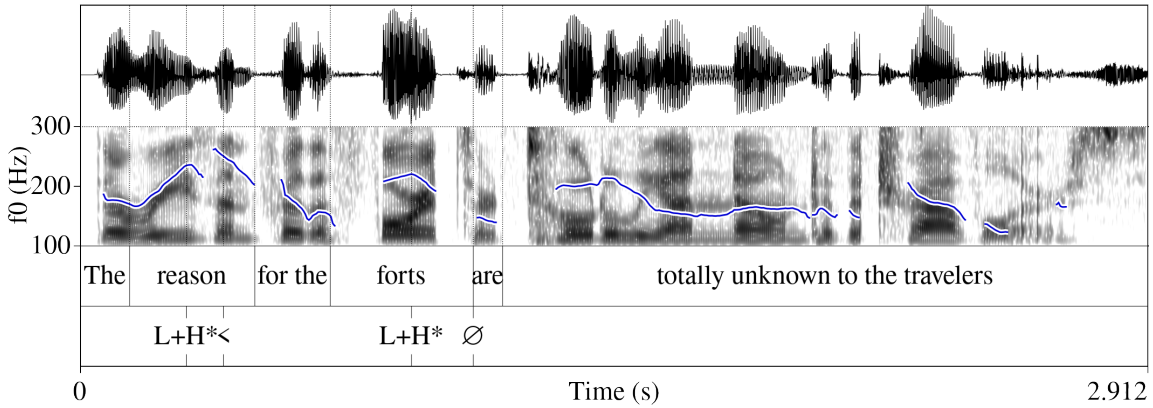
2.1.1.2 Production Experiment Stimuli Truncation

For the production experiment, the stimuli recordings were truncated immediately after the local noun. Since the verb was not present, this resulted in a 2x2 design (shown

[3]The local noun was sometimes downstepped to a !H* and L+!H*



(a) Waveform and spectrogram of an item where the local noun verb are in different IPs. Pitch track is shown in blue.



(b) Waveform and spectrogram of an item where the local noun verb are in different IPs. Pitch track is shown in blue.

Figure 2.1: Two examples of stimuli showing the difference between the SAME IP condition (2.1a) and the DIFFERENT IPS condition (2.1b)

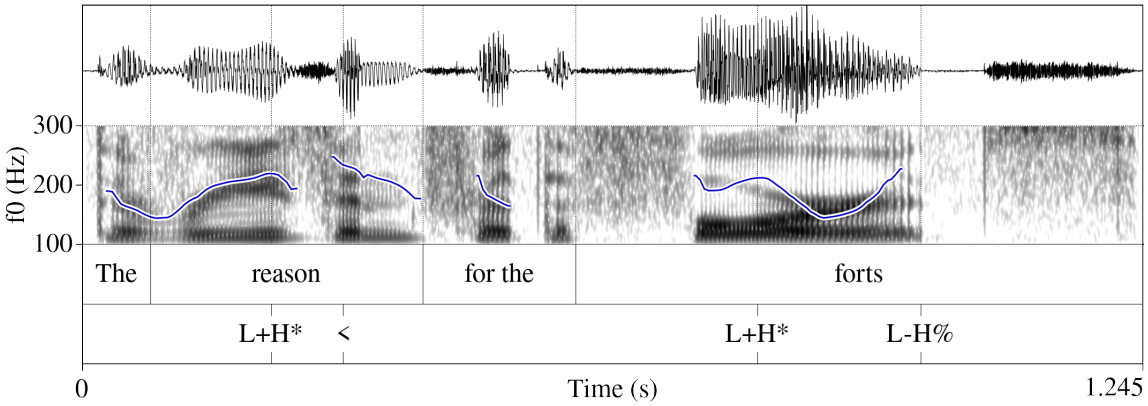
in Table 2.3) of the phrasing manipulation by the local noun plurality. Pitch tracks and spectrograms of the phrasing difference can be found in Figure 2.2.

The main difference between the two phrasing conditions is the durational difference of the last syllables of the local noun. This durational difference is due to phrase-final lengthening of the last syllable in the DIFFERENT IPS condition.

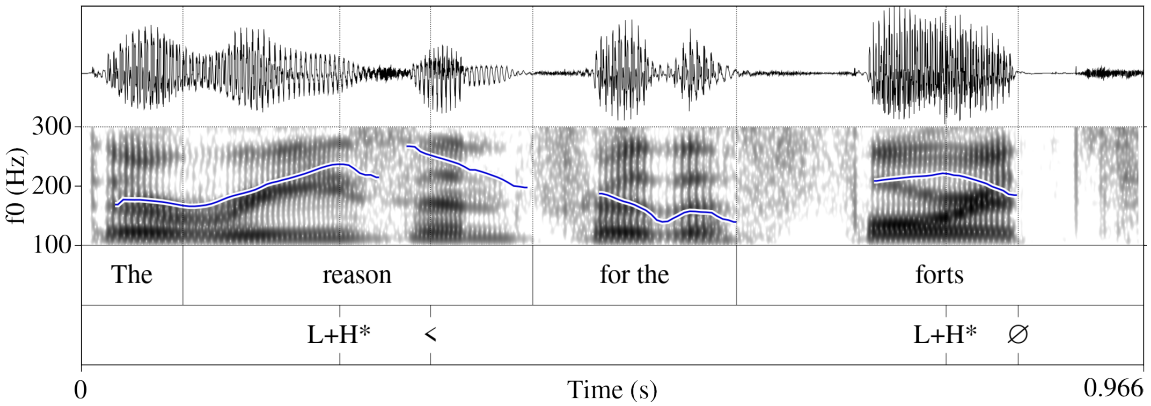
	Nouns MATCH in #	**Nouns MISMATCH in #**
Ends w/ word boundary juncture	The key to the cabinet—	The key to the cabinets—
Ends w/ IP juncture	The key to the cabinet$_{\text{L-H\%}}$	The key to the cabinets$_{\text{L-H\%}}$

Table 2.3: Example of stimuli differences

Filler items were truncated in the same way as the critical items such that participants always hear a complete NP in any given trial.



(a) Waveform and spectrogram of a truncated item where the local noun is preceding a regular word boundary. Pitch track is shown in blue.



(b) Waveform and spectrogram of a truncated item where the local noun is at the end of an IP. Pitch track is shown in blue.

Figure 2.2: Two examples of production stimuli showing the durational and f0 between the "same IP" condition (2.2b) and the "different IP" condition (2.2a).

2.1.2 Procedure

Once participants consented to participating in the experiment, the task was described to them. They were told that they would hear the beginning of sentence fragments and that their task was to think of a way to complete the fragment and form a complete sentence. They were told to come up with a completion to the fragment that was relatively short and sensible. For example, they would hear something like "The apartment with the leaks" and might complete it with "is not on the market anymore," "are on the other side of town," or "should have been fixed before the realtor's showing."

After reading the instructions, participants were prompted to test whether their microphone was working correctly or not via a test screen. Next, three practice trials followed to familiarize them with the task. During the task, participants were shown a play button that let them hear the sentence fragment, which they could replay an additional two times if needed. Once they had thought of a completion for the fragment, they pressed a button to continue. They then would advance to a waiting screen that showed a 3-second countdown and asked them to practice their sentence in their head. Finally, the recording screen appeared and immediately began recording. When the participant was finished saying their complete sentence, they pressed the stop button and continued to the subsequent trial.

2.1.3 Data Processing and Transcription

Recordings were collected for 48 critical and 48 filler trials from a total of 42 participants. Two participants were excluded for not completing the entire experiment, six participants for technical issues with the recordings, and one for failing to produce complete sentences with a verb. Thus, the data analyzed here are from 33 participants producing a total of 1584 critical trial recordings. Five recordings were excluded for not being a complete sentence, leaving 1579 recordings.

A team of research assistants had the task of generating orthographic transcriptions of the participants' recordings and putting those transcriptions into a spreadsheet. The spreadsheet contained columns for marking whether agreement attraction was present in

the transcription or not and any notes the transcriber had on the production itself.

Once the orthographic transcriptions were complete, an R script was used to convert them into text files for the Montreal Forced Aligner to use (McAuliffe et al., 2017). This aligner was used to find the edges of phones and words and mark them in a Praat TextGrid file. After the creation of the TextGrid files, the author used MAE_ToBI guidelines for intonational transcription of the files (Beckman and Hirschberg, 1994; Beckman and Ayers, 1997). The accenting of both nouns and prosodic junctures that occurred anywhere between the head noun and the verb were labeled.

2.2 Results

Participants produced a total of 43 instances of agreement attraction, as shown by the highlighted row in Table 2.4 (this table also shows all counts of subject NP and verb combinations attested). 605 recordings out of a total of 1473 (i.e., 41%) contained a singular head noun and a plural local noun, which is the configuration that has the potential to license agreement attraction. Of that subset of sentences where agreement attraction was possible, 397 (65.4%) of them contained an unambiguously singular verb, 165 (27.5%) contained a verb that was ambiguous in number, and 43 (7.1%) contained an unambiguously plural verb. Bock et al. found rates of attraction around 5-10% in their experiments, showing that our results align with theirs (Bock and Miller, 1991).

Another result that stands out, as seen in Table 2.4, is the number of cases in which the head noun was plural even though it was always singular in the critical stimuli. A consequence of (semi-)spontaneous production tasks is that sometimes participants do unexpected things that may or may not be of theoretical significance to the question at hand. A simple explanation for the changes in head noun number could be due to the participant paying insufficient attention to the original sentence fragment. Thus they just made an error due to a lack of attention. Since these recordings were conducted online, there is no way to know the environment in which the experiment was done. The rate of this head noun number changes (34/605, 5.6%) is similar to that of the number of agreement

Head noun	Local noun	Verb	Counts	Example
sg	sg	sg	564	"the actor in the film was good looking"
sg	pl	sg	397	"the slogan on the posters was very catchy"
sg	sg	amb	268	"the painting of the mountain looked realistic"
sg	pl	amb	165	"the door behind the curtains stayed locked"
sg	pl	pl	43	"the confession of the criminals were coerced"
pl	pl	pl	13	"the labels on the bottles were peeling off"
pl	pl	amb	9	"the leaders of the gangs got arrested"
pl	pl	sg	7	"the apartments with the leaks needs to be fixed"
pl	sg	pl	4	"the letters for the lawyer were delivered"
pl	sg	amb	3	"the bosses of the firm said the company was bankrupt"

Table 2.4: Counts of head noun, local noun, and verb number usage. The row highlighted in yellow represents the agreement attraction cases.

attraction cases (47/605, 7.1%). It is possible that the failure to faithfully reproduce the subject NP points to the type of representation that the participant built upon hearing the sentence fragment. In other words, the participant may have encoded the subject NP as conceptually plural, resulting in a plural head noun when it came time to produce their own sentence. I reiterate that we cannot know if this was indeed what was happening, but it is undoubtedly a phenomenon worth investigating further in the future.

2.2.1 Intonational analysis

The intonational analysis here is limited to sentences with singular head nouns and plural local nouns because they are the ones where agreement attraction is possible. For reference, since word boundaries are not usually associated with a tone of any sort, "0" was used in the labeling to indicate the absence of any tone. Figure (2.3) shows that most of the break sizes were larger than a word boundary (i.e., the ToBI break index label "1)", showing that the local noun and verb were more often phrased separately. Focusing on the usage of breaks larger than a word (i.e., ToBI break index labels 3 through 4), sentences with and without attraction seem to have similar rates of breaks larger than a word (i.e.,

68% and 75% respectively). This majority use of junctures larger than a word means that roughly 3/4 of the time, the subject NP was in its own intermediate or intonation phrase.

The tones associated with these larger junctures also seem to be relatively the same in attraction and non-attraction sentences. Figure (2.4) As one would expect, continuation plateaus (e.g., H-,!H-,H-L%,!H-L%) and rises (e.g., L-H%) vastly outnumber terminal low tones (e.g., L-, L-L%) since speakers still had more to say after the subject NP in order to complete their sentences.

As for the pitch accent types used on the head noun and local noun (see Figures 2.5a and 2.5b respectively), there was a comparable degree of similarity. The head noun in attraction and non-attraction productions was marked with either a L+H* or H* with a preference for H* in the attraction productions. As for the local nouns, accent type usage was relatively the same. When looking at the combinations of accent types on the head noun and local noun, we see the often there was a L+H* pitch accent on both nouns(see Figure (2.6)).

The combination of the last pitch accent in the phrase and boundary tone information was investigated to explore any connection between the accent type on the local noun and the rate of attraction. Recall Hypothesis 2: Phrase-Final Accent Boosts Interference, which posits a relationship between the local noun's nuclear/pre-nuclear accentual distinction and the rate of attraction. In this hypothesis, when the local noun carries a nuclear pitch accent, it is more highly activated in memory. It thus results in a higher degree of interference in the agreement process. Looking at the rate of nuclear pitch accenting illuminates whether the production data support this hypothesis. Figure 2.7 shows the most frequent to least frequent combinations. The prevalence of junctures larger than a word after the local noun and the local noun being accented show that it often receives a nuclear pitch accent.

When looking at the data as a whole, the two most frequent combinations are downstepped accents (i.e., L+!H* and !H*) preceding an intermediate phrase juncture. Separating out just the attraction sentences, the top 3 tunes are [L+!H* 0] (n=5), [L+!H* L-H%]

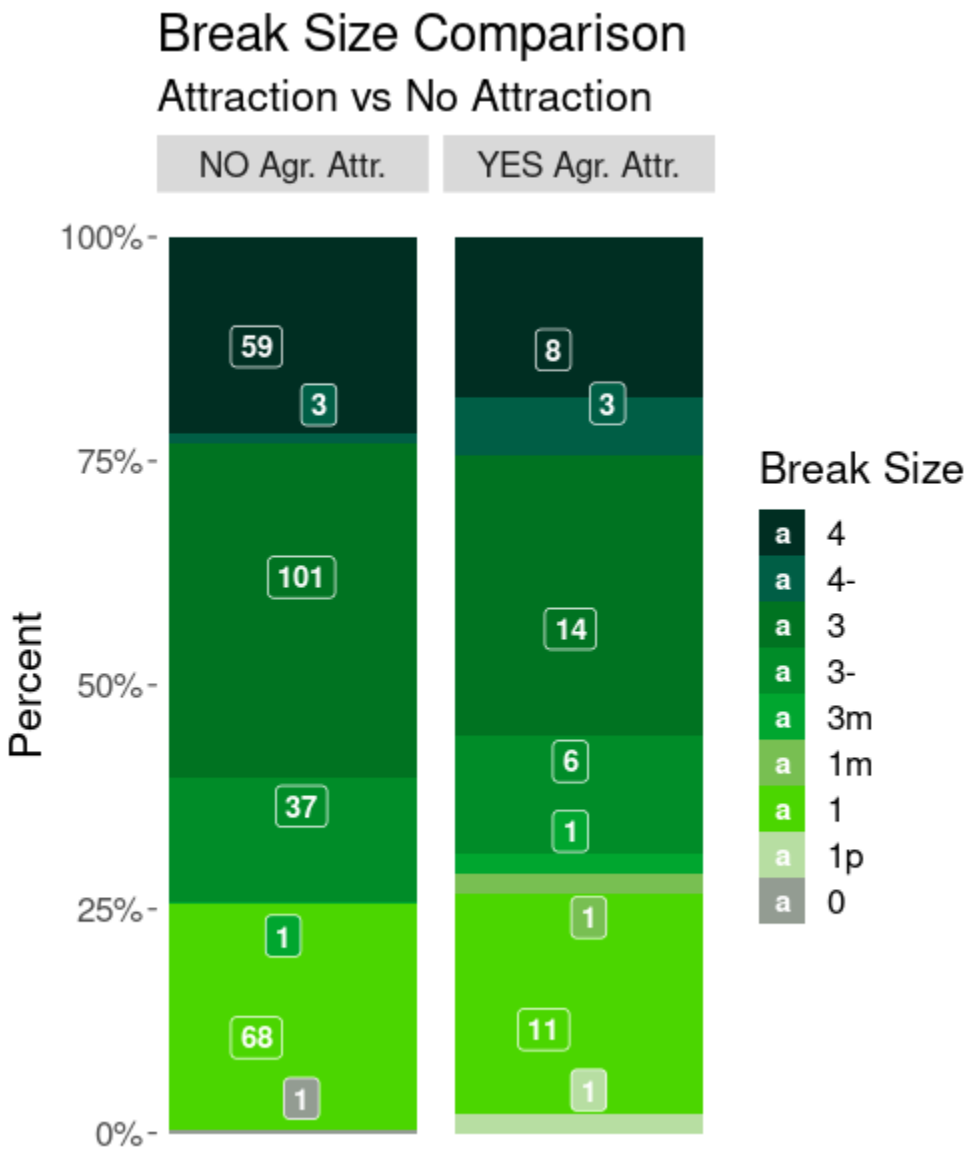


Figure 2.3: Comparison of the break sizes after the local noun. The number in the middle of each bar is the raw counts of that break size.

