2019-06-01

Using Eye Tracking to Examine Working Memory and Verbal Feature Processing in Spanish

Erik William Arnold
Brigham Young University

Using Eye Tracking to Examine Working Memory and Verbal Feature Processing in Spanish

Erik William Arnold

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Deryle W. Lonsdale, Chair
Steven G. Luke
Jeffrey R. Parker

Department of Linguistics

Brigham Young University

ABSTRACT

Using Eye Tracking to Examine Working Memory and Verbal Feature Processing in Spanish

Erik William Arnold
Department of Linguistics, BYU
Master of Arts

Second language acquisition (SLA) has been a dominant field in linguistics research over the past several decades. In this field, researchers have investigated what makes learning the grammar of a second language difficult and have identified many factors that may contribute to this difficulty, including agreement processing. In linguistic terms, agreement refers to the necessary covariation of grammatical features between two or more syntactic constituents. In early years, researchers examined how native speakers process varying grammatical features (e.g. number and gender) in agreement relations. In recent years, however, they have turned towards L2 learners and have investigated whether learners can attain native-like levels of processing agreement in a second language.

While some studies have investigated differences between learners and native speakers, other studies have examined the effect of individual differences on agreement processing. Of particular interest to this thesis is working memory capacity (WMC) and its effect during the different processing stages of agreement. Lastly, features expressed through agreement may affect individuals' processing behavior. Different features (e.g. person, number, gender) are regularly expressed in agreement relations by different manifestations of exponence. Many authors have investigated the effect different features have on processing agreement when those features are expressed by separative exponence. Fewer have examined the effect of cumulative exponence on agreement processing.

Eye tracking is a useful psycholinguistic tool to investigate these questions. Using eye tracking, I examine English learners of Spanish and their eye behavior as they processed Spanish verbal agreement and investigate whether they demonstrate sensitivity similar to native Spanish speakers while processing verbal agreement errors. I investigate if individuals demonstrate similar sensitivity when processing three different types of verbal agreement errors—number, person, and tense. Additionally, I examine whether individuals' sensitivity to agreement errors is affected by working memory capacity. Using a linear mixed effects model, I analyze the eye tracking data and share the results of the analyses and their implications for L2 research in agreement processing.

Keywords: eye tracking, agreement processing, second language acquisition, working memory capacity, Spanish

ACKNOWLEDGMENTS

I thank Dr. Deryle Lonsdale for his great support. He has instructed me in the majority of my classes and has been a true mentor during this entire process. I thank Dr. Steven Luke for his training and support in using the eye tracker, building the experiment, and analyzing the data. I thank Dr. Jeff Parker for his help in developing my research questions and research design as well as providing valuable insight into data analysis. I also thank all the students who were willing to come and participate in my experiment. Lastly, I thank my family for their support in every capacity. I am especially grateful for my wife who is my fiercest supporter, loyalest friend, and greatest companion. She has stood by my side through everything and I could not have done it without her.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Interest in eye tracking methodology has grown among linguistic researchers over the past several years. Eye tracking gathers eye behavior data by recording individuals' eye movements as they carry out some visual task (e.g. in linguistics, reading a sentence or paragraph). Similar to other psycholinguistic methodologies (e.g. self-paced reading and lexical decision task), eye tracking provides researchers with the ability to capture individual responses to online processing tasks. One of the more complex linguistic tasks individuals carry out is reading comprehension. Like self-paced reading, eye tracking is useful in gathering data from individuals during the time course of reading. However, its advantage over self-paced reading becomes clear when considering the information available from recording eye movements.

Individuals' eye movements provide detailed information regarding processing of linguistic material during reading. The breadth of variables from which to gather this information can be overwhelming. However, this large selection allows researchers to capture individuals' behavior during the entire time course of reading including early, late, and very late processing. Such an advantage gives further insight into lexical access, lexical information analysis and reanalysis, and repair costs (Clifton et al., 2007). Eye tracking is therefore a favorable tool to use when capturing second language (L2) learners' behavior during reading.

Prior investigations in second language acquisition (SLA) have examined a variety of linguistic phenomena from all linguistic domains (e.g. phonology, syntax/morphology, semantics) among adult learners. Of particular interest for many researchers, including the topic of this thesis, is the extent to which adult L2 learners demonstrate integrated knowledge of the grammar of their second language. What is more interesting is how learners show knowledge of the interface of morphology/syntax during processing. Such morphosyntactic knowledge incorporates important grammatical structures like agreement, a one-way dependent relation between two or more syntactic constituents. Even native speakers sometimes make errors while processing agreement in their first language, which has driven researchers to examine to what degree language learners can achieve native-like processing in a second language (Staub, 2009).

Several studies examining this question have found that some learners can demonstrate behavior (e.g. sensitivity to agreement errors) similar to natives while processing L2 agreement (Foote, 2011; Sagarra and Herschensohn, 2010); although, most only examined certain features expressed primarily through nominal agreement (e.g. number and gender). These studies have observed similar sensitivity exhibited by native and L2 learners to different features in agreement anomalies. Other studies, however, have found that learners cannot exhibit native-like processing behavior. Their claims are predominantly supported by the observation of qualitative differences (i.e. disparate underlying processing mechanisms) and quantitative mechanisms (i.e. limited cognitive resources) between L1 and L2 processing (Clahsen and Felser, 2006b,a; Jiang, 2004; Jiang et al., 2011). While researchers are continuing to explore differences between natives and learners regarding agreement processing, others have also examined differences between proficiency and university class level class level among L2 learners. Some studies have found measurable differences between beginning and advanced L2 learners regarding agreement processing. Whereas advanced learners have demonstrated sensitivity to agreement anomalies, beginning learners have not (Jiang et al., 2011; Keating, 2009; Lim and Christianson, 2015). However, other studies have suggested minimal differences between advanced and beginning learners as both groups have exhibited similar sensitivity to agreement errors (Foote, 2011; Sagarra and Herschensohn, 2010). With such disparate results from these studies it is clear this area of research warrants further investigation.

In addition to L1/L2 differences and class level differences, researchers have also examined the effect of individual differences on agreement processing. One particular individual difference that has garnered much attention is working memory capacity (WMC). Some studies have found that individuals' WMC has a measurable effect on agreement processing, whereas others have found individuals' WMC has limited or no effect (Coughlin and Tremblay, 2013; Havik et al., 2009; Sagarra and Herschensohn, 2010). Again, this area requires further research, particularly into the effect of WMC during the time course of agreement processing (i.e. early and late stages of processing, which will be discussed later). As a psycholinguistic tool, eye tracking provides researchers with detailed information of individuals' processing of morphosyntactic information in early and late stages.

Additionally, eye tracking methodology can be used to examine differences between the type of features present in agreement processing. Previous studies have indicated certain features may elicit different behavioral responses, or sensitivity, from individuals who are processing agreement. These studies have largely examined differences between features expressed by separative exponence and more salient (i.e. transparent and cross-linguistically common) features in nominal agreement (e.g. number and gender) (Barber and Carreiras, 2005; Romanova and Gor, 2017). However, cumulative exponence and less salient features in verbal agreement (e.g. person and tense), while common cross-linguistically, have received little attention. Few studies have investigated cumulative exponence and person agreement (see Mancini et al., 2011, 2014). This thesis seeks to address this gap in agreement processing research.

Because of its functionality and versatility, eye tracking is a useful tool to investigate the questions introduced above. In this thesis, I employ eye tracking to examine English learners of Spanish. I investigate differences between L2 learners and native speakers, among class levels, and in individuals' working memory capacity. Additionally, I use eye tracking to examine differences between Spanish verbal features expressed by cumulative exponence. I analyze the eye tracking data in a linear mixed effects model and interpret the results. Before introducing the design and results of this thesis, I explore agreement processing generally and the previous work in this field of research.

Chapter 2

Literature Review

Learning the grammar of a new language is difficult. This is particularly true for most second language (L2) learners despite having already accomplished this integral part of language learning in their first language (L1). This process in language learning is necessary not only for devoted language-learning students but even the most casual learners. In part, the grammar of a language represents a set of domains each comprised of combinatorial rules that specify some hierarchical structure of how words and clauses should be ordered. Such a complex representation allows a language to work and function to some extent with regulation. Language learners can then follow basic guidelines to have successful communication and exchange of information. It is therefore essential, even at the most rudimentary level, for individuals to learn the grammar of their second language.

Despite it being essential, accurately acquiring and precisely executing the grammar of an L2 in production and comprehension can be distressing for learners. A particular aspect of grammar that several researchers have noted L2 adult learners struggle to acquire is the morphology (i.e. word structure) of a language (Bosch and Clahsen, 2016; Clahsen and Felser, 2006b; Jiang, 2004; Keating, 2009; Sagarra and Herschensohn, 2010). Unfortunately, understanding what makes learning a language (particularly its morphology) difficult is itself a difficult question to answer. Prior studies have indicated a variety of factors that may account for learners' struggle with L2 morphological acquisition including age (Birdsong, 2006; DeKeyser and Larson-Hall, 2005), low L2 performance (Hopp, 2007, 2010), low L2 competence (Jiang, 2004, 2007), limited cognitive resources (Juffs, 2015; McDonald, 2006), and L1 influence (Bond et al., 2011; Gabriele et al., 2013).

Apart from the individual differences noted above and their effect on language learning, characteristics of the words themselves are contributing factors that make L2 acquisition more difficult. In particular, inflectional morphology presents a challenge to learners. Even advanced level, highly proficient L2 speakers have difficulty processing inflectional morphology correctly when presented with tasks such as reading comprehension (Bosch and Clahsen, 2016; Keating, 2009; Lopez Prego, 2012; Sagarra and Herschensohn, 2010). Part of the difficulty with inflectional morphology is that words may take varying forms and exhibit different types of grammatical information necessary for proper language comprehension. In most cases, L2 learners must obtain explicit knowledge of the different word forms and integrate that knowledge among other linguistic sources (e.g. phonology, syntax, and semantics). It is then especially taxing for learners to make this knowledge readily available and access it during linguistic tasks.

Some of the initial and continuing difficulties learners have with L2 inflectional morphology is acquiring a knowledge and integration of word paradigms, available for use in language comprehension. To acquire these word paradigms, learners often are not only required to acquire a single word form but a whole family of words related in form and meaning. Depending on the language, these word paradigms may be complex or simple, irregular or regular, highly ambiguous or not. A sample word paradigm is given for the Spanish verb *tomar* ('to take') in Table 2.1. The word paradigm for *tomar* exhibits the different grammatical information typically expressed by Spanish verbs. This includes features such as person, number, tense, aspect, and mood. These are the features even beginning learners are expected to have knowledge of and to integrate in tasks such as reading comprehension.

A challenging aspect for learners in integrating knowledge of word paradigms is that the grammatical information on words is often syncretic, demonstrating ambiguity of inflected form (Pickering, 1999). This ambiguity is demonstrated in Table 2.1 by the

Table 2.1: Spanish Word Paradigm for *tomar*

<table>
<tr><th></th><th>present indicative</th><th>present subjunctive</th><th>imperfect indicative</th><th>imperfect subjunctive</th><th>preterite</th><th>future</th></tr>
<tr><td>2SG</td><td>tomas</td><td>tomes</td><td>tomabas</td><td>tomases</td><td>tomaste</td><td>tomarás</td></tr>
<tr><td>1SG</td><td>tomo</td><td rowspan="2">tome</td><td rowspan="2">tomaba</td><td rowspan="2">tomase</td><td>tomé</td><td>tomaré</td></tr>
<tr><td>3SG</td><td>toma</td><td>tomó</td><td>tomará</td></tr>
<tr><td>1PL</td><td>tomamos</td><td>tomemos</td><td>tomábamos</td><td>tomásemos</td><td>tomamos</td><td>tomaremos</td></tr>
<tr><td>2PL</td><td>tomáis</td><td>tomeis</td><td>tomabais</td><td>tomaseis</td><td>tomasteis</td><td>tomaréis</td></tr>
<tr><td>3PL</td><td>toman</td><td>tomen</td><td>tomaban</td><td>tomasen</td><td>tomaron</td><td>tomarán</td></tr>
</table>

Note. This table was adapted from the Surrey Person Syncretism Database[1].

syncretic forms *tome* and *tomaba*. Both forms exhibit first singular and third singular for the present subjunctive and imperfect indicative respectively. Not only do some word forms exhibit syncretism, which creates difficult ambiguities for learners, but the types of grammatical information expressed by word forms are important for showing agreement between constituents in a sentence adding another complextity for learners acquiring L2 morphology. Perhaps the most challenging task for L2 learners is correctly processing agreement between sentence constituents.