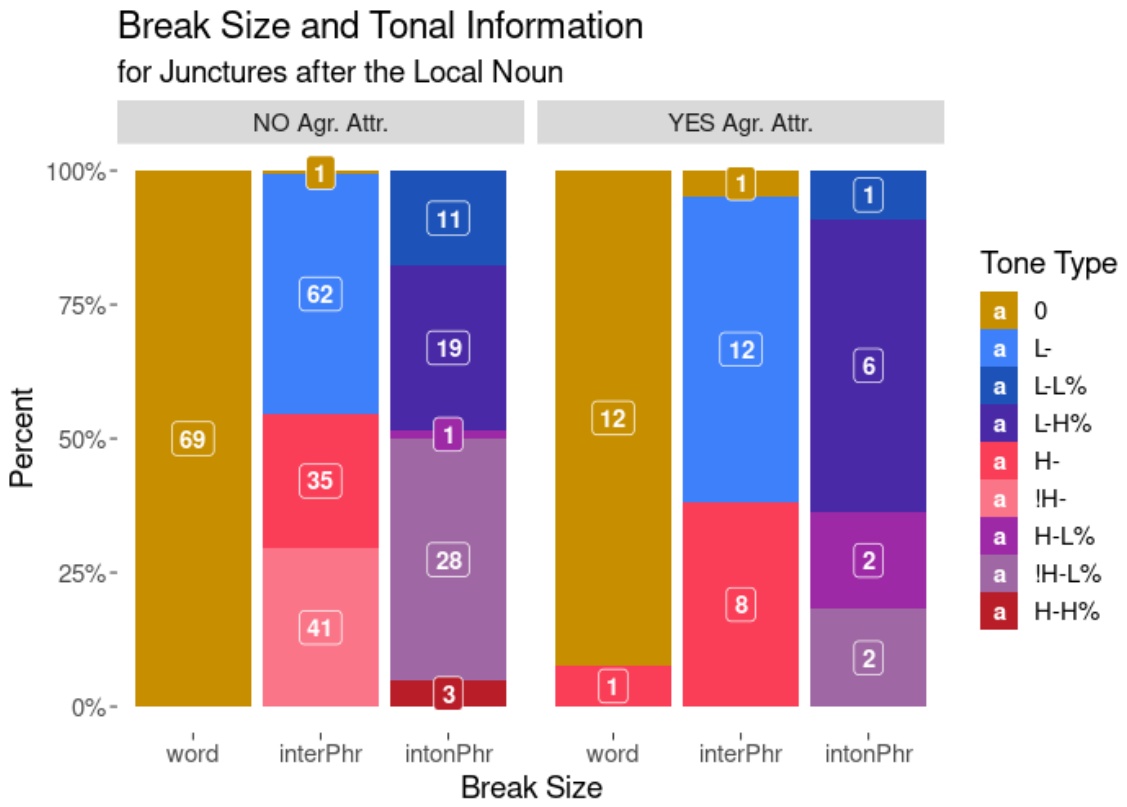


Figure 2.4: Comparison of the break size and tonal markings after the local noun. The number in the middle of each bar is the raw counts of that tonal type for that break size.

(n=4), and [L+H* L-] (n=3).

2.3 Discussion

2.3.1 Similar rate of attraction as Bock and Miller (1991)

Despite some of the methodological differences between this experiment and Bock and Miller (1991)'s experiment, a similar rate of attraction was found in both. Bock's design and instructions to participants emphasized the production of a response as quickly as possible. This kind of time pressure can help prevent speakers from monitoring their speech too closely and then revising errors during their planning of the sentence (Baars and Mackay, 1978). However, in the experiment presented here, participants were given unlimited time to think of their completions and were encouraged to practice their sentences silently before recording. Nonetheless, participants still produced attraction at a rate of about 7.5% of all possible instances. In terms of representational and retrieval models of attraction, while both can explain this robust effect of attraction, the *Marking and Morphing* model can account for variability in attraction through probabilistic means. This probabilistic mechanism explains variation in agreement without relying on the agreement process itself as a locus of some mistake. Therefore, in a representational account, since the agreement process is unimpeded, there are no predicted differences between attraction and non-attraction sentences concerning the difficulty of determining or checking agreement. Suppose the representation of the subject NP is slightly more plural due to the plural morphology on the local noun. In that case, a certain percentage of the time (based on S(r), or representation plurality of the NP), a plural verb is a grammatical choice for production or comprehension.

In the family of accounts based on cue-based retrieval, added time pressure to a task would predict more interference during retrieval due to the strain on cognitive resources. This view is based upon the idea of faulty and errorful retrieval in the agreement process. A decrease in accuracy results in an increase in time to do the task should give partici-

pants a better chance to self-monitor production/comprehension and thus lead to even less attraction. However, we see that this is not the case. Therefore, one point that still needs to be worked out in cue-based retrieval accounts of agreement attraction is how errorful productions are not revised in cases where speakers have the time to do so.

2.3.2 Implications of Explicit Prosody on Reading Studies' Results

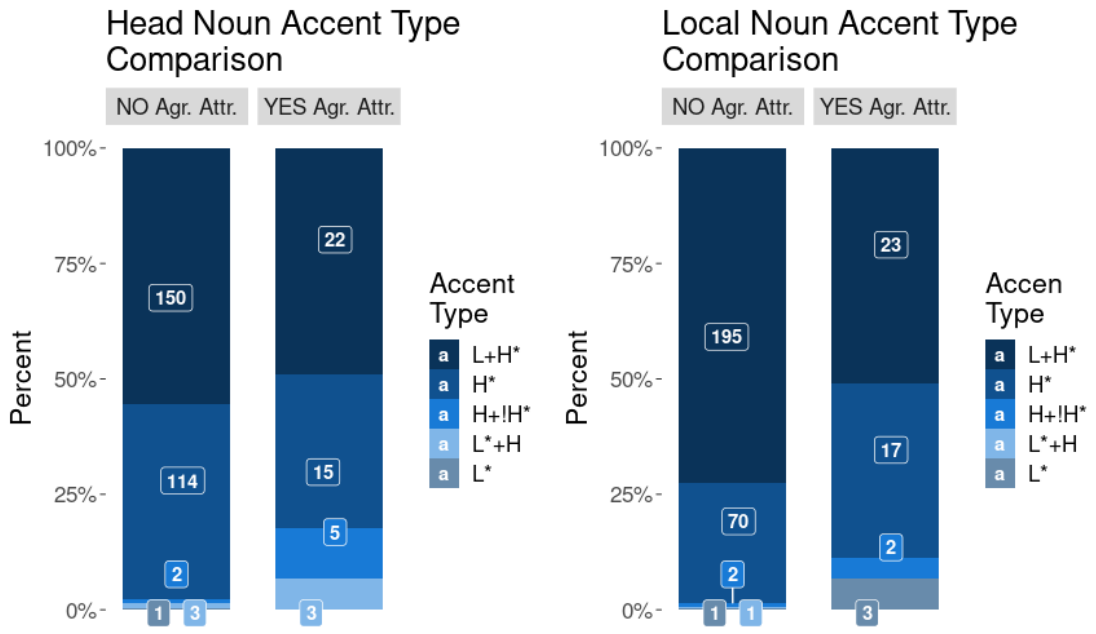
One of the main takeaways for how these PP-modified head noun subjects are produced intonationally is that the subject NP is often phrased in its own IP. However, it is important to note that the stimuli participants heard were the subject NP and nothing else. When thinking about everyday conversational speech, isolated incomplete sentences can be felicitously uttered, but they are necessarily in their own IP, even if they are only one word. Therefore, the prevalence of phrasing the subject NP separately might be an artifact of the fact that the fragment they heard could be interpreted as its own IP.

We know that there is a strong connection between the implicit prosody projected onto a sentence we read and the kind of explicit prosody the sentence tends to be produced with (Bishop, 2020). This could mean for the reading of agreement attraction sentences that there is a preference for a different prosodic phrase for the subject NP when projecting an implicit contour onto a sentence. This claim is undoubtedly substantiated by Pratt and Fernández (2019) in their work on the reading of attraction sentences. Pratt et al. found inhibitory effects of word-by-word presentation on comprehension and facilitation of agreement compared to phrase-by-phrase presentation. Therefore, these results must be taken into consideration when examining the results of studies that present attraction sentences through word-by-word RSVP, like Staub (2009). Word-by-word presentation may have hindered participants' processing of the subject NP due to the difficulty of projecting an implicit contour one word at a time. Thus, this difficulty resulted in higher rates of attraction errors and possibly longer RTs for deciding on the verb to use.

With such few agreement attraction sentences overall, it is difficult to make strong claims about what differences, if any, there are between the intonation of attraction and

non-attraction sentences. There does not seem to be a striking difference in the accenting or phrasing patterns between the types of sentences from the data we have. A more in-depth analysis would likely need to include a larger corpus of attraction sentences to make substantial claims. Ideally, this would be a corpus of spontaneous productions of agreement attraction.

The next chapter delves deeper into the connection between phrasing and agreement attraction by testing participant's judgments of attraction sentences when phrasing varies.



(a) Head noun accent types

(b) Local noun accent types

Figure 2.5: Comparisons of accent types on head and local nouns in attraction and non-attraction productions.

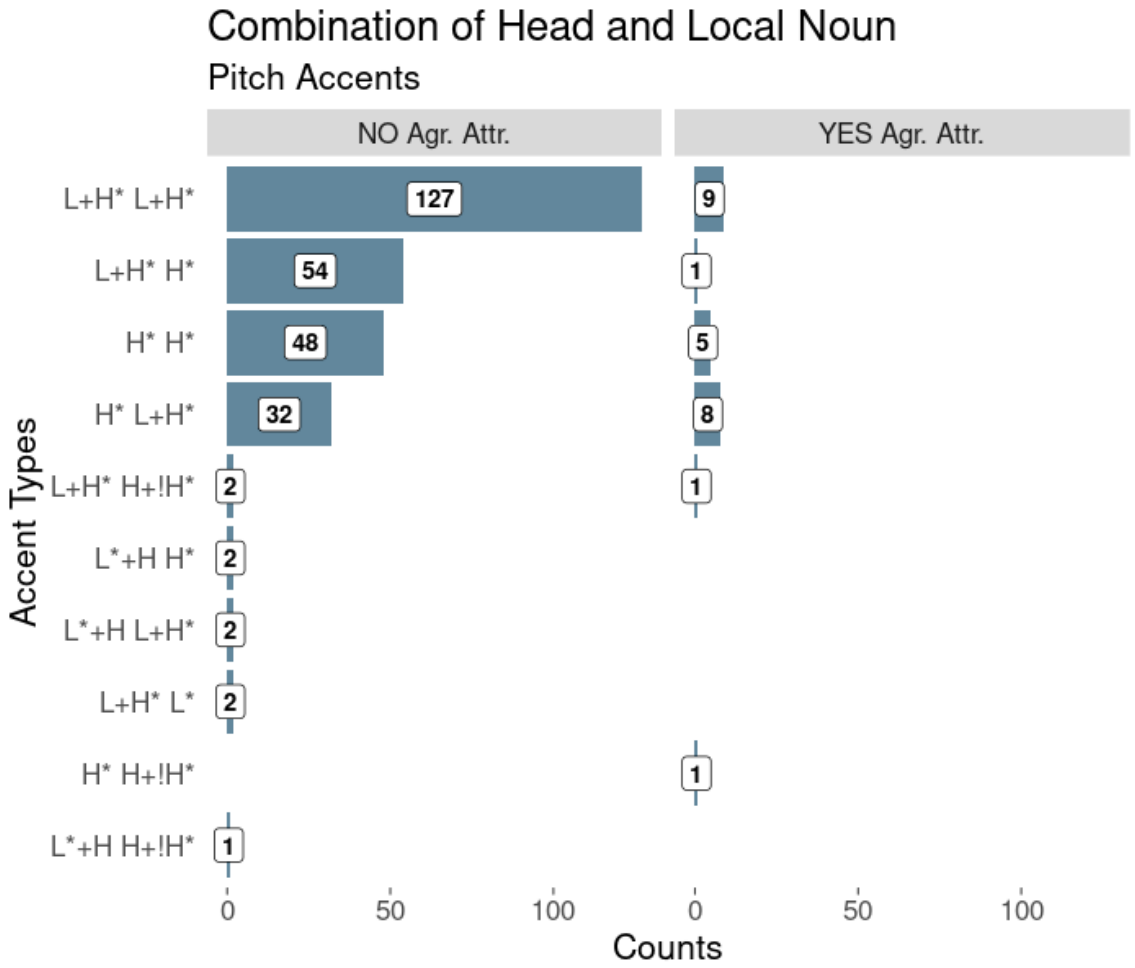


Figure 2.6: Counts of the different combinations of accent types that appeared on the head noun and local noun.

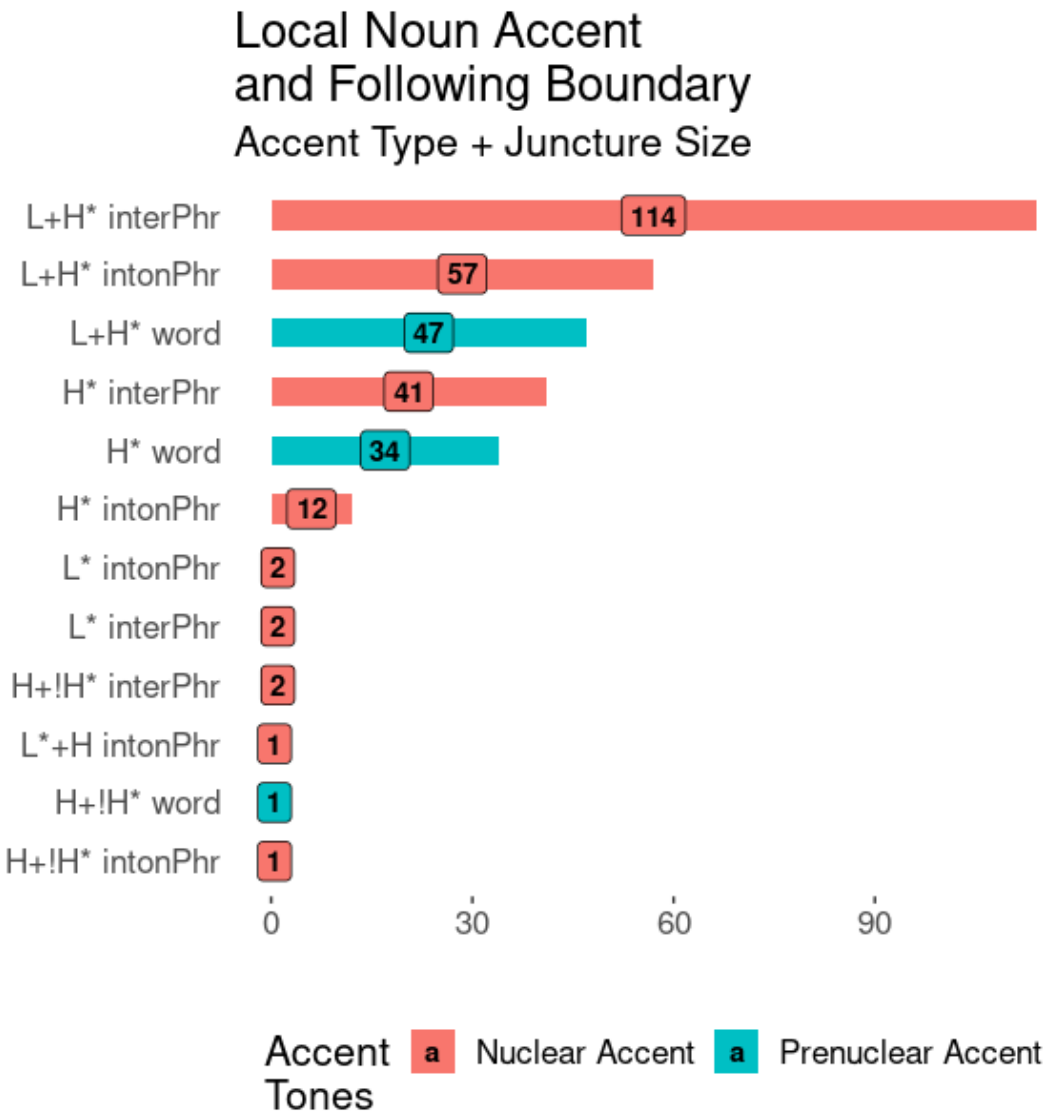


Figure 2.7: Counts of the accent and boundary size combinations used on the plural local noun and singular/plural verb. For clarity, the counts for nuclear pitch accents are in red and prenuclear accents are in blue.

CHAPTER 3

Experiment 2: Speeded Acceptability

In this chapter, I present details about the comprehension experiment. First, I present a brief reminder of the experimental hypotheses and predictions. Then I describe the methods used. Finally, I present the results of the experiment along with a brief discussion.

3.1 Review of Hypotheses

The two main hypotheses I put forth in this study revolve around phrasing and pitch accenting, respectively. Hypothesis 1 builds off of work by Schafer (1998) and others on the role of the intonation phrase in delimiting domains of interpretation. Hypothesis 2 focuses on the role of the nuclear pitch accent in retrieval from memory.

H1 - The Chunk-Internal Availability Hypothesis Linguistic elements within the same IP are more accessible from memory than elements in a preceding IP. Phrasing elements together facilitates syntactic parsing and meaning construction with those IP-internal elements.

H2 - Accentual Processing Boost Hypothesis A word that carries the last pitch accent in a phrase (i.e., nuclear pitch accent) is more accessible from long-term memory than other accented or unaccented words in the phrase.

Agreement Attraction by Phrasing	**H1 - Interpr. Domains**	**H2 - Accent Interference**
The key to the cabinets | are on the table	low acceptability	high acceptability
The key to the cabinets are on the table	high acceptability	low acceptability

Table 3.1: Predictions about the effect of phrasing on the acceptability of agreement attraction sentences for both prosodic hypotheses.

3.2 Methods

The comprehension study was designed as a two-alternative forced-choice (2AFC) task. Participants listened to complete sentences and were asked to judge whether the sentence was "acceptable" or "unacceptable" by pressing the corresponding key on their keyboard. Participants were told that if they did not respond within 3 seconds of the offset of the sound file, the software would automatically advance to the subsequent trial. The purpose of the time pressure was to discourage participants from overthinking prescriptive rules of English agreement and instead encourage them to rely on their intuitions as native speakers.

3.2.1 Stimuli Design

The stimuli design is the same as previously mentioned in Chapter 2 but will be repeated here for clarity. The items were comprised of 48 critical and 48 filler sentences for a 2x2x2 design, manipulating local noun number, verb number, and phrasing. The critical sentences consisted of a subject NP containing a PP-modifier (e.g. "The consultant for the firm") and some predicate. A majority of the subject NPs used in the design came from previous studies: 32 came directly from Staub (2009), 8 came from Wagers et al. (2009), and 8 were created from scratch by the researcher and their RAs. The predicate of each critical sentence was created from scratch. The researcher and RAs applied two constraints in creating the predicates to control for particular confounds and differences across sentences. Firstly, the predicate had to plausibly apply to both of the nouns in the

H1 - Interpr. Domains

"Acceptable" RT	SAME IP	*Comparison*	DIFF IPS	
Fastest	The key to the cabinet is ...	is the same as	The key to the cabinet	is on the table
	The key to the cabinets is ...	(acc. to MM) is slower than	The key to the cabinets	is on the table
		(acc. to CBR) is the same as		
Slowest	The key to the cabinets are ...	is faster than	The key to the cabinets	are on the table

H2 - Accent Interference

"Acceptable" RT	SAME IP	*Comparison*	DIFF IPS	
Fastest	The key to the cabinet is ...	is the same as	The key to the cabinet	is on the table
	The key to the cabinets is ...	(acc. to MM) is faster than	The key to the cabinets	is on the table
		(acc. to CBR) is the same as		
Slowest	The key to the cabinets are ...	is slower than	The key to the cabinets	are on the table

Table 3.2: Predictions

subject NP to not bias the final interpretation towards only the head noun or only the local noun. Secondly, the continuation of the sentence after the verb needed to be roughly equal in length and number of words across items.