Agreement Processing

Agreement in linguistics refers to a type of syntactic relation between two or more constituents within a given domain (the agreement domain). Agreement is important to identify and differentiate distinct grammatical relations in a language. The agreement relation is described as the covariance of a feature or features with certain constituents acting as controllers and other constituents acting as targets. Controller words determine the feature(s) that must covary, and the target words adhere to the feature specifications determined by the controllers (Haspelmath and Sims, 2010; Corbett, 1998). For example, the form of the Spanish adjective *corto* 'short' changes to reflect concordant gender agreement when its

[1] See `http://dx.doi.org/10.15126/SMG.10/2`

controlling noun expresses different grammatical genders (*examen corto* 'short test' or *falda corta* 'short skirt'). It is important to note that this description of 'controller' and 'target' does not present itself regularly in every language (i.e. consider pro-drop or underspecification in certain languages). However, for the purposes of this thesis and the explanation and presentation of morphosyntactic agreement in this thesis, I use the terms 'controller' and 'target'.

The complexity of morphosyntactic agreement varies widely cross-linguistically. Some languages have relatively simple systems of inflectional morphology and exhibit limited agreement relations, whereas other languages exhibit profoundly complex agreement relations. Even among languages that employ some rudimentary system of agreement rules, the manner in which agreement is regulated varies. For example, the lexical categories involved in agreement relations may vary, as well as the distance between constituents in agreement relations. The grammatical features that occur in agreement relations vary as well as the manner of exponence, or phonological representation in which those features are expressed (e.g. separative, simultaneous, or cumulative). Agreement manifests itself differently across languages.

Spanish Agreement

Compared to English, Spanish exhibits a relatively rich system of inflectional morphology and morphosyntactic agreement. Several lexical categories in varying controller and target word pairs exhibit multiple combinations of features in agreement that are not expressed in English. Additionally, these agreement relations regularly involve constructions of verbal exponence less pervasive in English (e.g. cumulative exponence). Verbs perhaps represent the most complex agreement relations in Spanish. As previously shown in the word paradigm for the verb *tomar*, Spanish verbs express grammatical features of person, number, tense, aspect, and mood.

Table 2.2: Spanish Verbal and Nominal Grammatical Features

	Person	Number	Tense	Aspect	Mood
Verbal	First Second Third	Singular Plural	Past Present Future	Perfect Perfective Imperfective	Indicative Subjunctive Imperative
	Number			Gender	
Nominal	Singular Plural			Masculine Feminine	

Each of the values expressed by these verbal features is shown in Table 2.2. Different verbal suffixes express each of the different combinations of these features. Spanish nouns, adjectives, and determiners also participate in agreement relations. These categories are typically marked for number and gender features. The values of number for these word classes are the same as those for verbs (i.e. singular and plural). The gender values are also binary, being masculine or feminine. Whereas number creates a relatively simple semantically determined marking system in Spanish, gender represents a more artibrary marking system, particularly for nouns. With the exception of a few animate nouns that are marked according to biological gender, the vast majority of nouns are classified arbitrarily as masculine or feminine. This arbitrary manner of marking gender, and the fact that gender is marked as a stem inherent feature on nouns adds another level of processing difficulty for learners.

Agreement relations commonly appear between nominal constituents with their determiners and articles and modifiers, and verbal constituents with their arguments and modifiers. Within these relations, three salient and productive features are commonly used: number, gender, and person. I will briefly discuss each of these prominent features and the variety of ways they are expressed in Spanish.

Number agreement. Number is perhaps the most salient feature marked in Spanish, particularly in cases of nominal agreement. Whereas the expression of number in verbal agreement can be opaque, in most cases of nominal agreement, number is marked transparently by *-s*, similar to English's marking of plural number. The agreement relation of grammatical number frequently appears as varying constituent pairs. In English and Spanish, grammatical number must agree between determiners and nouns.

Table 2.3: English and Spanish Nominal Agreement

	English		Spanish	
	Singular	Plural	Singular	Plural
Grammatical				
	this book	these books	este$_{\text{masc.}}$ libro	estos$_{\text{masc.}}$ libros
	this table	these tables	esta$_{\text{fem.}}$ mesa	estas$_{\text{fem.}}$ mesas
Ungrammatical				
	these book	this books	estos$_{\text{masc.}}$ libro	este$_{\text{masc.}}$ libros
	these table	this tables	estas$_{\text{fem.}}$ mesa	esta$_{\text{fem.}}$ mesas

Table 2.3 shows both grammatical and ungrammatical nominal number agreement in English and Spanish. In Spanish, determiners including articles, possessive pronouns, and demonstratives require agreement with the noun they control. If the controller specifies singular number, the target must spell out singular also. In cases of discordant (i.e. ungrammatical) agreement, as seen in the bottom row of Table 2.3, the controller specifies a certain value of number, but the target spells out the opposite value (i.e. singular instead of plural, or plural instead of singular). In these examples, plural is the marked value for number in both languages. The Spanish examples also provide a masculine and feminine example. This additional complexity of gender agreement will be shown later.

Another regular relation of nominal agreement in Spanish is that number must agree between nouns and their modifiers. This agreement relation is found in noun–adjective pairs. See the following examples:

(1)	falda-ø	larg-a-ø	*falda-ø	larg-a-s
	skirt.FEM-SG	long-FEM-SG	skirt.FEM-SG	long-FEM-PL
	'long skirt'		'long skirt'	
(2)	viaje-s	cort-o-s	*viaje-s	cort-o-ø
	flight.MASC-PL	short-MASC-PL	flight.MASC-PL	short-MASC-SG
	'short flights'		'short flights'	

In (1) and (2) the controller nouns establish the value for the number condition when determining agreement with their target adjective. The examples in the left column of (1) and (2) show concordant number agreement whereas the examples in the right column show discordant number agreement. The controller nouns specify singular and plural number in (1) and (2) respectively, but the target adjectives spell out differently, creating agreement anomalies.

Apart from nominal agreement, grammatical number is also an important feature in Spanish verbal agreement. The number of the verbal subject (the controller) requires agreement from the verb (the target), as seen in the following examples.