The critical stimuli were manipulated by local noun plurality, verb plurality, and phrasing of the subject NP. The head noun was always singular. The local noun plurality either matched the head noun or did not (Conditions SS vs. SP). The verb's number was either singular or plural, with only the singular form being grammatical (Conditions GRAMMATICAL vs. UNGRAMMATICAL). Finally, the phrasing of the sentences varied according to whether or not the verb was in the same intonation phrase as the subject NP or a different intonation phrase (Conditions SAME IP vs. DIFFERENT IPS). Examples of these manipulations can be found in Tables 3.3 and 3.4. Each critical sentence was produced such that there was one version for each of the eight condition cells.

Filler sentences differed in that the subject NPs included conjunctions (e.g. "The boots and sneakers"), proper nouns (e.g. "Jane"), genitive constructions (e.g. "The celebrity's neighbors"), nouns with quantifiers (e.g. "All of the sinks"), and simple unmodified nouns (e.g. "The farmer").

3.2.1.1 Filler items

Filler items were created to mask the experimental manipulations and elicit a range of acceptable and unacceptable responses that did not relate to subject-verb agreement. Sentences included in the fillers were grammatical, ungrammatical, semantically anomalous, and marginally grammatical. This wide array of sentence types was presented to participants for two main reasons. Firstly, the variety of sentence acceptability helped to keep participants' attention. Too many clearly grammatical or ungrammatical sentences could be too easy of a task, leading the participant's attention to wander. Therefore, these more questionable or marginal cases were included to keep participants engaged in the task since those sentences require more attention to be fully understood. The second reason for the variety of fillers was that participants had a more challenging time figuring

out which single grammatical feature was of interest in the study. After reviewing the participants' debriefing questionnaire, it was confirmed that very few (i.e., 10) of the 110 participants correctly identified the manipulation. In addition, participants were asked what they thought the experiment was manipulating in the questionnaire, and only ten mentioned noticing that subject-verb agreement stood out to them.

3.2.1.2 Analysis of Critical Item Stimuli

One of the main questions of this experiment is how phrasing influences the processing time of agreement attraction sentences. This experiment was designed to answer this question by measuring how long it takes a participant to make an acceptability judgment. In measuring the RT of an offline judgment, the timing and duration of stimuli are essential factors to consider since imbalances across items and conditions could confound results. Therefore, a durational analysis of the local nouns across conditions was conducted. Phrasal edges in English tend to be lengthened, making this analysis necessary since phrasing is one of the primary manipulations (Klatt, 1975, 1976; E.Turk and Shattuck-Hufnagel, 2007; Byrd and Saltzman, 2003). This phrase-final lengthening occurs at the right edge of intermediate and intonation phrases and can be seen in Figure 3.1. Therefore, in the DIFFERENT IPS condition, when the phrase ended after the local noun, any material after that would be much longer than the SAME IP condition with no phrase break afterward. The plural morpheme /s/ of the local noun in the DIFFERENT IPS condition (see Figure 3.1) is more than three times as long as the duration of the /s/ in the SAME IPS condition. The additional lengthening of the local noun, all else being equal, gives the listener more time to process the subject NP which is a confound when analyzing the time-to-decision of participants' judgments. This durational analysis aims to quantify the amount of lengthening present in the stimuli across the prosody conditions.

The duration of the local noun was marked on an interval tier in Praat, and then a script was used to extract the duration values. These measurements were modeled as a shifted log-normal distribution in brms using a Bayesian mixed model (Bürkner, 2017).

	Nouns MATCH in # **Local N = SG**	Nouns MISMATCH in # **Local N = PL**
GRAMMATICAL	The key to the cabinet is on the table	The key to the cabinets is on the table
UNGRAMMATICAL	The key to the cabinet are on the table	The key to the cabinets are on the table

Table 3.3: 2x2 table showing how the sentences varied across local noun number and the verb number conditions. Red boxes with a drop-shadow above indicate singular word forms and blue boxes with a drop-shadow below indicate plural word forms. The head noun and verb are in square boxes to indicate that the head noun is the the expected source of agreement for the verb. The intervening local noun is in a box with rounded edges to indicate is it not the expected source of agreement.

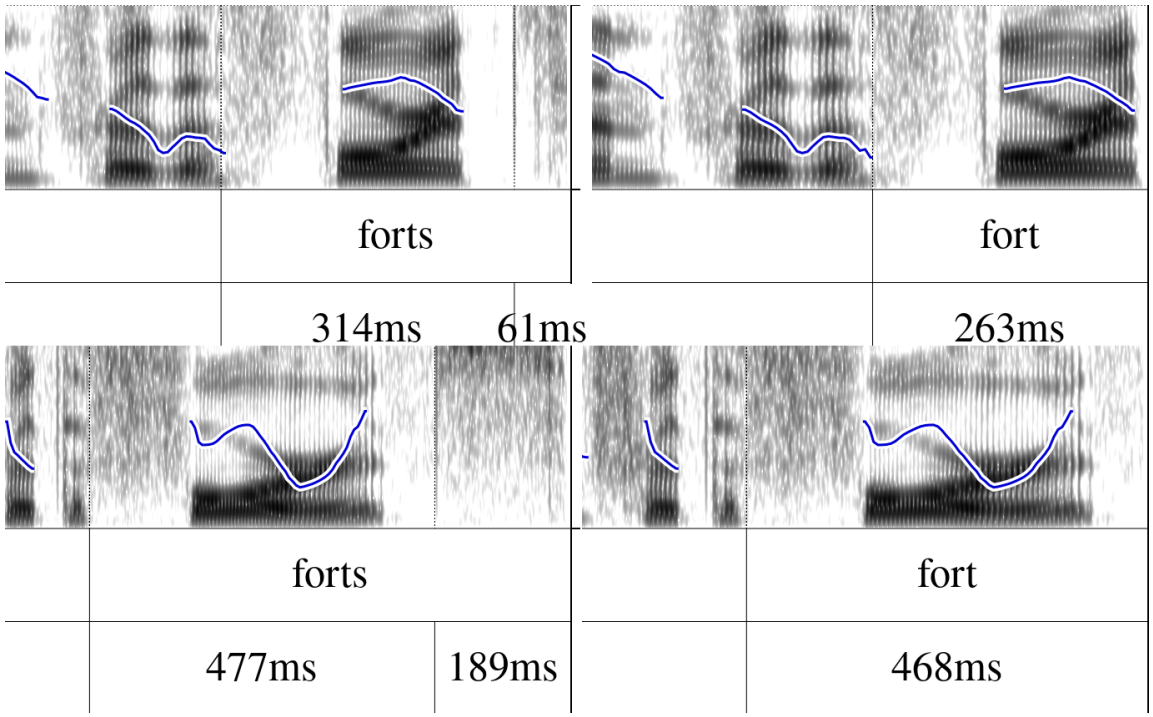


Figure 3.1: The spectrograms show the final syllable of the subject NP "The reason for the fort(s)". The top row shows stimuli from the SAME IP condition and the bottom row shows the DIFFERENT IPS condition. The left column shows the plural forms and the right column shows the singular forms.

SAME IP	DIFFERENT IPS
The key to the cabinets are on the table	The key to the cabinets | are on the table

Table 3.4: Schematic representation of the phrasing conditions using the agreement attraction condition (SP X UNGRAMMATICAL) as an example. The purple boxes surrounding the words represent IPs.

The data in Figure 3.2 show the duration of the local noun as a function of the local noun number and the break condition the item was in.

Bayesian methods were used to analyze the results of duration analysis. The benefit of using Bayesian statistical methods instead of Frequentist methods is that the credibility of the model estimates is explicitly calculated, giving more information about the presence, size, and direction of an effect. In addition, explicit calculations of the probability of an effect avoid issues that arise from the arbitrary dividing line between "significant/non-significant" effects. Bayesian methods also provide more easily interpretable results, particularly when it comes to the Credible Intervals (CrI) that surround model estimates. These intervals contrast with Confidence Intervals (CI), which are often erroneously interpreted in how CrIs are meant to be interpreted (i.e., there is some probability that the parameter's "true" value is within the interval). We used R and the brms packages to conduct the analyses (R Core Team, 2021; Bürkner, 2017). Figure 3.3 shows the model's parameter estimates with 89% credible intervals. The exact estimates and credible intervals can be found in Table 3.6.

A few apparent effects come out of the local noun duration model. Two of the most obvious effects are that local nouns with a plural morpheme are longer than those without one ($\beta = 0.23$, CrI = [0.353, 0.473]) and local nouns before IP breaks are also longer than those that are IP medial ($\beta = 0.276$, CrI = [0.397, 0.516]) .

What is evident from the data in Figure 3.2 is that the lengthening that adding a plural morpheme contributes is even greater when that plural local noun is at the end of an IP, due to final-lengthening (See how much longer the sibilant noise at the end of "laboratories" is in Figure 2.1b compared to Figure 2.1a). This additional lengthening means that listeners

Filler Item Examples

1. The specialty bookstore had an aisle full of books about magic.
2. Sarah's mom and dad have matching tennis outfits for the tournament.
3. The forecast predicts that the weather will get worse next week.
4. #The Browns won the last game that they lost.
5. #The train will leave the station early yesterday.
6. ?The bride loved the cake that her uncle baked and decorated it.
7. *The machine helps the workers broke down the recycled plastic.
8. *The beakers of chemicals will being analyzed by next week.
9. *Everyone should check out these three easy life hack.

Table 3.5: Examples of the filler items used. 1-3 are grammatical, 4-5 are semantically anomalous, 6 is marginally grammatical, and 7-8 are ungrammatical by Mainstream US English standards.

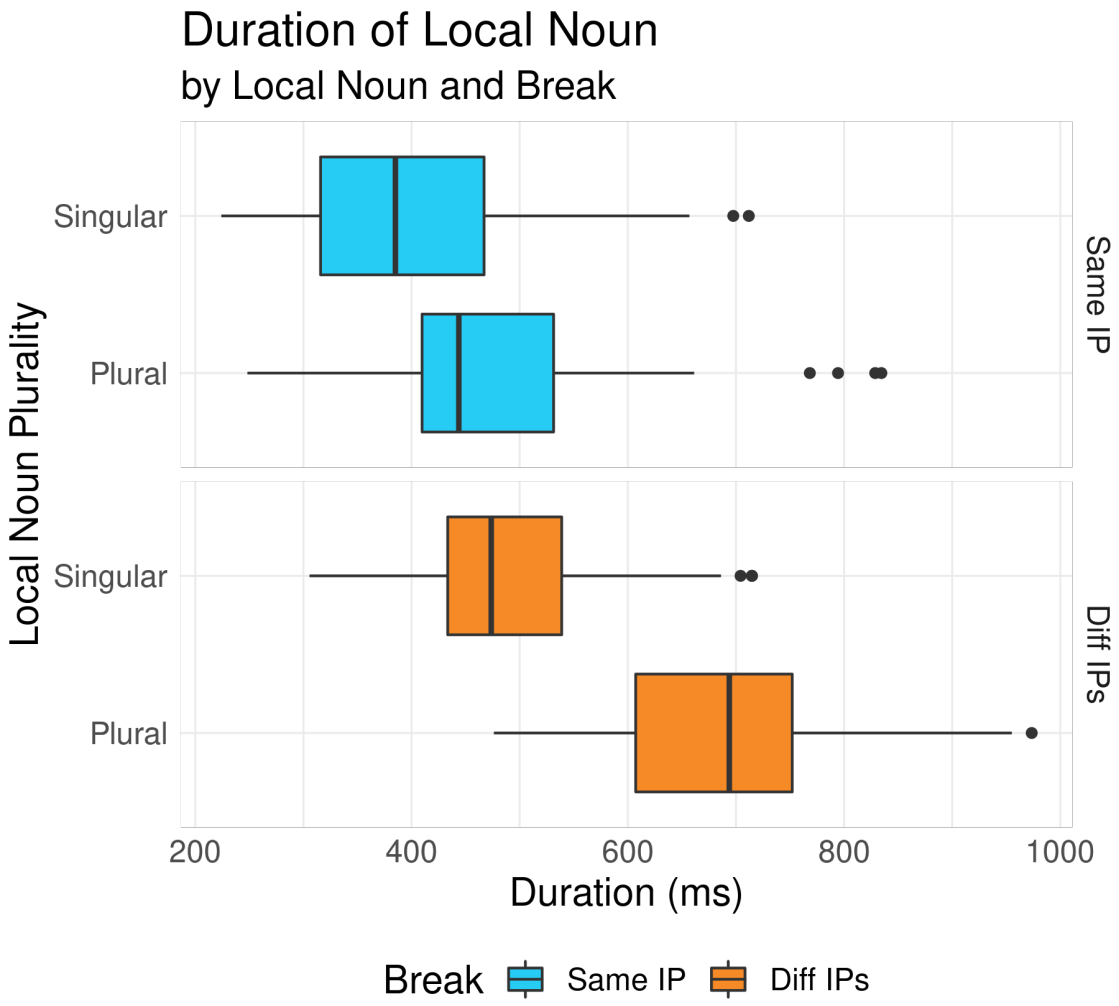


Figure 3.2: Duration of the Local Noun across prosody and plurality conditions.

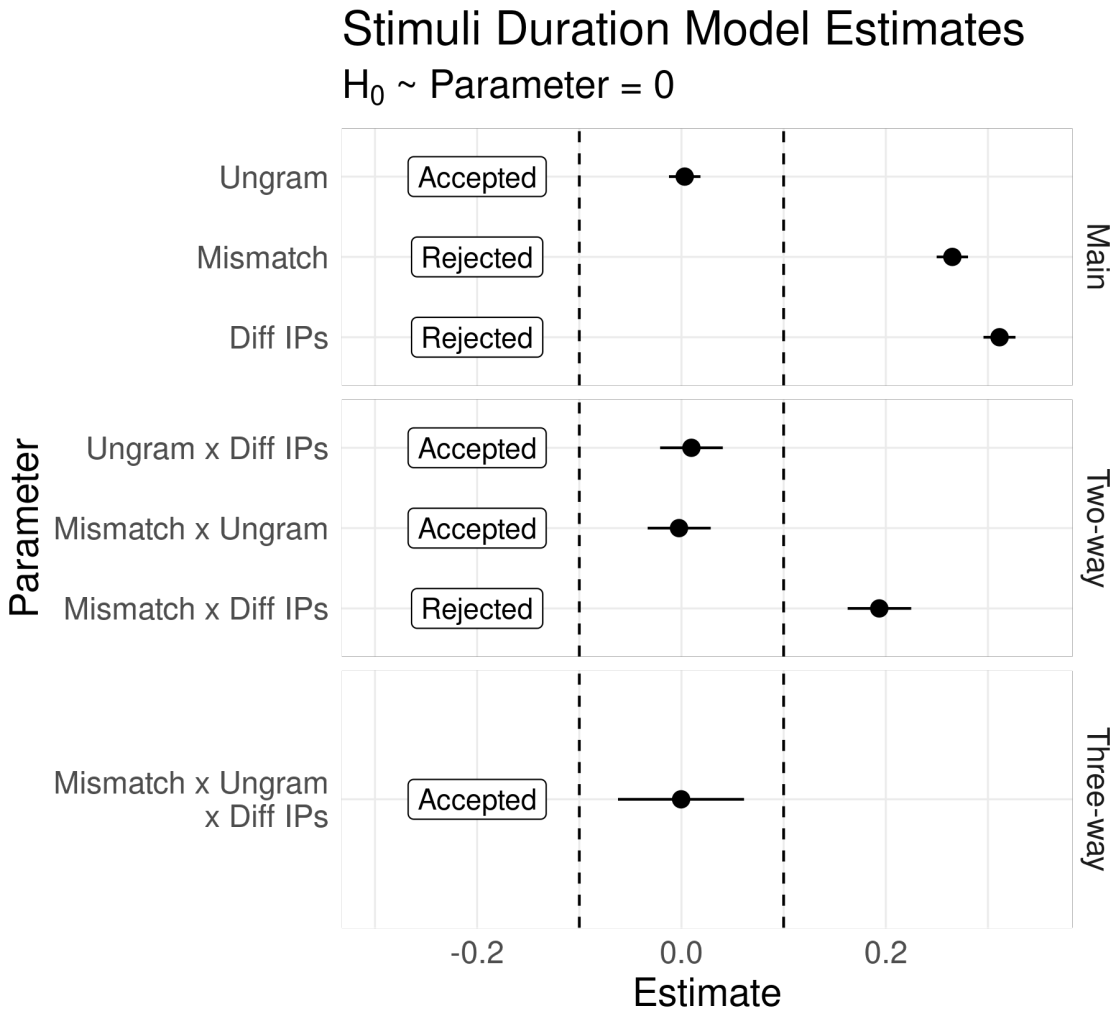


Figure 3.3: 89% credible intervals for the parameter estimates produced by the model. This includes main effects of local noun number, verb number/grammaticality, and phrasing condition. The result of a Test of Practical Equivalence.

Parameter	5.5%	Median Est.	94.5%	ROPE Equivalence
Mismatch	0.25	0.27	0.28	Rejected
Ungram	-0.01	0.00	0.02	Accepted
Diff IPs	0.30	0.31	0.33	Rejected
Mismatch x Ungram	-0.03	0.00	0.03	Accepted
Mismatch x Diff IPs	0.16	0.19	0.22	Rejected
Ungram x Diff IPs	-0.02	0.01	0.04	Accepted
Mismatch x Ungram x Diff IPs	-0.06	0.00	0.06	Accepted

Table 3.6: Coefficients of stimuli duration model

are getting slightly more time to process the subject NP in this condition, but it is inevitable due to the nature of final lengthening.

3.2.2 Procedure

Once participants consented to participate in the experiment, they were shown instructions (shown in Figure 3.4) for the task. The goal of the instructions was to encourage the listeners to rely solely on their native speaker intuition when judging the "acceptability" of an utterance they heard. They were explicitly instructed not to rely on prescriptive rules learned in academic settings. Because these explicit instructions can only do so much, the 3sec time limit per response was also motivation for not "overthinking" a given sentence.

After the instructions, three practice trials familiarized the participant with the task. A given trial proceeded as follows: A red screen with "Get Ready" appeared for 500ms to alert the participant to the incoming sound. Then, the screen turned yellow with an X in the middle, and the recording of the stimuli played. As soon as the recording finished playing, the screen turned green. Two buttons labeled "acceptable" and "unacceptable" appeared, along with the keyboard key (i.e., "F" and "J" respectively) corresponding to that answer.

The 48 critical trials were counterbalanced in a Latin square design across eight pseudo-randomized lists.

3.2.3 Participants

A total of 114 participants completed the experiments. On average, participants completed the experiment in 20min (sd 9.6min). 5 participants were excluded from the experiment for completion times above 2 standard deviations from the mean (i.e.,>39min). 6 participants were excluded for letting the trial auto-advance without selecting a response for 5% or more of critical trials. 12 participants were excluded for not providing an answer on 5% or more of the dialect survey questions. An additional 2 participants' data had to be excluded due to technical problems with the reported RTs. After exclusions, this left 89 participants for analysis.

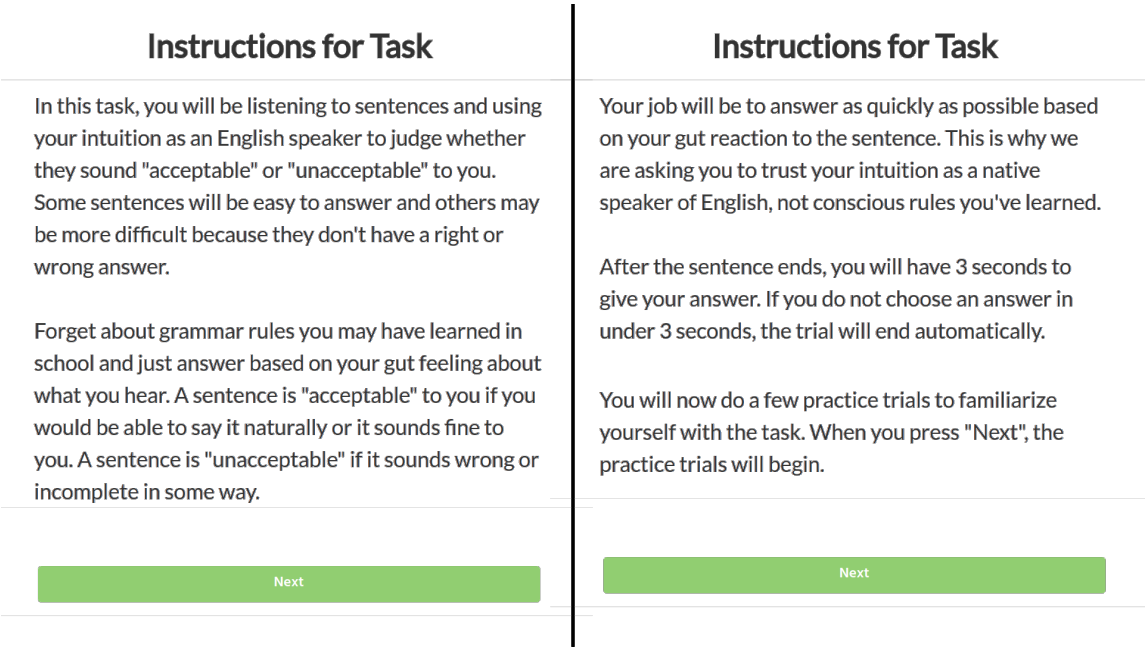


Figure 3.4: A screenshot of the two instruction pages that preceded the practice trials.

Figure 3.5: A diagram showing the sequence of actions in a given trial.

3.2.4 Data Analysis and Statistical Methods

Because of the nature of the acceptability task, two measurements were given for each trial: the binary acceptability judgment (binomial variable) and the amount of time it took for the participants to give their response (continuous variable).

The acceptability judgement data were coded as 0s (*unacceptable*)and 1s (*acceptable*) and were modeled in a Bayesian hierarchical model in brms (Bürkner, 2017; Bürkner, 2017b). Population-level effects included main effects of local noun number (SS/SP), verb (GRAMMATICAL/UNGRAMMATICAL), prosodic condition (SAME IP/DIFFERENT IP), and dprime score for sensitivity to default singular agreement as well as all their interactions up to 4-ways. The categorical factors were contrast coded (-0.5 & 0.5) such that the intercept of the model reflects the grand mean (UCLA Statistical Consulting Group, 2011). The statistical reference levels Match, Grammatical, Same IP were coded as -0.5. Group-level effects included by-participant and by-item intercepts and slopes for all three of the factors as mentioned earlier.

The binomial response data were modeled using a Bernoulli distribution. All priors were set to be weakly informative. Weakly informative priors are used to constrain the search space of the model without having an undue influence on the estimates of the model. The *brms* code for generating the model is given in Figure 3.6.

The continuous variable, RT, was winsorized with 2.5% tails in order to deal with outliers. Winsorization is a method of taking data points that fall outside some defined distance from the mean and recoding them as outer limit values (Dixon and Tukey, 1968). For the RTs, the 2.5% fastest and 2.5% slowest RTs for each participant within each condition were recorded (i.e., lowered or raised respectively) to the 2.5% threshold level. Also, the binary choice acceptability data are not affected by winsorization, so they were left untransformed.

The winsorized RTs were modeled in a Bayesian hierarchical model in brms (Bürkner, 2017; Bürkner, 2017a). Population-level effects included the same factors as previously mentioned in addition to a main effect of trial response (ACCEPTABLE/UNACCEPTABLE)

(see Section 3.3.3.1 of this chapter for a more detailed discussion of this addition) and all interactions with the rest of the factors up to 4-ways. Trial response was contrast coded with "acceptable" coded as -0.5.

The same group-level effects were included in the RT model as were in the judgment model. One difference, however, is that the by-item random slope and intercept for dialect were left out due to problems with model convergence.

In order to model the logarithmic tail of RTs, a shifted log-normal likelihood was selected. A shifted-log normal distribution is used to model "time-to-event" data and includes a parameter for non-decision time (NDT). NDT is a positive number that shifts the logarithmic distribution rightward by some amount (Bürkner, 2021). In their paper on the relationship between RT mean and RT standard deviation, Wagenmakers and Brown (2007) lay out the characteristic "nonnormal" nature of RT distributions and some of the types of descriptive distributions used to characterize them. While both ex-Gaussian and shifted log-normal distributions can fit RTs well (Ratcliff and Murdock, 1976), I chose to use a shifted log-normal because its theoretical assumption is more in line with modeling psychological processes like response time data (Bürkner, 2020). Namely, it assumes that the shifted log-normal distribution reflects a decision time process plus a certain degree of shift in the distribution, which accounts for non-decision time (NDT) and relates to task difficulty.

As in the previous model, priors were weakly informative, and the *brms* code for the model can be found in Figure 3.7. Priors are the a priori expected distribution of estimates of a parameter. The model uses the priors to constrain the search space and rule out highly improbable values. The priors are updated after the observation of data, however, which gives us our posterior distribution. The priors are "weakly informative" in that they are relatively wide (2 standard deviations from the mean), centered on 0, and do very little to change the posterior distribution with even the observation of only a few data points. Therefore, we do not have to worry about the priors strongly influencing our model estimates.

Factor Levels	Stimuli Example
Subject NP: SG-SG/SG-PL	The key to the cabinet... vs. The key to the cabinets...
Verb: GRAMMATICAL/ UNGRAMMATICAL	The key ... is... vs. The key ... are...
Phrasing: SAME IP/DIFF IPS	|The key to the cabinets are...| vs. |The key to the cabinets| |are...|

Table 3.7: Examples of the levels of each categorical factor in the model.

```
Ratingmodel <- brm(Acceptability ~ LocalNoun * Verb * Phrasing  +
                 (1 + LocalNoun * Verb * Phrasing | Item) +
                 (1 + LocalNoun * Verb * Phrasing  | Subject),
               data = Exp2Data,
               family = bernoulli(),
               prior = set_prior("normal(0,2)", class = "b"),
               cores = 8,
               iter = 4000,
               warmup = 2000,
               chains = 8)
```

Figure 3.6: The *brms* code used to run the Bayesian hierarchical model for participant acceptability judgements.

3.3 Results

3.3.1 Review of Main Questions and Hypotheses

Before discussing the experimental results, a review of the main questions of this study follows. Firstly, this comprehension study used auditory stimuli to test how overt prosodic phrasing affects the acceptability of agreement attraction sentences and the speed with which those judgments are made. The two prosodic hypotheses under consideration are the *Chunk-Internal Availability* and *Accentual Processing Boost* hypotheses.

```
RTmodel <- brm(winsorRT ~ LocalNoun * Verb * Phrasing
                                                  * Response *
                 (1 + LocalNoun * Verb * Phrasing * Response | Item) +
                 (1 + LocalNoun * Verb * Phrasing * Response | Subject),
               data = Exp2Data,
               family = shifted_lognormal(),
               prior = set_prior("normal(0,2)", class = "b"),
               cores = 8,
               iter = 4000,
               warmup = 2000,
               chains = 8)
```

Figure 3.7: The *brms* code used to run the Bayesian hierarchical model for participant RTs.

The *Chunk-Internal Availability* is based on the fact that the prosodic phrases form processing chunks that can aid parsing and the construction of the meaning of elements that occur within the same phrase. In this hypothesis, elements within the same ip or IP are easier to resolve a dependency between than two elements in different IPs.

On the other hand, the *Accentual Processing Boost* hypothesis relates more to nuclear pitch accenting and how its phonological prominence boosts retrieval of the accented word. Here, a word that carries a nuclear pitch accented is more deeply encoded in memory and thus easier to retrieve later.

The two hypotheses make opposing predictions about the rate of acceptability of agreement attraction and the RTs for the judgments. For *Chunk-Internal Availability*, agreement attraction sentences should be rated as acceptable more often when the local noun and verb are in the SAME IP as compared to DIFFERENT IPS. Since the local noun's retrieval is facilitated by being in the same prosodic processing chunk, the interference caused by the local noun is greater than when phrased separately. On the other hand, when the local noun is phrased separately from the verb, it occurs at the right edge of an IP and carries a nuclear pitch accent. The nuclear pitch accent leads to the local noun being processed more deeply, which allows it to be chosen as the source of agreement. Therefore, we

would expect to see more interference of the local noun in the DIFFERENT IPS condition compared to the SAME IP condition.

3.3.2 Ratings

3.3.2.1 Model Convergence and Fit

Model fit convergence was checked by examining the Gelman-Rubin diagnostic, Rhat, and the Effective Sample Size of the model parameters. Rhat is a diagnostic used to check the mixture of Markov Chain Monte Carlo (MCMC). The ideal Rhat value is 1, and any value above 1.1 indicates that the chains may not have converged, which would require further iterations of the model to be run (McElreath, 2020). The ESS of a parameter estimate is used to check the auto-correlation of the Markov chain posterior samples. If the ESS is significantly lower than the number of total iterations, then the estimates are likely unreliable (Stan Development Team, 2020). However, a general rule is that if the ESS is great than 1000, estimates are stable (Bürkner, 2018). For the parameter estimates of this model, all Rhats were 1, and ESSs were between 4000 and 6500 samples out of 8000 post-warmup samples.

Another way of evaluating the model's fit is by comparing a posterior predictive check (PPC) to compare the model's predictions to the actual data collected. A PPC is a simple visual to see if the model is roughly capturing the distribution of the data (McElreath, 2020). The PPC for the Rating model is seen in Figure 3.8. The dark blue circle represents the average predicted count of both response times across 100 simulated data sets from the model. The light blue bar shows the actual counts observed in the collected data. The variation in simulated counts across the 100 datasets is so tiny that the 89% CrI for the average is not visible on the graph. Our actual data fit so well within the simulated dataset could be a sign of model over-fitting.

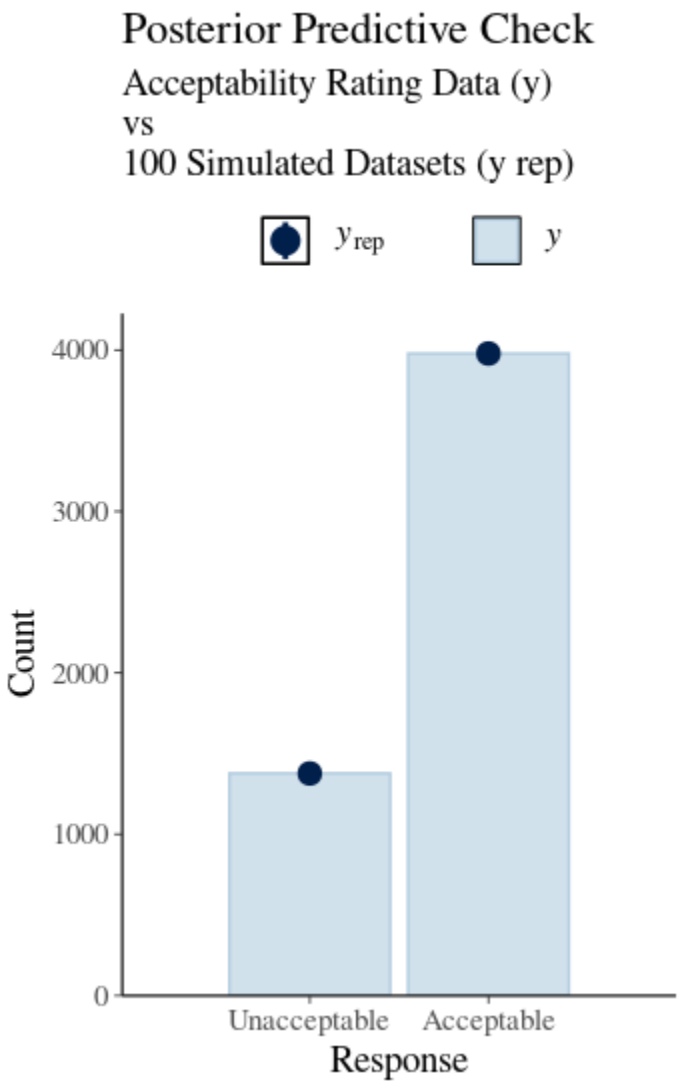


Figure 3.8: Posterior Predictive Check comparing the actual data, y, with 100 simulated data based on the model predictions, y_{rep}.

			Response Type	
Subject NP	**Verb**	**Phrasing**	acceptable	unacceptable
Sg-Sg (Match)	Sg (Gram)	Same	494	39
		Diff	493	40
	Pl (Ungram)	Same	226	307
		Diff	230	299
Sg-Pl (Mismatch)	Sg (Gram)	Same	485	48
		Diff	467	66
	Pl (Ungram)	Same	390	141
		Diff	368	165

Table 3.8: Counts of acceptable and unacceptable responses broken down by condition.

3.3.2.2 Model Results

The samples from the posterior predictive distribution are shown in Figure 3.9 and estimates of the parameter values are shown in Figure 3.10. When all of the probability mass lies above or below 0, we can be very confident in the presence of an effect in a particular direction Kruschke (2018). In this case, the arbitrary distinction between "significant" and "non-significant" effects (see Vasishth et al. (2018) and van Zwet and Cator (2021) for more on the perils of the “significance filter”) is replaced instead with a continuous valuation of our confidence of the presence, size, and direction of the effect using 89% Credible Intervals (i.e., CrIs). 89% CrIs were used as opposed to 95% CrIs because 95% CrIs are often unnecessarily conservative and 95% itself is only significant due to its convenience of being close to the percentage of data in a normal distribution that is within ±2 standard deviations from the mean (McElreath, 2020).

The posterior predictive distribution shown in Figure 3.9 exhibits a strong, and expected, effect of grammaticality such that ungrammatical sentences (i.e., those with a plural verb) were rated much lower than those that were grammatical, (β = -3.24, CrI = [-2.79, -2.35])). This effect of grammaticality can be confirmed by looking at the counts of re-

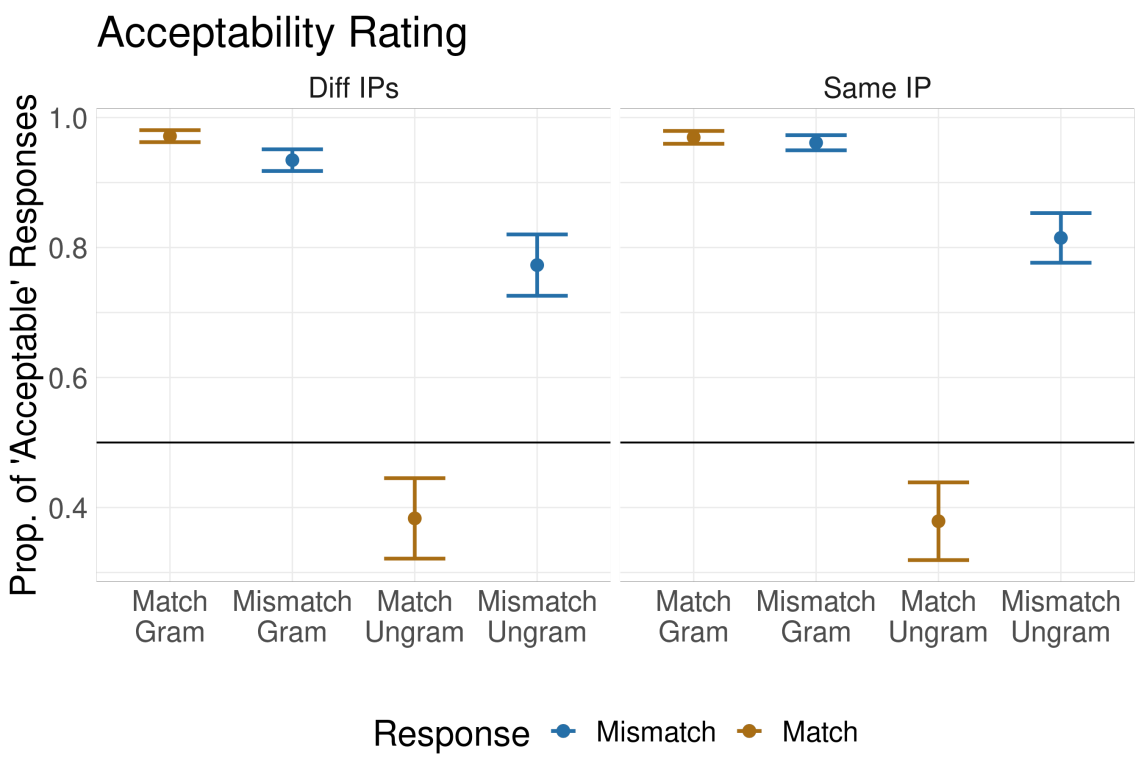


Figure 3.9: Average acceptability rating for each condition. '1' is an acceptable response and '0' is an unacceptable response.

sponses by condition in Table 3.8.