(3)	los	perro-s	beb-en	agua
	the	dog.MASC-PL	drink-3PL	water
	'the dogs drink water'			

(4)	el	gato-ø	beb-e	la	leche
	the	cat.MASC-SG	drink-3SG	the	milk
	'the cat drinks the milk'				

Examples (3) and (4) show verbal number agreement between subject-verb pairs for plural number and singular number respectively. It may be worthwhile noting that whereas number marking on nominals and their modifiers is transparent, number marking on verbs is more opaque, perhaps adding an additional complexity to verbal agreement processing.

Gender agreement. Grammatical gender is a salient and regular feature primarily expressed in Spanish nominal agreement. The classification of grammatical gender may follow semantic or phonological rules; however, in Spanish, its classification is mostly arbitrary. Gender commonly appears in agreement relations between nouns and their determiners and articles, as seen in the following examples.

(5)	la	casa	*la	caso
	the.FEM	house.FEM	the.FEM	case.MASC
	'the house'		'the case'	

(6)	el	caso	*el	casa
	the.MASC	case.MASC	the.MASC	house.FEM
	'the case'		'the house'	

The examples in (5) and (6) show concordant and discordant gender agreement in the left and right columns respectively. In the case of both (5) and (6) the controller word is the noun and the target is the determiner. When the noun specifies the gender as feminine, the determiner must be classified as a feminine noun, otherwise a gender mismatch occurs. Grammatical gender agreement also presents itself in noun-adjective pairs. In the following examples, the controller word is the noun and the target is the adjective. Whichever gender the noun inherently expresses, the following adjective must also express to avoid discordant agreement.

(7)	chica	bonit-a	*chica	bonit-o
	girl.FEM	pretty-FEM	girl.FEM	pretty-MASC
	'pretty girl'		'pretty girl'	
(8)	niño	gracios-o	*niño	gracios-a
	boy.MASC	silly-MASC	boy.MASC	silly-FEM
	'silly boy'		'silly boy'	

Person agreement. Grammatical person is the third prominent feature that appears in agreement relations, often between subject-verb pairs. Whereas number and gender may be expressed separately on nouns, determiners, and adjectives, grammatical person in Spanish cannot be expressed without simultaneously expressing number. It is an inherent feature of nominal stems (any non-pronominal subject necessarily spells-out as third person), and pronominal subjects. These characteristics make it rather unique when considering its role in agreement relations. The following examples show the expression of person agreement in Spanish subject-verb pairs.

(9)	yo	com-o	una	torta
	1SG.SUBJ	eat-1SG	a	sandwich
	'I eat a sandwich'			

(10)

tú	com-es	pan	cada	noche
2SG.SUBJ	eat-2SG	bread	every	night

‘You eat bread every night’

(11)

nosotros	mir-amos	un	pájaro
1PL.SUBJ	watch-1PL	a	bird

‘We watch a bird’

(12)

ellos	mir-an	al	cuadro
3PL.SUBJ	watch-3PL	the	painting

‘They look at the painting ’

Apart from these three features, other grammatical features may also appear in agreement relations including case, definiteness, and tense (Corbett, 1998). Of particular interest to this thesis is the consideration of tense as a grammatical feature in Spanish agreement relations. While tense may not appear to be as salient a feature as person or number, its role as a feature in verbal agreement is important to consider when factoring in the presence of a temporal adverb. It’s true that subject-verb agreement receives more attention than adverb-verb agreement, and for good reason. The features expressed in subject-verb agreement (i.e. person and number) are more salient and a subject is typically required to make a clause grammatical, whereas a temporal adverb is not. However, if a temporal adverb is present, it expresses some value of the tense feature, and this value must agree with the tense marking on the verb. Consider the following temporal agreement relations in Spanish.

(13)

Ayer,	el profesor	d-io	una	tarea
yesterday	the professor	give-3SG.PAST	a	homework

‘The professor assigned homework yesterday’

(14)

Mañana,	los paquetes	llegar-án
tomorrow	the package	arrive-3PL.FUT

‘The packages will arrive tomorrow’

Examples (11) and (12) both have a temporal adverb that controls tense for the verb. In both examples, the main verb agrees in tense with the adverb.

(15)

*Anoche,	yo	sal-go	con	mis	amigos
last night	1SG.SUBJ	leave-1SG.PRES	with	1SG.POSS	friends

'Last night, I go out with my friends'

(16)

*Mañana,	ella	regres-ó	a	casa
tomorrow	3SG.FEM.SUBJ	return-3SG.PAST	to	house

'Tomorrow, she returned home'

The examples in this section demonstrate the complexity and variety of nominal and verbal agreement in Spanish. They depict the types of features and feature relations learners of Spanish encounter when acquiring a knowledge of Spanish grammar and agreement.

Native Processing

The ability to process grammatical agreement correctly is a critical step in language acquisition not only for second language learners but also for native speakers. Knowledge of agreement rules is crucial for language users to parse and access lexical items for proper language production and comprehension.

Several studies have investigated native speakers' behavior while processing agreement (Barber and Carreiras, 2005; Lim and Christianson, 2015; Molinaro et al., 2011; Romanova and Gor, 2017). Much of the prior work investigated whether native speakers demonstrate integrated knowledge of grammatical agreement rules and, if they do, how they process grammatical agreement. These studies investigated these questions by examining how native speakers responded to various types of agreement violations, the primary characteristic of the grammatical violation paradigm. From this research, it is hardly disputed that native speakers do show integrated knowledge of agreement rules while processing agreement mismatches. However, by examining native speakers' behavior and brain responses to agreement violations, researchers have found interesting, yet disparate results. While ongoing research argues for and against both qualitative and quantitative differences between the types of agreement violations, a variety of measures have demonstrated native speakers do exhibit different behaviors in response to agreement violations and agreement non-violations (Barber and Carreiras, 2005; Lim and Christianson, 2015; Mancini et al., 2011, 2014).

When native speakers process grammatical agreement, they demonstrate agreement sensitivity through congruency and incongruency effects during online processing (Barber and Carreiras, 2005; Keating, 2009; Romanova and Gor, 2017; Sagarra and Herschensohn, 2011; Tokowicz and Warren, 2010). Researchers have captured congruency and incongruency effects in participants' behavior when presented with different agreement scenarios. When natives are presented with syntactic items that agree in all features, they exhibit congruency effects through shorter reaction times (latencies), insignificant amplitudes in P600 and N400 effects[2], and quicker reading times. When syntactic items do not agree, natives exhibit incongruency effects through longer reaction times; greater amplitudes in LAN, P600, and N400 effects; and longer reading times. These congruency and incongruency effects have been demonstrated using a variety of online processing methodologies: lexical decision tasks (LDT) (Romanova and Gor, 2017); event-related potentials (ERP) (Barber and Carreiras, 2005; Fraga et al., 2017; Molinaro et al., 2011); self-paced reading (Sagarra and Herschensohn, 2011; Tokowicz and Warren, 2010); and eye tracking (Foucart and Frenck-Mestre, 2012; Keating, 2009; Lim and Christianson, 2015). It is assumed that longer reading times indicate natives focus more attention on processing agreement; therefore, discords in agreement require greater cognitive effort than concordant agreement. The indication that native speakers take longer to read and process sentences with agreement anomalies strongly suggests they are sensitive to those types of grammatical violations and therefore are likely to have an integrated knowledge of the rules and forms associated with grammatical agreement.

L2 Processing

Looking beyond native speaker agreement processing, several researchers have investigated whether second language learners exhibit native-like congruency and incongruency

[2]P600, N400, and LAN are event-related potentials, or neural metrics that I will refer to throughout this thesis. Larger amplitudes in these effects are taken to indicate greater cognitive effort.

effects during online agreement processing. In particular, they examined whether learners show qualitative or quantitative similarities to natives in their sensitivity to agreement violations. Qualitative similarities are indicated by similar fixation areas, rereading and skip areas for self-paced reading and eye tracking methodologies, or similar location amplitudes (positive or negative) in EEG for ERP methodologies. Quantitative similarities are indicated by similar measures of reading durations (i.e. fixation time, gaze duration, total reading time), or similar measures of amplitude in EEG. Researchers have explored whether learners demonstrate similar congruency and incongruency effects to native speakers as described above, when processing concordant and discordant agreement respectively. However, the findings on this topic are still inconclusive.

In the past decade, several theoretical accounts have been presented which posit contrasting theories to explain differences between L1 and L2 grammatical processing, including agreement processing. These theories converge on a single point: L1 and L2 processing cannot be similar. One of the primary hypotheses to emerge during this time was Clahsen and Felser's (2006a; 2006b) Shallow Structure Hypothesis (SSH). In their hypothesis, the authors claim that L2 speakers exhibit reduced sensitivity to grammatical information during morphosyntactic processing. This reduced sensitivity makes it difficult for learners to develop qualitatively similar mechanisms to fully parse and comprehend grammatical dependencies. While learners may approach native-like processing, the underlying qualitative mechanism they use yields to computational difficulty and doesn't allow them to process fully representational syntactic structures. Instead, learners create "shallow" or "local" representations (i.e. adjacent dependencies in agreement processing or processing of full words instead of using decomposition), relying more on lexical and semantic information in order to process the complete hierarchical structures as natives do.

Clahsen and Felser's (2006a; 2006b) SSH presents a problem, however, in that only learners' qualitative differences in native-like processing are taken into account. As described above, qualitative differences are not the same as quantitative differences. Both types represent separate utilities developed and used by learners. Qualitative differences suggest learners employ different processing mechanisms, whereas quantitative differences suggest the learner and native mechanisms are the same but are used to different extents (i.e. a more limited extent for learners than for natives). Some studies have refuted the SSH because on-line processing data has been gathered from behavioral (SPR, eye tracking) and brain (ERP, fMRI) studies which indicate that learners do exhibit similar qualitative behavioral and neural data (Cunnings, 2017; Marull, 2017; Reichle et al., 2016). This thesis also expects to produce similar behavioral data from natives and learners with eye tracking methodology.

Ullman (2001) also presented a knowledge theory to account for differences in L1 and L2 processing. His Declarative/Procedural model describes two memory systems that subserve storage and access to different types of lexical and grammatical information. Natives draw upon both memory systems during morphosyntactic processing. They rely on a declarative memory domain where information about monomorphemic words, inflectional irregularities, and idioms are stored; and a procedural memory where complex words are stored and accessed through decomposition due to their regular inflection patterns. Whereas native speakers have access to both memory systems, L2 learners only access their declarative memory through knowledge and memorization of grammatical rules. It is proposed that as learners gain greater proficiency, proceduralization of grammatical rules occurs and learners garner greater and greater access to procedural memory as natives do instead of relying too heavily on declarative memory alone. This mechanism of accessing morphosyntactic knowledge represents a qualitative difference between L1 and L2 speakers in processing (Cunnings, 2017; Reichle et al., 2016). In a similar manner to Clahsen and Felser's SSH, Ullman's model can perhaps be refuted through evidence that underlying mechanisms involved in L1 and L2 processing are more similar than they appear.

Whereas Clahsen and Felser's (2006a; 2006b) and Ullman's (2001) theories only account for qualitative differences between L1 and L2 processing, other theories posit that L1/L2 differences are more quantitative in nature. These theories claim that native speakers and learners use qualitatively similar underlying mechanisms for morphosyntactic processing. McDonald (2006) and Hopp (2006; 2010) proposed that L2 processing requires greater cognitive resources from L2 learners. These resources are finite and limited. Learners' limited cognitive resources like working memory capacity are therefore due to a lack of L2 knowledge, proficiency, or proceduralization of L2 grammatical processes. Therefore, during a taxing cognitive task like morphosyntactic processing, learners' cognitive supply becomes exhausted quickly, despite employing similar processing mechanisms as native speakers. This limited supply of resources demonstrates a quantitative difference between L2 and native processing. This difference exhibits measurably slower or reduced responses in both behavioral and neural imaging data.

While the question of whether L1 and L2 processing differences are more qualitative or quantitative in nature has been debated, several studies have demonstrated that in contrast to native speakers, learners have reduced or no sensitivity to agreement errors in their second language. Learners have exhibited dissimilar incongruency effects to native speakers (Bosch and Clahsen, 2016; Clahsen and Felser, 2006b; Jiang, 2004; Jiang et al., 2011; Keating, 2009). This lack of sensitivity to agreement violations has been attributed to a variety of different factors including: age of acquisition, working memory capacity, language proficiency, engagement of cognitive mechanisms, L1 transfer of linguistic features, and transparency of morphophonological cues (DeKeyser, 2005).