A main effect of local noun number was detected, such that plural local nouns were rated higher than singular local nouns (β = 0.64, CrI = [0.39, 0.9]) .

In an interaction of local noun and verb number (to be referred to as the 'agreement attraction' interaction), the rating data show that the difference between ungrammatical and grammatical verb is much smaller for plural local nouns than singular local nouns. This result is evidence of the robust agreement attraction effect. As for phrasing's influence, we cannot decide on whether or not the effect is non-null. However, this effect seems to be that plural local nouns are rated lower when phrased separately from the verb as compared to when they are phrased together.

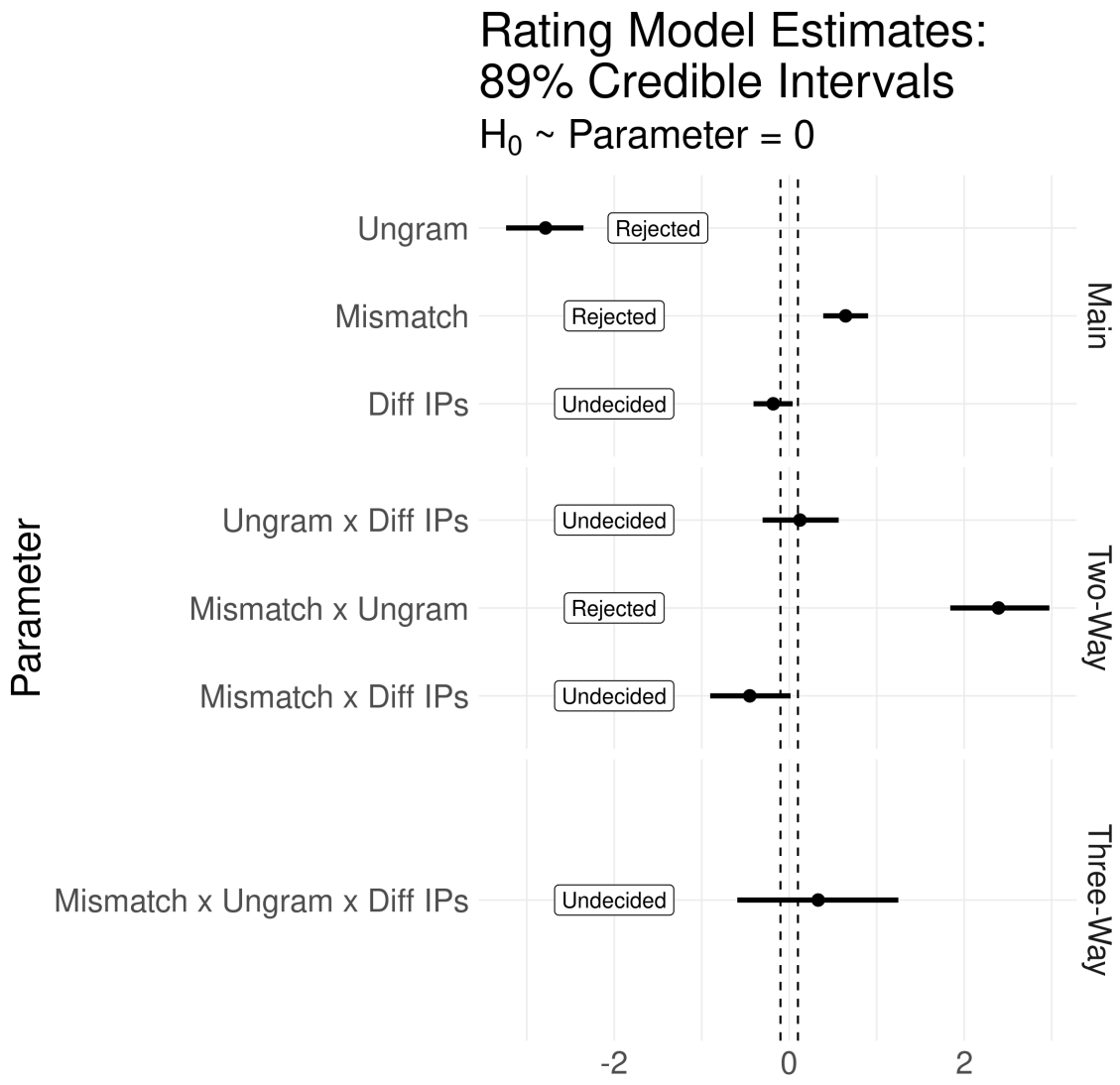


Figure 3.10: 89% credibility intervals for estimates of the Ratings parameters.

Parameter	5.5%	Median Est.	94.5%	ROPE_Equivalence
Mismatch	0.39	0.64	0.90	Rejected
Ungram	-3.24	-2.79	-2.35	Rejected
Diff IPs	-0.41	-0.18	0.04	Undecided
Mismatch x Ungram	1.84	2.39	2.97	Rejected
Mismatch x Diff IPs	-0.90	-0.45	0.02	Undecided
Ungram x Diff IPs	-0.30	0.12	0.57	Undecided
Mismatch x Ungram x Diff IPs	-0.59	0.33	1.25	Undecided

Table 3.9: Parameter estimates from the ratings model.

3.3.3 Time-to-decision RTs

3.3.3.1 Decision on model parameter selection

A baseline model including all and only the parameters of the Rating model was first fit and compared to a larger model that also included ACCEPTABILITY as a parameter (including all interactions). The larger model coded ACCEPTABILITY as being the expected or unexpected response given the verb's number in the particular condition—this coding of the responses allowed for a simpler interpretation of the 3- and 4-way interactions. Since the head noun was always singular across all conditions, an "acceptable" response was expected in the GRAMMATICAL and an "unacceptable" response was expected in the UNGRAMMATICAL. These two types of responses were coded together and are referred to for the rest of the paper as "correct". If the participant gave an "unacceptable" response to a GRAMMATICAL condition or an "acceptable" response to an UNGRAMMATICAL condition, the response was coded as "incorrect". The correct/incorrect coding is not meant to reflect prescriptive values of agreement but rather an easily digestible label for later interpretation and discussion of the results.

The larger of the two models is presented here because the effect between the two models was not drastically different from one another, and the larger model is more con-

ceptually in line with the research questions of this study. In particular, questions about the speed of accepting or rejecting agreement attraction sentences require the coding of the response type. Thus, if these two types of responses were lumped together into a smaller model, we would miss out on effects that may address this question of speed of decision making.

3.3.3.2 Model Convergence and Fit

For the parameter estimates of this model, all Rhats were 1, and most ESSs were between 5000 and 10000 samples out of 12000 post-warmup samples. Only two parameters had ESSs under 5000, but their values were still far above the 1000 sample heuristic.

The PPC in Figure 3.11 shows the actual RT data plotted in dark blue as a density graph with the 100 simulated datasets from the model in light blue. While the general shape of the simulated datasets aligns with the observed RTs, the model does not seem to capture the RT peak around 400ms. Additionally, the model seems to over-generate RTs between 600ms and 1200ms.

3.3.3.3 Model Results

Offline judgments were not influenced by prosody; however, the online RTs tell a different story. Table 3.12 shows the median, 89% credibility intervals, and the Region of Practical Equivalence for each parameter estimate. Additionally, a table of the mean winsorized RTs (and standard deviations) for each condition appears in Table 3.10.

Figure 3.13 shows the parameter estimates for the main and two-way interactions effects. With regard to main effects, number MISMATCHes in the subject NP resulted in overall longer RTs than number matched subject NPs (β = 0.17, CrI = [0.13, 0.22]) (see Figure 3.12). Additionally, incorrect responses took much longer than correct ones (β = 0.16, CrI = [0.11, 0.22]) . However, the entire probability mass of phrasing's estimate was within the bounds of the ROPE, meaning we can be reasonably confident in saying that the phrasing effect is negligible.

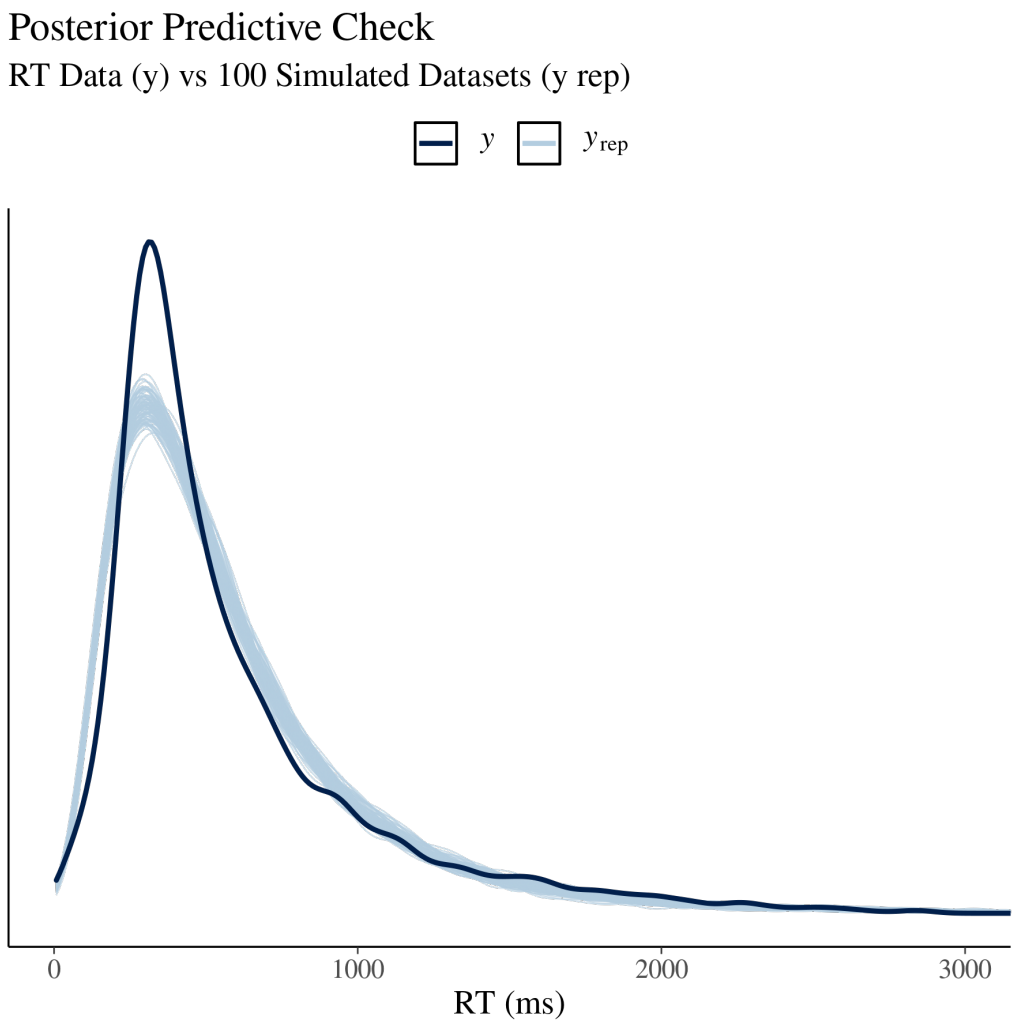


Figure 3.11: Posterior Predictive Check comparing the actual RT data, y, with 100 simulated datasets based on the model predictions, y_{rep}.

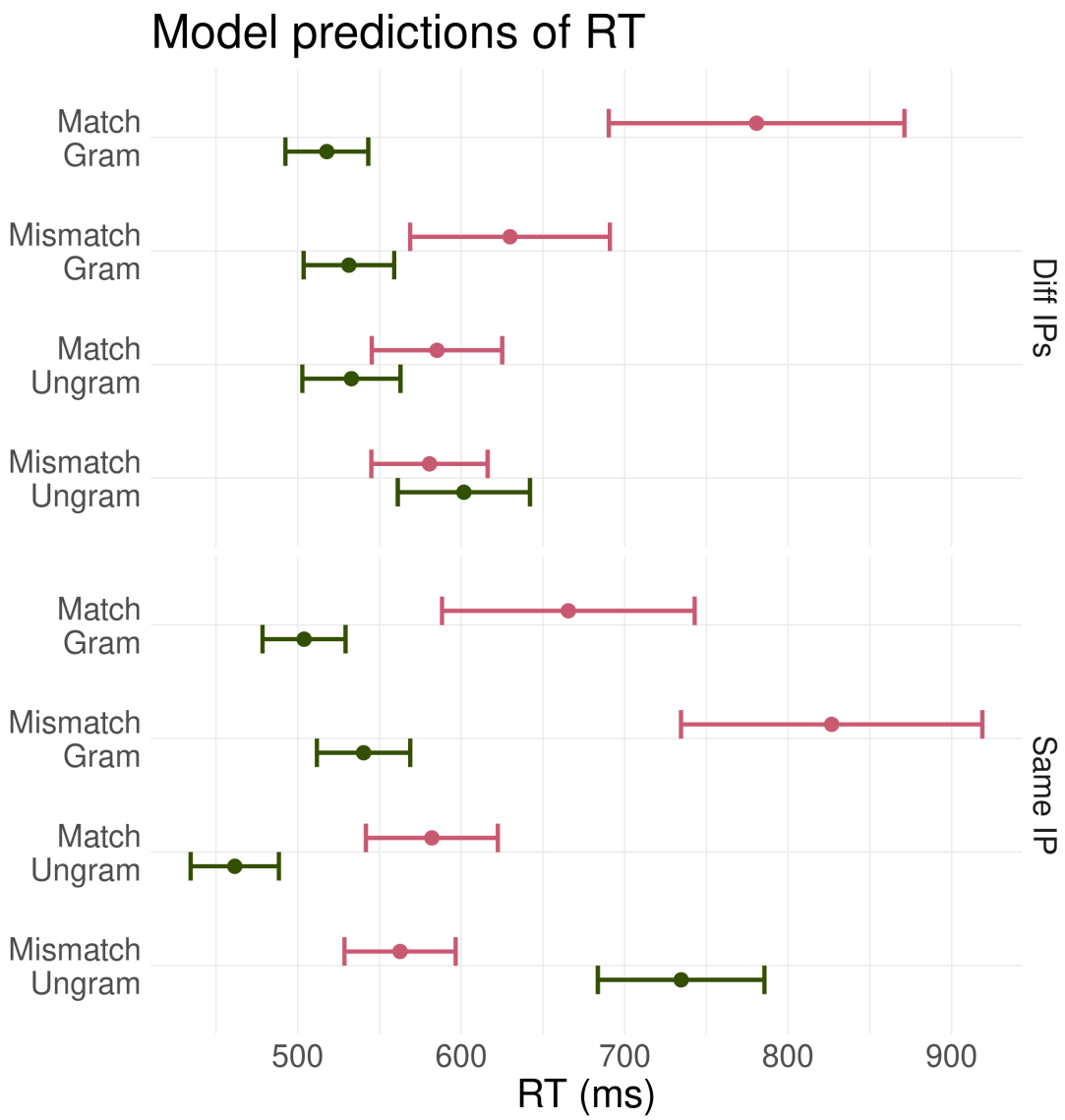


Figure 3.12: Samples from the posterior predictive distribution showing for participant RT by condition.

LocNoun	Verb	Break	correct	incorrect
Match	Gram	SameIP	537 (18)	694 (74)
Match	Gram	DiffIP	542 (17)	849 (94)
Match	Ungram	SameIP	533 (25)	595 (28)
Match	Ungram	DiffIP	587 (26)	591 (30)
Mismatch	Gram	SameIP	597 (21)	856 (82)
Mismatch	Gram	DiffIP	559 (18)	682 (60)
Mismatch	Ungram	SameIP	794 (49)	579 (21)
Mismatch	Ungram	DiffIP	658 (41)	620 (24)

Table 3.10: Mean winsorized RTs in milliseconds and their standard errors in parentheses by condition.

For interactions, the package emmeans (Lenth, 2021) was used to obtain the estimated marginal means of contrasts. Computing the estimated marginal means allows for a pairwise comparison of conditions. For example, the interactions between local noun plurality and response type seem to exclude the entire width of the ROPE, meaning we can be confident that there is some effect there. After computing a pairwise comparison of conditions using emmeans (see Table 3.11), it seems that this effect is driven by the fact that there is no difference between the RT of correct and incorrect responses when the verb is ungrammatical, (β = -0.01, CrI = [-0.09, 0.08]) . However, there is a difference between the two conditions in when the verb is grammatical such that correct responses are faster than incorrect responses, (β = -0.32, CrI = [-0.44, -0.22]) . The coefficients reported will be from the pairwise comparison of estimated marginal means for the rest of these results. A table of the model's parameter estimates, however, can be found in Table 3.12.

As for the speed of the agreement attraction time-to-decision, when the local noun is singular, the RT for grammatical and ungrammatical verbs is the same, (β = -0.01, CrI = [-0.10, 0.09]) . When the local noun is plural, however, responses to ungrammatical verbs are much slower than responses to grammatical verbs (β = -0.12, CrI = [-0.22, -0.03]) .

Verb	Contrast	Est.	5.5%	94.5%
Ungram	correct - incorrect	-0.01	-0.09	0.08
Gram	correct - incorrect	-0.32	-0.44	-0.22

Table 3.11: Pairwise comparison of the estimated marginal means of the interaction between local noun number and response type.

Additionally, the interaction between local noun number and response types shows that correct responses were faster than incorrect responses when the local noun was singular (β = -0.26, CrI = [-0.35, -0.16]) and nearly so in the local noun was plural (β = -0.07, CrI = [-0.16, 0.02]) . This interaction's probability mass does not completely exclude the ROPE, which is why the response type difference for plural local nouns does not entirely exclude 0.

Interestingly, there also seems to be an effect of phrasing on the time-to-decision. For singular local nouns and all verbs that are phrased in the same IP produce faster responses than when they are phrased separately (β = 0.08, CrI = [-0.01, 0.17]) . This effect goes in the opposite direction when the local noun is plural however, (β = -0.11, CrI = [-0.20, -0.03]) .