In their 2011 study, Jiang et al. investigated late L2 learners' competence in select linguistic domains. They examined the morphological congruency hypothesis, which questions whether L2 learners integrate non-L1 morphological knowledge during L2 morphological processing. To investigate this question, the researchers set up a self-paced reading experiment with English native speakers, and Russian and Japanese ESL speakers as subjects. They used violations in plural marking and verb subcategorization, and tested participants' sensitivity while processing each condition. The results indicated the native English speakers were sensitive to both types of violations, and Russian ESL speakers were sensitive to both violation types, but Japanese ESL speakers were only sensitive to verb subcategorization errors. The results were consistent with findings from Jiang's (2004; 2007) earlier studies, where Chinese ESL speakers did not show sensitivity to plural agreement errors in English. Overall, the three studies indicate that L2 learners do not show similar sensitivity to morphological agreement errors when processing non-L1 morphological information. Such findings suggest that L2 linguistic competence is selective and that underlying mechanisms for processing non-L1 morphosyntactic information must be acquired separately from underlying mechanisms that process L1 morphosyntactic information, implying a qualitative difference between L1 and L2 processing.

Similar to Jiang et al. (2011), Keating (2009) investigated learners' processing of non-L1 morphological knowledge. Using eye tracking, he examined native and non-native speakers' sensitivity to gender agreement errors in Spanish. He examined whether non-native speakers of Spanish exhibit acquisition of gender agreement through similar patterns of sensitivity to agreement anomalies as native speakers. He tested three groups of learners from varying levels of Spanish: beginning, intermediate, and advanced. The results indicated that, like native speakers, advanced-level learners of Spanish showed sensitivity to gender agreement violations. The findings demonstrated that adult English learners of Spanish can acquire gender agreement (a non-L1 feature) in L2 acquisition. However, advanced learners were only sensitive to violations in short dependency agreement relations, unlike natives who were sensitive to both short and long dependency relations. This suggests that while advanced learners may be sensitive to agreement errors, they do not show similar patterns of sensitivity to native speakers. Additionally, only advanced learners showed sensitivity to violations; intermediate and beginning learners did not show sensitivity to gender agreement violations. This suggests that gender agreement is not acquired until advanced levels in L2 acquisition.

While a lack of learners' sensitivity to agreement mismatch has been shown in some studies, other authors have cited learners who show similar sensitivity to agreement errors as native speakers. Foote (2011) investigated integrated morphosyntactic knowledge in early and late English-Spanish bilinguals. Using a moving window paradigm in a Spanish reading task, she examined participants' sensitivity to subject-verb agreement errors and noun-adjective agreement errors. She found that both groups of bilingual speakers (i.e. early and late) showed similar incongruency effects as native speakers to both types of agreement errors. The results suggested that in contrast to other findings, late L2 learners develop a similar type of linguistic knowledge for their L2 as seen in their L1. This further demonstrates that late learners can acquire integrated knowledge of linguistic features not present in their L1 and they can use similar mechanisms as natives for processing that information. This was seen with late learners of Spanish who were sensitive to gender agreement errors in noun-adjective pairs, a feature not present in English, their L1.

Sagarra and Herschensohn (2010) investigated the effect of language proficiency on the processing of Spanish morphosyntactic information. They were interested in whether L2 learners at different levels of Spanish proficiency would demonstrate similar sensitivity to agreement errors of different types of Spanish nominal agreement violations (i.e. number and gender). They employed an agreement violation paradigm to observe learners' sensitivity to different features in agreement violation conditions. The authors recruited beginning and intermediate learners of Spanish and native Spanish speakers for a study using offline and online methodologies. Participants engaged in a grammaticality judgment task and a self-paced reading task. The results of both tasks suggested that L2 learners can exhibit sensitivity similar to native Spanish speakers while processing agreement violations. However, only the intermediate learners showed this sensitivity to both types of agreement anomalies. Beginning learners of Spanish did not show sensitivity to agreement errors of either type (i.e. number or gender) in either the offline or online tasks. The authors found converging evidence with Foote (2011) that L2 learners can demonstrate sensitivity to agreement errors in Spanish, particularly gender agreement errors.

The prior studies make it clear that many factors (similar or different) may be involved in L1 and L2 morphosyntactic processing. While the precise mechanisms and functions of these factors or interaction of factors modulating agreement sensitivity remain unclear, several researchers recognize it is possible for learners to exhibit native-like sensitivity. In light of the previous research examining differences between L1 and L2 agreement processing, this thesis also investigates whether L2 learners of Spanish demonstrate sensitivity similar to native Spanish speakers while processing agreement violations. I further investigate not only processing differences between L1 and L2 speakers, but also differences between L2 learners at different language learning levels. Previous research has demonstrated that novice learners show less sensitivity to agreement errors than advanced learners (Keating, 2009; Sagarra and Herschensohn, 2010; Tokowicz and Warren, 2010). In this thesis, I also ask whether learners from different levels of Spanish exhibit similar sensitivity to agreement violations.

Working Memory Capacity

Apart from language proficiency and language level, other individual differences such as working memory capacity may also account for differences in L2 processing ability. As an individual cognitive resource, working memory and its capacity to store and process information is important for individuals to carry out complex cognitive tasks. Researchers have demonstrated that this storage has limited capacity and limited functionality for processing information during certain tasks (Baddeley, 2003, 2007). For verbal working memory, the storage space integrates information drawn from several different lexical domains (e.g. phonology, syntax/morphology, and semantics). For certain cognitive tasks, this integration of information can be taxing on individuals and limit their ability to process lexical information. This is true for complex cognitive tasks such as reading comprehension.

Several studies have shown that WMC has a significant effect in cognitive tasks such as L2 sentence processing (Dai, 2015; Juffs, 2015; Tagarelli et al., 2015) and L2 morphological processing (Coughlin and Tremblay, 2013; Havik et al., 2009; Sagarra and Herschensohn, 2010). Other studies, however, have shown that WMC does not have an appreciable effect on L2 morphosyntactic processing (Felser and Roberts, 2007; Foote, 2011).

Some studies that have observed an effect for WMC suggest that while WMC does have an effect on L2 processing, the effect is only present with low-to-mid-proficiency learners and is lost with learners of higher proficiency. In Sagarra and Herschensohn's (2010) study of gender and number agreement processing, they found that lower proficiency learners of Spanish demonstrated greater demands on their WMC while processing agreement anomalies. The results of this study suggest that WMC may have an effect on L2 processing, but the effect is seen with lower proficiency learners; the limitations for WMC storage and demands for information integration during L2 processing are greater for lower proficiency learners than higher proficiency learners.

Contrary to the results of Sagarra and Herschensohn (2010), Coughlin and Tremblay (2013) found that WMC displayed a greater effect for high-proficiency learners than for low-proficiency learners, although this effect was only a weak predictor. They examined the role of WMC in processing number violations between object clitics and their left-dislocated antecedents by L1 English L2 learners of French. For certain anomaly conditions, WMC was a predictor for high-proficiency learners' sensitivity to errors. Also, in Havik et al.'s (2009) study of subject-object relative clause ambiguities in Dutch, they found that the WMC effect was greater for higher proficiency German learners of Dutch than for lower proficiency learners. These two studies are in line with others that suggest while WMC is likely language independent, it is dependent on proficiency and in fact covaries with proficiency in a second language (i.e. as L2 proficiency increases, WMC also increases). It is clear from the studies highlighted in this section that the interaction between L2 proficiency and WMC shows disparate effects with WMC's influence on L2 processing. This area of research warrants further investigation.

There is also a dearth in the literature of studies that have directly investigated the effect of WMC on morphological processing in regards to early and late processing stages during the time course of reading comprehension. It may be true that WMC is more important for some learners during a period of reading processing than for other learners at a different period of reading processing. With this lack of consensus on the effect of WMC generally with L2 processing, it is clear there is much to be explored. Investigations in L2 agreement processing with WMC as a factor may clarify the interaction of WMC with proficiency and early/late processing.

Varieties of Exponence in Agreement Processing Studies

In parallel with studies on L1/L2 differences and individual differences in agreement processing, many researchers have investigated the effect of feature type on agreement processing. They have examined whether different features expressed in agreement relations elicit different behavioral and neural responses from native speakers and learners. As presented before, several different features (number, gender, person, and tense) may participate in an agreement relation between different word pairs (noun-adjective, noun-determiner, subject-verb). In these agreement relations, varying types of exponents (phonological realization of grammatical features) are used to express grammatical features. Exponence is used here to mean the expression of inflectional values, or grammatical features. Two primary types of exponence that are used in agreement relations are separative exponence and cumulative exponence. In separative exponence, a single grammatical feature is mapped to a single exponent. In cumulative exponence, multiple grammatical features are mapped to a single exponent. Table 2.4 demonstrates both types of exponence from Spanish examples.

Separative exponence is characteristic of expressing a single feature as seen in the word *buena* ('good'). Here, the suffix -*a* expresses a single feature of gender with the value feminine. Some words may exhibit separative exponence multiple times, in an agglutinative way. The word *buenas* is a good example where two instances of separative exponence are used separately to express two features: gender and number. The word *hablé* ('I spoke') exemplifies cumulative exponence by syncretic expression of three features. The differences in

Table 2.4: Exponence Types in Spanish

Separative exponence	Cumulative exponence
$-a \mapsto Feminine$ $-s \mapsto Plural$	$-é \mapsto [1^{st}person, Singular, Past]$

exponence and how grammatical features are expressed has warranted extensive investigation into how speakers (native and L2) interpret and process those features during reading.

Separative exponence in agreement processing. Several studies have investigated native and L2 speakers' processing of agreement features expressed by separative exponence. In these studies, researchers primarily examined individuals' sensitivity to agreement errors of features that are expressed by a single exponent or zero exponents. This is characteristic of separative exponence. That is, the target constituent in an agreement relation is constrained to a single grammatical feature with the absence of simultaneous or cumulative exponence. For instance, the productive plural suffix in English *–s* is a single exponent that represents a single feature—number—when attached to nominal stems. The primary purpose for many of these studies was to investigate the processing behavior of native and second language speakers and identify possible behavioral processing differences between both groups, as described in the previous section. These pioneering studies were important because of their early research into processing differences between concordant and discordant agreement of a single feature, including number (Jiang, 2004; Lim and Christianson, 2015) and gender (Bates et al., 1996; Grosjean et al., 1994; Guillelmon and Grosjean, 2001).

Separative exponence of multiple features. Instead of primarily examining a single feature in agreement processing as noted before, several researchers further investigated whether different types of features in agreement relations are processed similarly. The primary method in this line of research was to investigate whether two features expressed on two separate exponents might elicit similar congruency or incongruency effects from both native and second language speakers (Barber and Carreiras, 2005; Gillón-Dowens et al., 2008). An example of this construction in Spanish is depicted in Table 2.4 with the word *buenas*. This extensive literature examined differences between gender and number processing. In most every case of languages studied, gender and number were expressed on separate exponents in a linear order, generally with gender more proximal to the stem than number.

Several researchers have further investigated whether different types of morphological information are represented and processed differently. Many of these researchers hypothesized that gender would elicit different behavioral and neural responses from participants, indicating participants were more sensitive to gender agreement anomalies than number agreement anomalies. These hypotheses were based on the assumption that violations of a stem-inherent morphological feature (e.g. gender) would cause greater disruption to agreement processing mechanisms and would require greater repair costs. Some studies have indicated that processing gender agreement is more cognitively taxing than number agreement for native speakers (Barber and Carreiras, 2005; Gillón-Dowens et al., 2008). In an ERP study Barber and Carreiras (2005) investigated differences in processing gender and number agreement in Spanish. Native Spanish speakers were tested in the violation paradigm by reading gender and number agreement anomalies in determiner-noun and noun-adjective word pairs. The results of the ERP indicated that both violation types elicited a LAN-P600 effect, suggesting that natives were sensitive to both gender and number mismatches. However, the authors found different EEG patterns in the late phase of P600. They discovered a greater amplitude for gender than for number mismatches, indicating that even though mismatch detection may not be different between the features, repair strategies are different. Gender agreement violations require greater repair costs than number violations.

However, Sagarra and Herschensohn (2011) found no significant differences between gender and number agreement processing with beginning and intermediate Spanish learners.