Concerning 3-way interactions, the interaction between local noun number, verb number, and response type seems to show that the two-way interaction of local noun and verb is driven by very fast (<550ms) correct responses, which counteract the slow (>550ms) incorrect responses (see Figure 3.14). Incorrect responses for agreement attraction configurations get a speedup, which nullifies the main effect of response type (i.e., correct faster than incorrect).

The effect on phrasing for the RT of agreement attraction sentences seems to follow an interesting pattern. The 2-way effect of local noun and phrasing seems to be solely driven by the ungrammatical verb condition, as seen in Figure 3.15.

Finally, the data do support the four-way interaction as well. This effect is born out in the difference between correct and incorrect responses in the agreement attraction condi-

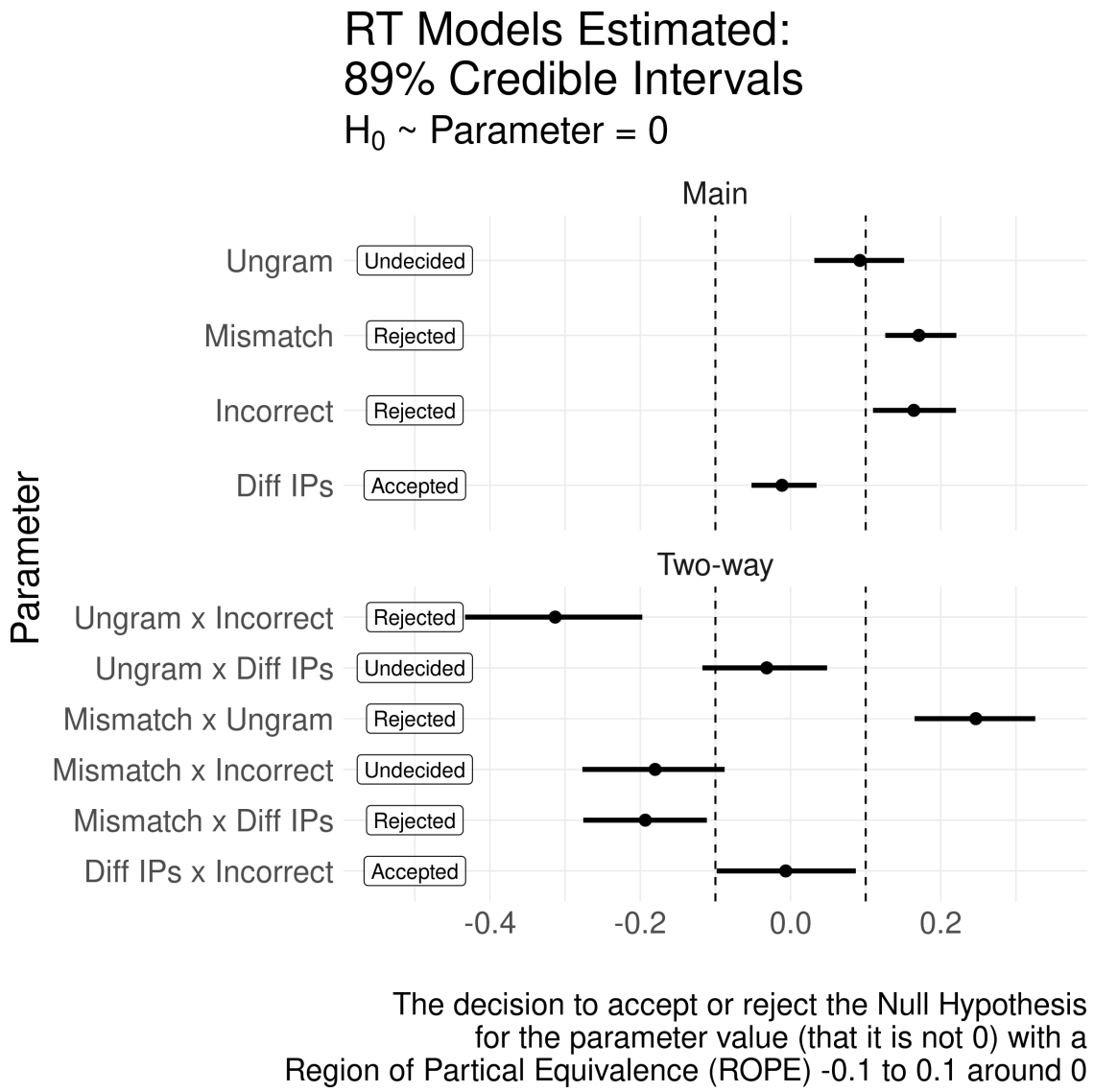


Figure 3.13: 89% Credible Intervals for estimates of the RT parameters.

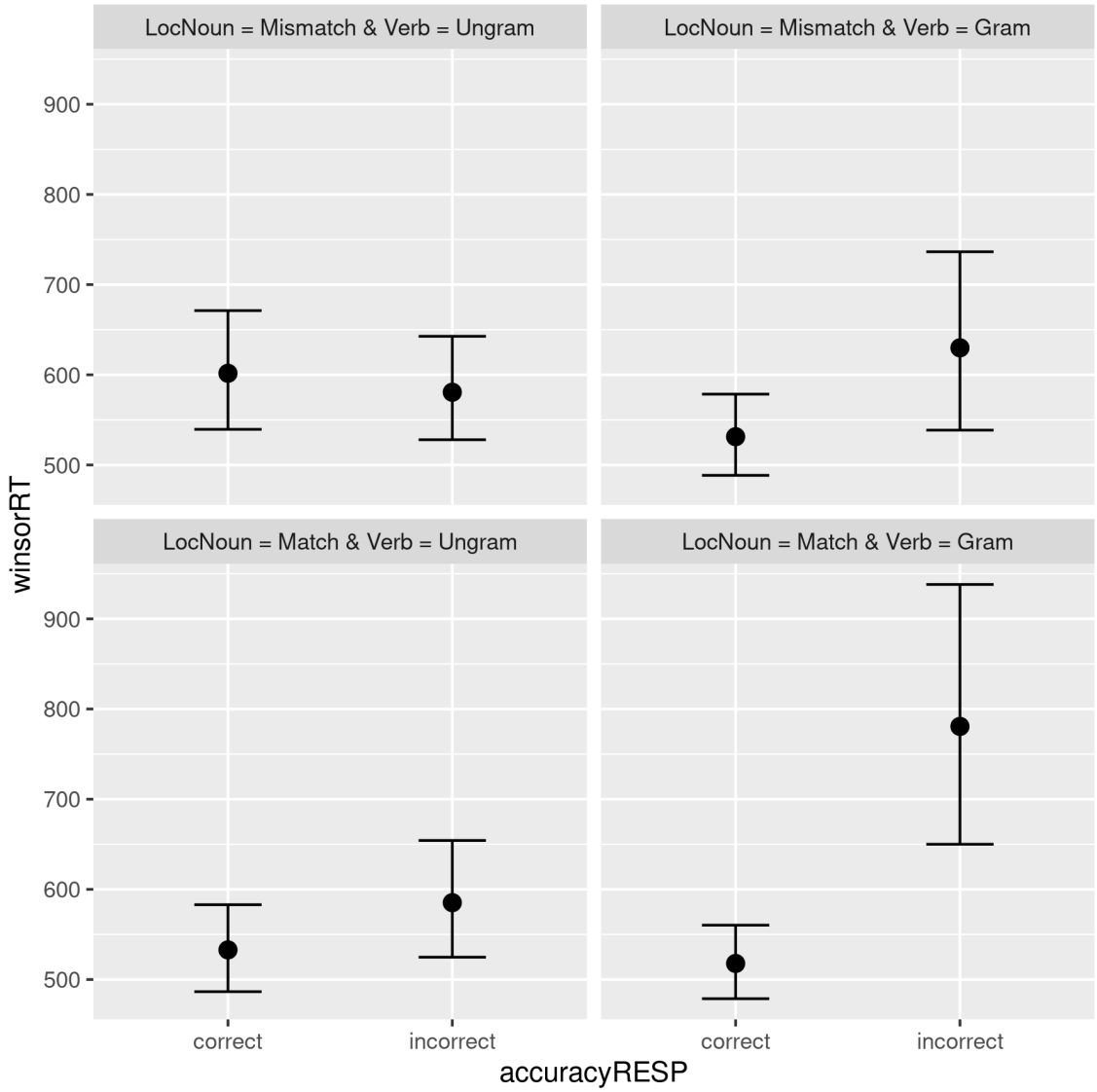


Figure 3.14: Estimated marginal means pairwise comparisons for Mismatch x Grammaticality x Response Type

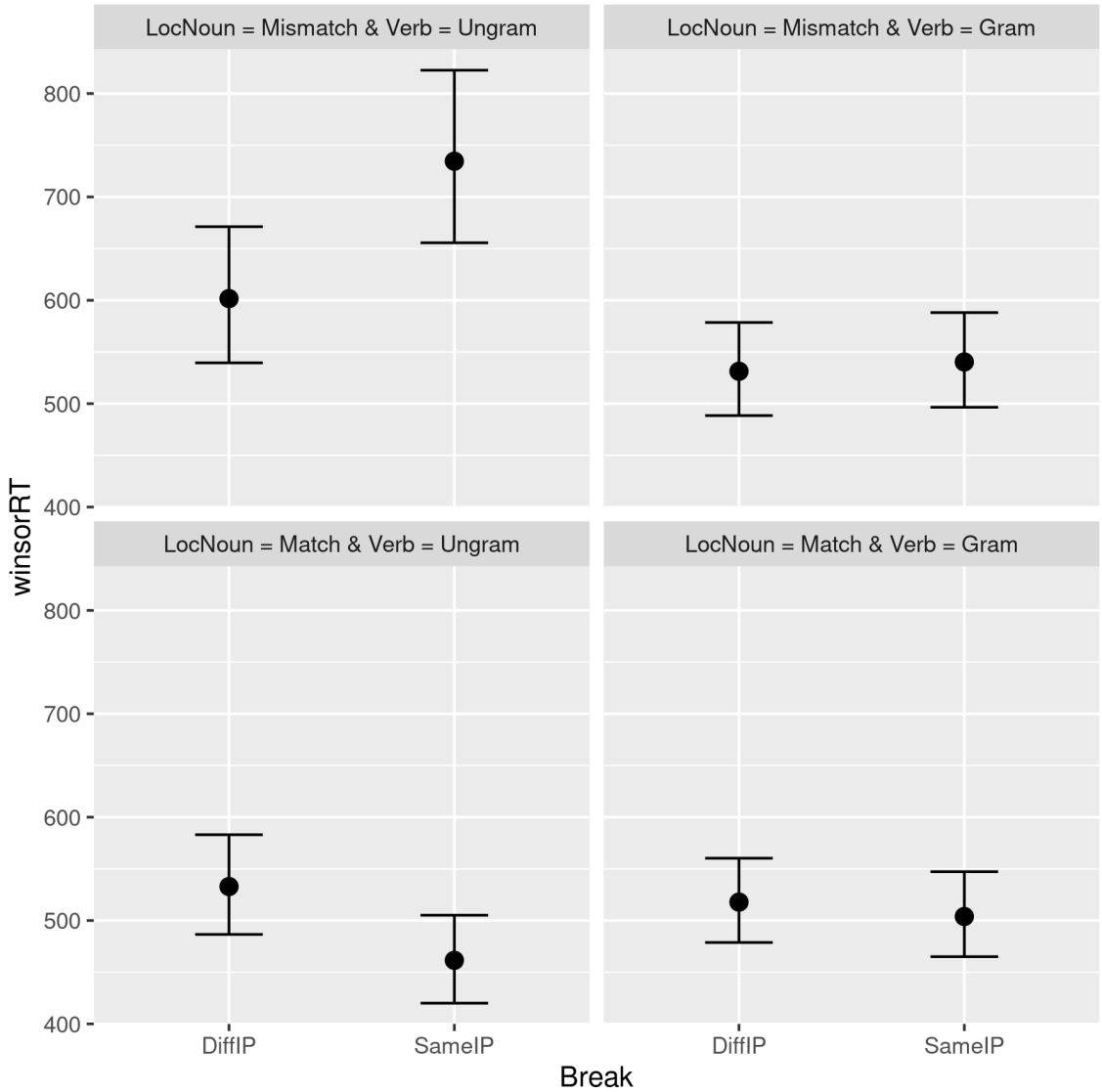


Figure 3.15: Estimated marginal means pairwise comparisons for Mismatch x Phrasing x Response Type

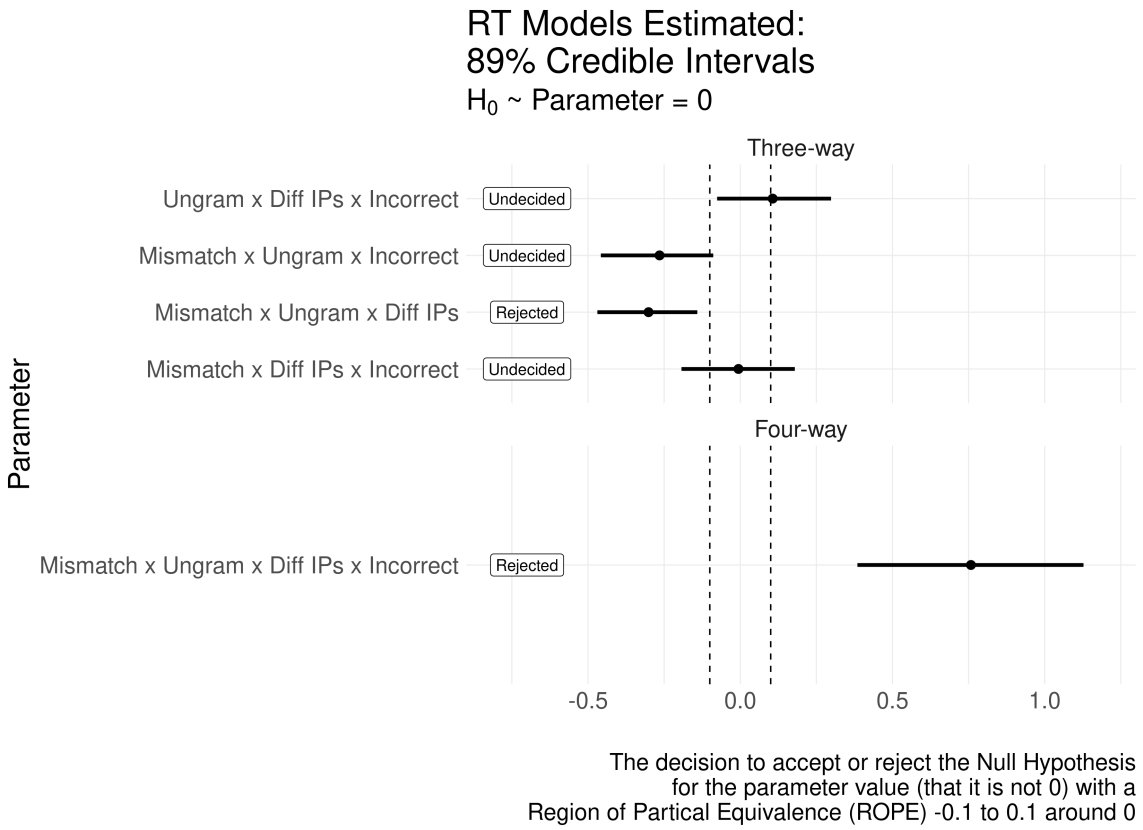


Figure 3.16: 89% Credible Intervals for estimates of the RT parameters.

tion, which can be seen in Figure 3.12. Correct responses are much slower than incorrect responses when the plural local noun and plural verb are phrased together. In this case, that means that unacceptable responses were slower than acceptable ones. However, correct/unacceptable responses are much faster, in fact, as fast as incorrect/acceptable responses, when the local noun and verb are in separate IPs.

Parameter	5.5%	Median Est.	94.5%	ROPE Equivalence
Mismatch	0.13	0.17	0.22	Rejected
Ungram	0.03	0.09	0.15	Undecided
Diff IPs	-0.05	-0.01	0.03	Accepted
Incorrect	0.11	0.16	0.22	Rejected
Mismatch x Ungram	0.17	0.25	0.33	Rejected
Mismatch x Diff IPs	-0.28	-0.19	-0.11	Rejected
Mismatch x Incorrect	-0.28	-0.18	-0.09	Undecided
Ungram x Incorrect	-0.43	-0.31	-0.20	Rejected
Diff IPs x Incorrect	-0.10	-0.01	0.09	Accepted
Mismatch x Ungram x Incorrect	-0.46	-0.27	-0.09	Undecided
Mismatch x Diff IPs x Incorrect	-0.19	-0.01	0.18	Undecided
Ungram x Diff IPs x Incorrect	-0.08	0.11	0.30	Undecided
Ungram x Diff IPs	-0.12	-0.03	0.05	Undecided
Mismatch x Ungram x Diff IPs	-0.47	-0.30	-0.14	Rejected
Mismatch x Ungram x Diff IPs x Incorrect	0.38	0.76	1.13	Rejected

Table 3.12: Table of median parameter estimates with 89% Credible Intervals from the RT model.

3.4 Discussion

3.4.1 Ratings support grammaticality asymmetry

The classic agreement attraction effect was replicated in the rating data; there was a sharp decrease in acceptability in the conditions with an UNGRAMMATICAL verb ("The key ... are ..."), but this decrease was ameliorated in the presence of a local plural noun ("The key to the cabinets are"). This increase in acceptability of ungrammatical sentences is evidence that participants did indeed experience an agreement attraction effect.

Representational accounts of attraction predict a symmetrical effect of attraction such that sentences with grammatical verbs and plural local nouns ("The key to the cabinets is") are rated lower in acceptability than those same sentences with singular local nouns ("The key to the cabinet is"). The reasoning behind this is that the entire subject NP is slightly plural due to the plural morphology on the local noun. Thus, it should occasionally be treated as plural for agreement. If the subject NP is treated as plural in comprehension, encountering a singular verb should result in an "unacceptable" or "ungrammatical" response. However, the data here more strongly support a grammaticality asymmetry, where agreement attraction only occurs in sentences with plural verbs (i.e., UNGRAMMATICAL verbs). Ratings for sentences with SG-PL subject NPs and GRAMMATICAL verbs did not result in a detectably lower acceptability rating that we can be confident in, which is predicted by Wager's cue-based retrieval account of agreement attraction.

3.4.2 No detectable effect of phrasing on ratings

The ratings of the sentences across phrasing conditions suggest a small negative effect of phrasing interacting with local noun number. However, based on these data, we cannot make firm conclusions about this effect. The lack of a robust prosodic effect in acceptability ratings could be evidence of a limited phrasing role in agreement processing. More specifically, phrasing may not affect the degree of interference caused by the local noun in agreement attraction sentences. In both of the previously proposed prosodic hypotheses,

phrasing was predicted to either attenuate or amplify the effect of interference and thus influence participants' judgments of the sentences. The lack of such an effect means that phrasing may not bear on the grammatical factors of agreement attraction. The effects we detect in the rating data are driven by morphological factors, with negligible influence from phrasing.

These results do not, however, tell us what kind of final interpretation the participants came to. One of the limitations of this design is that it does not address this question, but this choice to exclude questions about the subject NP was made purposefully. In order to disguise the main manipulation of this experiment, questions probing the participants' understanding of the plurality of the subject NP were avoided. On the one hand, this design choice was successful because few participants mentioned the subject-verb plurality mismatches in their debriefing surveys responses. On the other hand, the only information we have about the representation of the subject NP that the participants built is the acceptability rating that they gave. This information alone does not give us a clear picture of how the participants interpreted these NPs. For example, let us consider the case when a participant hears "The key to the cabinets are on the table", with the subject NP and verb in the same IP, and judges this sentence to be "acceptable". In the *Marking and Morphing* account of attraction, this "acceptable" answer is because that entire NP is slightly plural. The rate at which the subject is treated as grammatically plural is determined probabilistically as a function of the subject NP's gradient plurality value (i.e., *S(r)*). Therefore, if the participant were asked whether there was a single key or multiple keys, they may respond that there were multiple because the representation they built for the entire subject NP was plural. Let us now consider a *Cue-based Retrieval* type account like the one proposed by Wagers. The participant may still respond that there was only a single key despite answering that the sentence is "acceptable" with a plural verb. This seeming contradiction is not a concern because the issue of attraction in Wagers' account is in the retrieval of the local noun as the controller of agreement on the verb. Therefore, the actual plurality of the subject NP is singular, which would be reflected in their answer to the interpretation question.

3.4.3 Grammatical symmetry in RTs: Slowdown for mismatched subject NPs

As for RTs, the findings present a bit more of a complicated picture of the comprehension of agreement attraction and prosody's role. In particular, a slow down in the time-to-decision was observed for mismatching subject NPs ("The key to the cabinets") for both grammatical and ungrammatical sentences. While the final judgment of the participants did not reflect interference of the local noun in grammatical sentences ("The key to the cabinets is") the time that it took those participants to make that decision did indeed reflect some difficulty in processing. However, it is not clear how the mismatched subject NPs can slow down the time-to-decision without influencing the acceptability rate of these grammatical sentences. For ungrammatical sentences, the explanation for slowed responses is that interference of the local noun (whether that be through making the whole NP slightly plural or through competition with the head noun for the role of the source of agreement) results in a slower or longer agreement process, and ratings reflect that interference as well. Since this same reasoning cannot apply to grammatical sentences, another element must be at play.

3.4.4 Phrasing facilitates time-to-decision for acceptability

Despite a lack of evidence for an effect of phrasing on ratings, an effect was detected for the RT data. In particular, separate phrasing ("| The key to the cabinets | | are ... |") seemed to mitigate the slowdown from subject NPs with mismatched number. The clearest case of this is seen in Figure 3.12 when comparing agreement attraction sentences across phrasing and response type conditions. While the response time for acceptable (incorrect) responses was unaffected by phrasing, this was not the case for unacceptable (correct) responses. When the local noun was phrased with the verb, unacceptable responses took much longer. This comparison of time-to-decision across response type is needed because few previous works have investigated response time differences with whole sentences. Self-paced reading experiments generally have not looked at the RT for a final decision on acceptability. In production tasks such as Staub's (2009), we do not know what a

rejection/unacceptable response looks like when participants are producing a verb for a given subject NP. Presumably, every answer they give is an acceptable production. We can, however, compare the RTs for acceptable responses in these data to the results of Staub (2009).

In Staub's experiment, the RT for the selection of a grammatical or ungrammatical verb did not differ when the head noun was singular and local noun was plural. The selection of a singular verb would align with the acceptable/correct responses for the MISMATCH GRAMMATICAL condition, and the selection of a plural verb would align with the acceptable/incorrect responses for the MISMATCH UNGRAMMATICAL condition. In Figure 3.17, these four conditions have been separated out for easier visual inspection. The result of a pairwise comparison of the conditions within their phrasing conditions shows that while we cannot definitively conclude that the RTs for the comparisons differ, the size of the difference is larger for the DIFFERENT IPS condition (β = -0.9, CrI = [-0.19, 0.02]) than the difference in the SAME IP condition (β = -0.4, CrI = [-0.14, 0.06]) . The magnitude of these differences is similar to what Staub (2009) reports in their experiments as well.

As for an explanation for the slowdown in correct/unacceptable responses for agreement attraction sentences in the SAME IP condition, some number of participants could have had to reanalyze their interpretation sometime after encountering the verb, which resulted in a slow down just for the unacceptable responses. In the SAME IP condition, people who ultimately decide that the sentence is unacceptable at first believe that the local noun is the controller for agreement, but at some point between the verb and the time they press the button, reevaluate the agreement relationship. This sort of revision would take some time to accomplish and thus result in longer RTs. In the case of the DIFFERENT IPS condition, either a reevaluation process does not occur or is facilitated by the presence of the prosodic juncture between the local noun and verb. Since the need for reevaluation can only arise after the verb is encountered, it must be that the prosodic break has some lasting effect that helps downstream in the processing of the sentence. When considering Schafer's (1998) work on prosodic junctures, a possible explanation could be that the juncture signals to the processor to wrap up any ongoing processes, which then frees up

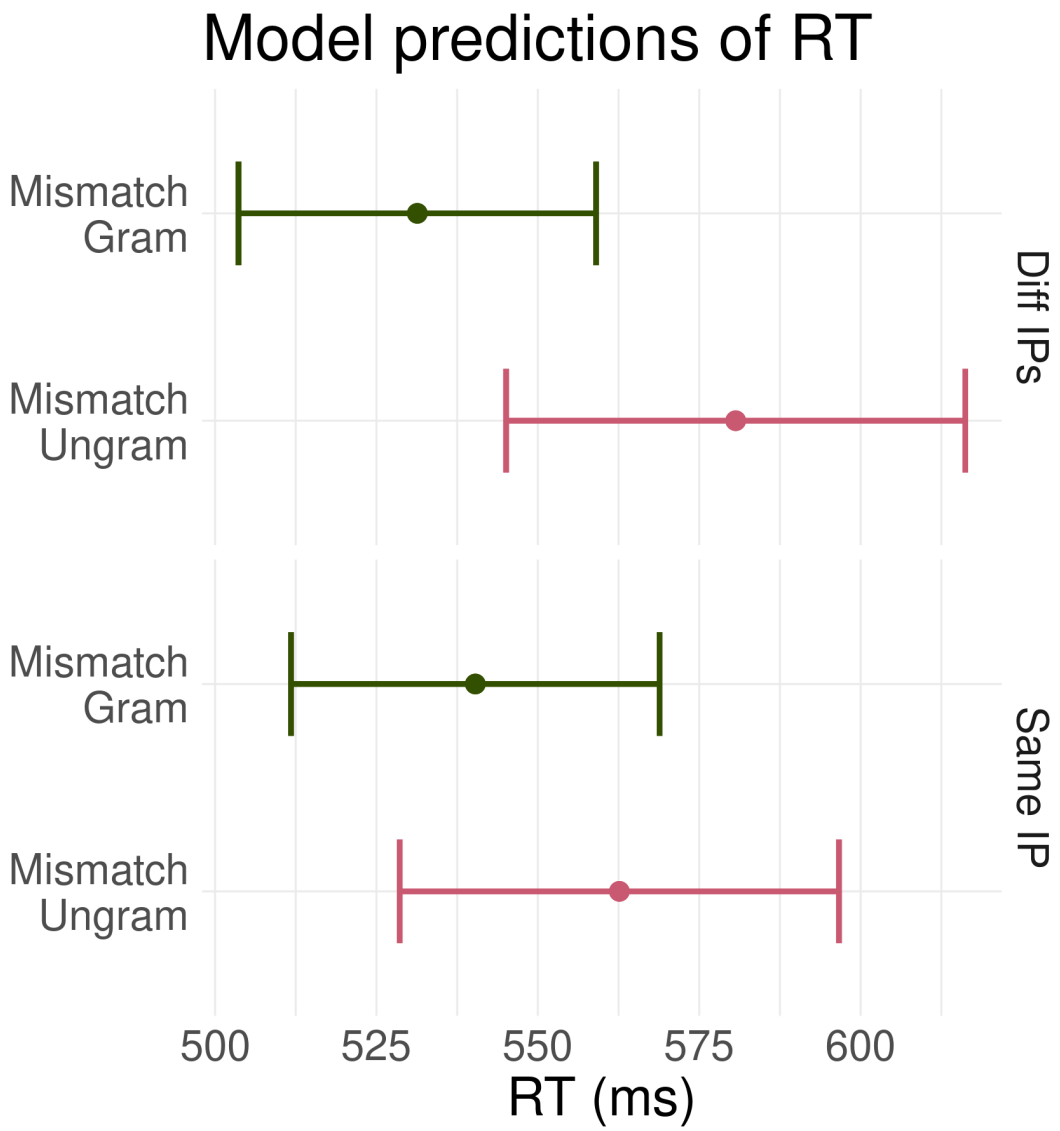


Figure 3.17: Samples from the posterior predictive distribution just for Mismatch conditions.

mental resources for other upcoming processes. That means that in the case of the DIFFERENT IPS condition, there is no slow down in time-to-decision for unacceptable/correct responses because the resources needed to reevaluate the agreement relationship are available. In the SAME IP condition, when the processor encounters the verb, it could still be processing the subject NP and thus have fewer resources for reevaluating the source of agreement as the sentence goes on. This explanation would be strengthened if the data here could bear on how speakers determined agreement mid-sentence because this could validate the idea that this reevaluation process is occurring.

Another explanation, however, is that phrasing reduced the facilitatory interference in the MISMATCH conditions. In a cue-based retrieval model like the one presented by Lewis and Vasishth (2005), when both nouns in the subject NP share the same cues (i.e., both nouns are singular, one is the subject, and the verb needs a singular subject noun) that the verb requires, then they compete for retrieval, causing a slowdown in retrieval. When those nouns both overlap with the cues required by the verb, but those overlapping cues are different (i.e., one noun is a singular subject, the other noun is plural, and the verb needs a plural subject noun), you get facilitatory interference because the cues pick our unique nouns. For correct/unacceptable responses to agreement attraction sentences, the SAME IP has much faster RTs than the DIFFERENT IPS condition. This difference could be a lack of facilitatory interference in the SAME IP condition. However, the problem with this account is that the acceptable/incorrect responses appear to be unaffected by phrasing. Both estimates for the time-to-decision for DIFFERENT IPS and SAME IP acceptable/incorrect responses are around 550 and 575ms respectively, whereas the unacceptable/correct responses' estimates are approximately 600 and 740ms, respectively. If facilitatory interference were at work in agreement attraction sentences in the DIFFERENT IPS condition, cue-based retrieval would predict that the acceptable responses might be faster than the acceptable responses in the SAME IP condition.

In the following chapter, the results of both experiments will be discussed together and how they bear on the prosodic hypotheses discussed previously.

CHAPTER 4

General Discussion

4.1 Overview of findings

To begin, let us review the main research questions outlined in Chapter 1 of this dissertation and answer them in light of the results of the experiments presented in Chapters 2 and 3.

4.1.1 What patterns of phrasing and accenting are typical for agreement attraction sentences? Does the intonation of agreement attraction sentences differ from non-attraction sentences?

From the ToBI analysis of the sentences produced by naive experimental participants, the phrasing and pitch accenting patterns used for agreement attraction sentences do not differ substantially from non-attraction sentences. In both kinds of sentences, the prevalence of intermediate and intonation phrase breaks after the local noun suggests that participants frequently project the same phrasing onto those sentences while reading. This phrasing falls in line with work by Pratt and Fernández (2019) which found that encouraging a phrasal "chunked" reading of attraction sentences facilitated the occurrence of agreement attraction in comprehension. Unfortunately, the limited number of attraction sentences collected from participants prevents us from drawing definite conclusions about the phrasing of attraction sentences versus non-attraction sentences.

As for the consistency in phrasing of the subject in its own prosodic phrase across conditions, a boundary at the edge of the subject NP would be expected given grammatical constraints on the length and phonological weight of ips and IPs (Selkirk, 1986; Truck-

enbrodt, 1995; Frazier et al., 2004). Utterances with many syllables are often broken up into smaller prosodic phrases, and this can often, but not always, align with syntactic constituent boundaries. A break after the end of the subject NP would align with the end of that syntactic constituent and often close a prosodic constituent that may be too phonologically heavy. The PP modifying the head noun in the critical sentences increases the phonological weight of the subject NP.

Another possible reason that influenced the phrasing is due to the way that the cuing stimuli were presented to participants. The sentence fragment recording may have primed participants to put the subject into its own phrase because that is the way they heard it. That is, each participant heard the critical subject NP by itself with no other material following, which in and of itself could be considered a felicitous utterance prosodically.

4.1.2 How does phrasing influence the acceptability of agreement attraction sentences?

The acceptability rating results replicated previous findings supporting the grammaticality asymmetry, as predicted by cue-based retrieval accounts of attraction. Sentences with a plural verb and plural local noun were rated much higher than those same sentences with a singular local noun, despite the head noun being singular in both cases. While there was little conclusive evidence for a three-way interaction, the interaction of local noun number and phrasing condition does present an interesting result that suggests a need for nuance in understanding prosodic effect. Pratt and Fernández (2019)'s work would make us predict that separate phrasing encourages the acceptability of agreement attraction sentences, but these data do not substantiate this. Phrasing the local noun and verb separately did seem to have a slight negative to null effect on ratings of conditions with plural nouns. These sentences with plural nouns showed a numerically lower overall rating in the DIFFERENT IPS condition compared to the SAME IP condition. With these data, we cannot decide whether or not this effect is null, but it does provide an interesting future avenue for research. Neither Marking and Morphing nor a cue-based retrieval account, as they

currently stand, can explain why plural local nouns might lower acceptability for both grammatical and ungrammatical sentences, but only when the local noun and verb are in different IPs.

4.1.3 How does phrasing affect the speed with which participants determine the source of agreement?

Phrasing did play a role in how fast participants judged sentences. In particular, phrasing seemed to have the biggest effect for UNGRAMMATICAL verbs RTs. While sentences like "The key to the cabinet are" are faster to judge when the local noun and verb are phrased together compared to when they are phrased separately, when the local noun is plural like in "The key to the cabinets are", participants are faster to judge the DIFFERENT IPS condition. The speed with which participants judged the agreement attraction sentences to be unacceptable when the local noun and verb were phrased together was much slower than when phrased separately. The change in the direction of these effects can be clarified when breaking down response type. In the SAME IP condition, unacceptable/correct responses are much slower than acceptable/incorrect responses. As for the DIFFERENT IPS condition, however, the RTs for both response types seem to be the same. Therefore, this effect is mostly being driven by the slowdown for unacceptable responses.

4.2 The Bigger Picture

The main prosodic hypotheses of this study presented two possible accounts for how phrasing could influence the processing of agreement attraction. According to the Accentual Processing Boost Hypothesis, words that carry the last, most prominent pitch accent of an intermediate phrase are more readily accessible in memory and thus exert a stronger influence on agreement processes. This proposal builds off of work showing that prominent pitch accents increase the likelihood of recall in serial recall tasks and generally contribute to the salience of carrier words. In this account, plural local nouns that carry a nuclear pitch accent (e.g. "cabinets" in ("The key to the **cabinets** are on the table"

) are more readily accessible or salient and thus cause a greater degree of interference in the agreement process than when they do not carry a nuclear pitch accent, as in the SAME IP condition. This hypothesis for the accent's influence generates the prediction that the acceptability rating of agreement attraction sentences increases in the DIFFERENT IPS condition relative to the SAME IP condition.

The other hypothesis, the Chunk-Internal Availability Hypothesis, framed phrasing's role as the facilitator of processing. In this hypothesis, syntactic dependencies are more reliably and quickly resolved when elements of that dependency are contained within the same IP. This hypothesis takes a much different view of how phrasing could influence the acceptability of agreement attraction. The main point is that plural local nouns in DIFFERENT IPS cause less interference with the agreement process than those in the SAME IP because phrases delimit processing chunks. The head noun does not undergo the same difficulty in retrieving because it is processed more deeply in its own IP and thus is easier to retrieve. About what predictions this makes for acceptability, a plural local noun would cause more interference in the SAME IP conditions resulting in increased acceptability and faster RTs. Partial evidence supporting Chunk-Internal Availability Hypothesis was found in the time-to-decision measures of participant's acceptability ratings. More work is needed, though, to understand how the phrasing could affect the time-to-decision, but not the ultimate decision the participant came to.

Based on the results of the comprehension experiment, in particular, both hypotheses have problems accurately characterizing the data. Firstly, the results of the rating data showed that there was little to no effect of phrasing on the influence of the acceptability of agreement attraction sentences. However, despite this lack of influence on ratings, phrasing was found to moderate the speed with which participants came to an "unacceptable" decision for agreement attraction sentences. This lack of an effect in the ratings but an effect in time-to-decision causes problems with current models of agreement attraction interference. Cue-based retrieval models, in particular, are not equipped to handle this lack of a rating effect because the interference of the local noun is thought to be what influences the decision on acceptability. That decision on the acceptability, in turn, must then

necessarily affect the speed with which participants make their decision. What this means for phrasing's role in attraction is that it may not necessarily have a direct relationship with the identification of the source of agreement or the carrying out of the agreement mechanism. However, it may facilitate other processes in language processing that then, in turn, affect time-to-decision measures. One example of this might be that the presence of the prosodic phrase break after the local noun signals to the processor to wrap up any ongoing processes. With those ongoing processes taken care of, more resources are available for the agreement mechanism to carry out its job in a fast and accurate way. When the local noun and verb are in the same IP, however, those ongoing processes of encoding the subject NP and parsing the structure would leave the processor with fewer resources for agreement processing, which may delay the decision-making process. What would help to investigate this phenomenon further is a measure of online processing difficulty. Pupillometry would be a suitable candidate methodology for getting an online measure of processing difficulty as the participant listens to the sentence. This measure could help distinguish the difficulty of processing agreement once the verb is encountered. It could also tell us about the nature of the prosodic breaks' facilitatory effect on processing.

4.2.1 Representation matters

We do not find unequivocal support for either of the agreement attraction theories in this study. In support of *Marking and Morphing*, two findings from the production results seem to indicate that the locus of problems in agreement is in the representation of the noun phrase. Firstly, as has been previously found, the rate of plural verb production when a local noun carries plural morphology is relatively low but non-zero. This distribution of agreement attraction errors lends itself to a probabilistic account of attraction production, put forward by Eberhard et al. (2005). This account stipulates that plurality is not binary (e.g. [+SINGULAR] vs. [-SINGULAR] or [-PLURAL] vs. [+PLURAL]) or privative (i.e., Ø vs. [PLURAL]), but rather a gradient value ranging from -1 (unambiguously plural) to 1 (unambiguously singular). The plural morphology on the local noun in an NP like "The key to the cabinets" makes the whole NPs' plurality value slightly plural, which

thus results in cases in which the entire NP is grammatically treated as plural. This slightly plural value is directly related to the probability of a plural verb being selected to complete a sentence with said NP. Therefore, a slightly plural NP results in a small degree of plural verb production. In this respect, the production results replicated this very gradience in responses.

In terms of results that counter the predictions of *Marking and Morphing*, there was not a detectable effect of attraction in grammatical sentences. The only detectable effect of the local noun influencing the resolution of agreement was in the ungrammatical sentences, as predicted by a cue-based retrieval account of attraction. As previously mentioned, though, this grammaticality asymmetry is predicted by Wagers et al. (2009), among others.

4.3 Future directions

4.3.1 Incorporation of prosody into computational models

While the *Marking and Morphing* model by Bock et al. (2001) and cue-based retrieval ACT-R model proposed by Lewis and Vasishth (2005) were mentioned in Chapter 1, they were not utilized in the modeling of the responses by participants. Thus, one clear avenue for future research would be to account for this study's results in both models and understand where they fail. These failures would be potential places to improve the models and generally grow our understanding of the agreement process.

4.3.2 Variety of syntactic configurations

The main goal in the design of the stimuli was to make the likelihood of agreement attraction as high as possible. Since reported rates of attraction in production are typically in the single digits or a bit higher, avoiding a floor effect of attraction was a major concern. Therefore, the PP-modifier was chosen to carry the local noun since it has been found to engender the highest rates of attraction. However, one of the drawbacks is that the ease of processing the sentences with PP-modifiers may have meant that processing was not hard

enough to see a strong effect of prosody. In the previously mentioned study by Pratt and Fernández (2019), the local noun appeared in a relative clause. The relative clause was either structurally simple (Example 20) or structurally complex (Example 21).

(20) The reporter [who called the senators every so often] writes awful stories for the newspaper.

(21) The reporter [who called the senators that Scott supported] writes awful stories for the newspaper.

What they found was that the complex sentences reduced comprehension accuracy. However, an interaction with the presentation method that helped the reader segment the sentence into phrases mitigated the cost of that complexity. In the current study, the relative ease of parsing the sentences may have meant that intonational information was not necessary or unhelpful in processing them. By increasing the difficulty of the sentences, phrasing may end up being recruited to help participants even more than was observed in the comprehension task.

Another well-known effect of attraction is that of syntactic depth, as mentioned in Chapter 1. One way of better understanding the effect of phrasing could be to test sentences with two PPs such as "The key to the cabinets by the door" and manipulate the grouping of the lower noun with the verb and the morphological number of the nouns. In such an experiment, the phrasing manipulation would need three levels: The higher and lower local nouns are phrased together with the verb, only the lower noun is phrased with the verb, and neither local noun is phrased with the verb. Since attraction is much weaker for the lower noun than the higher noun, phrasing could increase the interference of the lower local noun when phrased with the verb and weaken the interference from the higher local noun.

4.4 Conclusions

4.4.1 What have we seen, and what does it all mean?

The influence of phrasing on the processing of agreement attraction sentences seems to be limited in nature. In this study, phrasing was found to facilitate time-to-decision for sentences where a plural local noun and verb were in separate IPs, but only when the participant ultimately judged the attraction sentence to be unacceptable. While this attraction effect did show up in the time-to-decision for acceptability, the actual acceptability ratings themselves were unaffected by phrasing. We use this as evidence that phrasing itself may not play a direct role in the agreement mechanism but rather a more general role in facilitating processing.

Bibliography

Anderson, J. R., Matessa, M., and Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462.

Baars, B. J. and Mackay, D. G. (1978). Experimentally eliciting phonetic and sentential speech errors: Methods, implications, and work in progress. *Language in Society*, 7(1):105–109.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11(1-2):17–67.

Beckman, M. E. and Ayers, G. M. (1997). *Guidelines for ToBI labeling*. Ohio State University. manuscript.

Beckman, M. E. and Hirschberg, J. (1994). The tobi annotation conventions. *Ohio State University*.

Beckman, M. E. and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, 3(01):255–309.

Bishop, J. (2020). Exploring the similarity between implicit and explicit prosody: Prosodic phrasing and individual differences. *Language and speech*, page 23830920972732.

Bock, J. K. and Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, 11(1):64–76.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., and Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive psychology*, 43(2):83–128.

Bock, K. and Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, 23(1):45–93.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2):37–50.

Bürkner, P.-C. (2017a). Advanced bayesian multilevel modeling with the r package brms. *arXiv preprint arXiv:1705.11123*.

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

Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2):149–180.

Bürkner, P. (2021). Parameterization of response distributions in brms.

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

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411.

Bürkner, P.-C. (2020). Bayesian item response modeling in r with brms and stan.

Carlson, K. (2009). How prosody influences sentence comprehension. *Language and Linguistics Compass*, 3(5):1188–1200.

Carlson, K. and Tyler, J. C. (2018). Accents, not just prosodic boundaries, influence syntactic attachment. *Language and speech*, 61(2):246–276.

Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, 40(2):141–201.

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

Dixon, W. J. and Tukey, J. W. (1968). Approximate behavior of the distribution of winsorized t (trimming/winsorization 2). *Technometrics*, 10(1):83–98.

do Carmo Lourenço-Gomes, M. (2017). Psycholinguistics is definitely tied up to prosody. *Journal of Speech Sciences*, 5(2):87–100.

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

E.Turk, A. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in american english words. *Journal of Phonetics*, 35(4):445–472.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., and König, P. (2017). Labvanced: a unified javascript framework for online studies. In *2017 International Conference on Computational Social Science IC2S2 (Cologne)*.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27(2):285–319.

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

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

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2010). Recognition memory reveals just how contrastive contrastive accenting really is. *Journal of Memory and Language*, 63(3):367–386.

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and aging*, 27(1):88–98.

Frazier, L., Carlson, K., and Clifton, Charles, J. (2006). Prosodic phrasing is central to language comprehension. *Trends in cognitive sciences*, 10(6):244–249.

Frazier, L., Clifton, C., and Carlson, K. (2004). Don't break, or do: prosodic boundary preferences. *Lingua*, 114(1):3–27.

Frazier, L., Clifton, Charles, J., Carlson, K., and Harris, J. A. (2014). Standing alone with prosodic help. *Language, Cognition and Neuroscience*, 29(4):459–469.

Gee, J. P. and Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, 15(4):411–458.

Greene, R. D. (2010). *Language, Ideology and Identity in Rural Eastern Kentucky*. PhD thesis, Stanford University.

Guest, O. and Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, page 1745691620970585. PMID: 33482070.

Hammerly, C. and Dillon, B. (2017). Restricting domains of retrieval: Evidence for clause-bound processing from agreement attraction. In *30th CUNY Sentence Processing Conference*.

Harris, J. and Jun, S. (2019). Using pupillometry to assess prosodic alignment in language comprehension. In *Proceedings of he International Congress of Phonetic Sciences (ICPhS) 2019*.

Hazen, K. (2014). A new role for an ancient variable in appalachia: Paradigm leveling and standardization in west virginia. *Language Variation and Change*, 26(1):77–102.

Hessick, C. B. (2021). [@cbhessick] when the grammar police get it wrong.

Hirotani, M., Frazier, L., and Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54(3):425–443.

Ito, K. and Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. *Methods in empirical prosody research*, pages 229–257.

Jun, S.-A. and Bishop, J. (2015). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and speech*, 58(4):459–473.

Kjelgaard, M. M. and Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40(2):153–194.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, 3(3):129–140.

Klatt, D. H. (1976). Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221.

Kreiner, H. and Koriat, A. (2005). The role of reading prosody in syntactic and semantic integration: evidence from eye movements. In *Proceedings of the International Symposium on Discourse and Prosody as a Complex Interface*, pages 1–17.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2):270–280.

Labov, W., Ash, S., and Boberg, C. (2008). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Lee, E. K. (2012). *Prosody and syntactic processing*. PhD thesis, University of Illinois at Urbana-Champaign.

Lee, E.-K. and Watson, D. G. (2011). Effects of pitch accents in attachment ambiguity resolution. *Language and cognitive processes*, 26(2):262–297.

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.6.2-1.

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

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal forced aligner: trainable text-speech alignment using kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3):817–835.

Nespor, M. and Vogel, I. (2007). *Prosodic phonology*, volume 28. Walter de Gruyter.

Nespor, M. and Vogel, I. (2012). *Prosodic Phonology*. De Gruyter Mouton.

Palan, S. and Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27.

Parker, D., Shvartsman, M., and Van Dyke, J. A. (2018). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. *Language Processing and Disorders*, page 121.

Patson, N. D. and Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, 69(5):950–971. PMID: 25624171.

Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, volume 271, page 311. MIT Press.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, Massachusetts Institute of Technology. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2020-06-24.

Pratt, E. and Fernández, E. M. (2019). Implicit prosody and cue-based retrieval: L1 and l2 agreement and comprehension during reading. *Frontiers in Psychology*, 7:17.

Prolific (2021). *Prolific: Revolutionizing people research*.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Ratcliff, R. and Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological review*, 83(3):190–214.

Sanford, A. J., Price, J., and Sanford, A. J. (2009). Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition*, 37(6):880–888.

Sanford, A. J. and Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6(9):382–386.

Schafer, A. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, 11(1-2):135–164.

Schafer, A. and Jun, S. (2002). *Effects of accentual phrasing on adjective interpretation in Korean*, pages 223–255. Sentence Processing in East Asian Languages. Stanford, CSLI.

Schafer, A. J. (1998). *Prosodic Parsing: The Role of Prosody in Sentence Comprehension*. PhD thesis, UMass Amherst.

Selkirk, E. O. (1986). *PHONOLOGY AND SYNTAX: THE RELATION BETWEEN SOUND AND STRUCTURE*. Cambridge, MA: The MIT Press. Date revised - 2003-10-01; SuppNotes - Edition date: 1986; Last updated - 2016-09-27.

Shattuck-Hufnagel, S. and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, 25(2):193. Last updated - 2013-02-23.

Simpson, H. E. (2017). *The role of intonation units in memory for spoken English*. PhD thesis, University of California, Santa Barbara.

Stan Development Team (2020). Stan modeling language users guide and reference manual, version 2.26.

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

Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136(1):64–81.

Truckenbrodt, H. (1995). *Phonological phrases: Their relation to syntax, focus, and prominence*. PhD thesis, MIT.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, 30(2):219–255.

UCLA Statistical Consulting Group (2011). *R Library Contrast Coding Systems for Categorical Variables @ONLINE*.

van Zwet, E. and Cator, E. (2021). The significance filter, the winner's curse and the need to shrink. *Statistica Neerlandica*.

Vasishth, S., Mertzen, D., Jäger, L. A., and Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, 103:151–175.

Vigliocco, G. (1996). One or more labels on the bottles? notional concord in dutch and french. *Language and Cognitive Processes*, 11(4):407–442.

Wagenmakers, E.-J. and Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological review*, 114(3):830–841.

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

Wagner, M. and Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7):905–945.

Watson, D., Arnold, J., and Tanenhaus, M. K. (2006). Acoustic prominence and reference accessibility in language production. In *Speech Prosody 2006*.

APPENDIX A

Appendix: Stimuli

Critical Items

1. The actor in the film(s) was/were popular with both young and old fans
2. The apartment with the leak(s) is/are too much for one repairman to handle
3. The argument over the plate(s) was/were presented by the leader of the group
4. The composer of the opera(s) was/were typically regarded as masterful
5. The confession of the criminal(s) was/were coerced by the shady police officers
6. The consultant for the firm(s) is/are preparing the offer for the potential buyer
7. The defect in the car(s) was/were pricey to fix before being sold
8. The dream about the castle(s) was/were totally forgotten by the king and queen
9. The entrance to the laborator(y/ies) is/are tough to find in the Psychology building
10. The garden with the spike(s) is/are cleverly hidden behind the tall shrubs
11. The key to the cabinet(s) was/were placed by the entrance to the storage room
12. The label on the bottle(s) is/are tiny compared to previous product designs
13. The lab with the computer(s) is/are personally owned by the boss of the company
14. The letter from the lawyer(s) was/were taken to the jail house after trial
15. The map of the creek(s) was/were treated with care by the park ranger

16. The mistake in the program(s) was/were costing the company billions of dollars
17. The name on the billboard(s) is/are potentially visible from the highway
18. The pamphlet from the agenc(y/ies) was/were terribly uninformative to consumers
19. The photo of the politician(s) was/were kindly received by the general public
20. The picture on the postcard(s) was/were created by the skillful artist
21. The purpose of the kit(s) is/are to help wounded hikers mend small cuts
22. The reason for the fort(s) is/are totally unknown to the travelers
23. The record from the singer(s) is/are prominent in the world of pop music
24. The ruling about the lot(s) was/were probably ignored by the land owners
25. The scale of the map(s) was/were posing a problem for the cartographer
26. The slogan on the poster(s) was/were picked by city council for the parade
27. The smell of the sock(s) was/were completely removed by the mother
28. The star of the musical(s) was/were praised by the critics in the newspaper
29. The story behind the myth(s) was/were told to generations of children
30. The typo in the book(s) has/have been changed thanks to the editor
31. The presentation from the accountant(s) has/have boosted the confidence of the shareholders
32. The manager of the archive(s) has/have been featured in the documentary
33. The leader of the gang(s) was/were being followed by the FBI for months on end
34. The advisor of the student(s) has/have to leave campus later this afternoon
35. The soldier by the tank(s) was/were behind enemy lines for two days

36. The assistant for the lab(s) is/are being supported by university funding

37. The painting of the mountain(s) is/are a breathtaking sight to behold

38. The office with the photocopier(s) has/have been inspected several times

39. The announcement about the game(s) is/are being broadcast on ESPN

40. The proposal for the building(s) has/have been reviewed by the architects

41. The advertisement for the store(s) was/were marketed towards millennials

42. The contract for the employee(s) is/are protected by the trade union's lawyer

43. The bill from the contractor(s) has/have caused a lot of discussion by the investors

44. The program for the concert(s) was/were already the talk of the town

45. The director of the movie(s) is/are famous for their action sequences

46. The opening to the tunnel(s) has/have been widened to allow the new machinery to enter

47. The song in the commercial(s) has/have been uploaded to YouTube recently

48. The door behind the curtain(s) was/were mentioned in the home decor magazine

Fillers Items

49. Mary and the rest of the cooking club % were making dinner.

50. The boots and sneakers % were supposed to be on the same shelf.

51. The poster and pamphlet % both say that children should sleep more.

52. Sarah's mom and dad % have matching tennis outfits for the tournament.

53. The laundromat asks that slacks and jeans % are in separate bags.

54. The pianist thinks the keys and pedals need replacing on the organ

55. The bus driver and teacher % eats lunch together every week.

56. Leo and Mark thinks they parked the car in Lot B % by the mall

57. The doctor guarantees that both infants and adults % is insured.

58. Trina and the other dancers % is planning to go to Italy together.

59. Red and green % is Bobby's favorite colors to paint landscapes with.

60. The teas and coffees that the local cafes offer is outstanding

61. The costume designer and director watches the footage of the play.

62. Their cousin might could be the one who left them the treasure map.

63. The play might could be canceled because the power keeps going out.

64. The three horses were startled % by the snake crawling towards them

65. Everyone should check out these three easy life hacks.

66. Some of the trees fell over during the storm last year.

67. Lucy thought that owning thirty pairs of socks was too much.

68. The display case was empty except for the last three muffins.

69. The robbers got away with ten televisions and a speaker system.

70. Mark ran so many marathons that he got a sponsorship for shoes.

71. Kat owned so many pair of jeans % that she bought a new dresser.

72. The breakfast catering company set up ten griddle station.

73. All of the sink were being used to fill up water balloons.

74. Each floor of the building had five suite and a break room.

75. Penny and her family % don't never get no sleep when her upstairs neighbors throw a party.

76. Mary never got none of Deb's texts % so she didn't know the concert was canceled.

77. Julia has so much free time % that she reads in the park anymore.

78. My grandma is always sitting in her living room knitting anymore.

79. The magician will try and cast a spell over the man to make him disappear.

80. Many birds gather at the bench % to fight over bread crumbs.

81. The meeting was delayed % due to technical issues with the speakers.

82. Everyone should read that self help book about saving money.

83. The lawyer wins all of the cases where arson is involved.

84. The lion that bit the zoo keeper % stay in a different cage now

85. The chef's favorite dish % are from his rival restaurant across town.

86. The baker made the cake that all the customers fights over

87. The voice actor is finished his voice-overs for this week's episode

88. The company welcomed the new interns % by giving them gift cards.

89. The Canadian diplomat asked his assistant to make up an excuse for him to leave the meeting early.

90. My tweet was retweeted by the actress I love and she even followed me.

91. Franklin wished he didn't have to come into work on Saturday mornings.

92. Tell Nancy to call the credit card company when she wakes up.

93. The power line that was knocked down by the storm caused a fire.

94. Last summer I visit Sweden with my brother for two weeks

95. The museum exchanges all of their exhibits for new ones last year.

96. The crossing guard talk with the students and teachers on the corner