The authors investigated whether native speakers and learners process noun-adjective gender and number agreement differently. In a self-paced reading study, their participants read plausible sentences in Spanish and answered comprehension questions after each sentence. They were presented with three types of experimental sentences: gender and number concord, gender discord, and number discord. Their results indicated that beginners were not sensitive to gender and number agreement violations. While intermediate learners and natives showed incongruency effects, results further showed there were no processing differences between gender and number agreement. In a similar study, Romanova and Gor (2017) showed that Russian monolinguals and advanced Russian learners do not process gender and number agreement differently, further supporting the results of Sagarra and Herschensohn (2011).

Cumulative exponence in agreement processing. Investigations into processing differences between different features of separative exponence has led researchers to further explore processing differences between features of cumulative exponence. Until recently, few authors have tasked themselves with empirical pursuits of native and learner differences in agreement processing of cumulative exponence. These studies largely emerged only in the past decade and the variety of methodological tools, while admirable, is still sparse. Three major studies in this area are highlighted here, each study using a separate psycholinguistic or neurolinguistic tool.

In two self-paced reading experiments, Mancini et al. (2014) investigated the person-number distinction in processing subject-verb agreement by native speakers of Italian. They hypothesized differences in natives' reading behavior during processing of person anomalies and number anomalies. Specifically, they expected to observe greater reading times for person anomalies compared to number anomalies. Their first experiment yielded no significant results between natives' sensitivity in processing the two feature anomalies, thus converging with Silva-Pereyra and Carreiras' (2007) ERP study. The authors ascribed their lack of significant differences to their construction of anomalous stimuli, in particular with the subject pronouns. In their second experiment, they used third person singular and third person plural subject nouns instead. This second experiment found significant quantitative differences between natives' sensitivity to person and number anomalies. In line with their hypotheses, natives spent longer reading person anomalies than number anomalies. The authors attribute this greater perturbation of person anomalies to different anchoring requirements of feature interpretation. Person information relies both on morphosyntactic and speech act (or participant roles) information for interpretation. Breaking the anchoring mechanism between morphosyntax and speech act is disruptive and requires greater repair, seen by longer reading times for person anomalies. On the other hand, number does not rely on speech act anchoring mechanisms and therefore causes less perturbation than person because number anomalies can be repaired quicker.

Mancini et al. (2011) also conducted an ERP Spanish study on individuals' differences in sensitivity to person and number agreement violations. They expected to find similar LAN-P600 patterns for number mismatches as had been seen with prior studies (Barber and Carreiras, 2005; Molinaro et al., 2008). They expected to see qualitative and quantitative differences between person mismatches and number mismatches, namely a greater N400 effect for person mismatches due to the persistent difficulty with repair costs. Their results indicated measurable differences between person and number mismatches. As expected, a LAN-P600 was elicited for number anomalies, demonstrating similar patterns to other agreement mismatch ERP study findings. An N400 effect was also elicited in person mismatches, which was expected but also a new finding for ERP agreement violation studies. The amplitude differences for the P600 effect were greater for person than for number mismatches. The N400 effect and greater P600 effect indicated individuals were more sensitive to person anomalies than to number anomalies. Mancini et al. (2011) interpreted these findings to mean that person mismatch likely required more costly repair operations than number mismatch. Overall, the quantitative and qualitative differences in person and number mismatches from Mancini et al.'s (2011) study support prior evidence for feature dissociation, particularly features expressed by cumulative exponence.

Mancini et al. (2017) further investigated feature dissociations at the neuroanatomical level using an event-related fMRI paradigm. They again examined differences in person and number features in Spanish subject-verb agreement. Using native Spanish speakers, the researchers presented participants with concordant and discordant examples of person and number agreement on verbs. They examined certain cortical areas—left inferior frontal gyrus (LIFG) and middle temporal gyrus (MTG)—involved in comprehension of syntactically correct sentences and investigated whether natives would demonstrate quantitatively and qualitatively different responses to person and number violations.

As expected, both quantitative and qualitative differences were found in those cortical areas. While both person and number violations generated sensitivity from the left MTG, person violations elicited greater responses and more selective responses. These results were in line with the researchers' hypotheses and show that a feature dissociation was observed at the neuroanatomical level. The findings further support previous evidence from behavioral and neural studies demonstrating the existence of different processing complexities of features expressed by cumulative exponence.

Eye Tracking

Linguistic studies employing eye tracking methodology have become increasingly popular over the past decade (Keating, 2009; Lim and Christianson, 2015; Qing et al., 2018; Zeyrek and Acarturk, 2014). Using eye-tracking methodology, this thesis will investigate whether L2 learners are sensitive to different information when that information is represented by cumulative exponence. In Spanish, person, number, and tense information are expressed by a single form in verb endings as discussed and shown above in Table 2.4. This thesis investigates whether English learners of Spanish are equally sensitive to the different types of agreement errors that occur with verbal features expressed on Spanish verbs (i.e. person mismatch, number mismatch, and tense mismatch).

The advantage of eye tracking over other psycholinguistic methodologies (e.g. self-paced reading and lexical decision tasks) is clear. More so than the other tools, eye tracking captures the time course of reading from start to finish. In this way it approaches the most natural reading environment for participants. Researchers choose from many reading time and reading behavior measures for which they can gather and record reading data. This greater pool of measures from which to draw allows researchers to closely examine small details in reading behavior. They can capture reading behavior of individual morphemes and characters. Researchers can also track eye movements to gather information about the words individuals skip and the areas individuals reread.

Not only is eye tracking useful for examining minute details of words and morphemes, it also captures individuals' behaviors during the entire time course of reading. Whether at the discourse or sentence level, researchers use eye tracking to record data from different stages of reading comprehension. This entails detailed recordings of early processing, late processing, and very late processing. These stages line up with different cognitive skills during reading comprehension. That is, early processing entails lexical access, late processing entails information integration and reanalysis and repair, and very late processing also entails information reanalysis and repair. Certain eye tracking measures nicely provide data for these different stages of reading comprehension. These measures are regularly employed in eye tracking studies of reading behavior and are also used in this thesis. The definitions of each measure are provided in the methodology section, but here it suffices to provide a simple diagram of reading measures and how they line up with different processing stages of reading comprehension, as seen in Table 2.5. This is how the measures will be discussed later with the results.

Table 2.5: Eye Tracking Variables Matched with Early and Late Reading Measures

First fixation	Gaze duration	Go-past time	Total time	Fixation count	% Regression out	Spill-over area
Early processing		Late processing				Very late processing

Eye tracking has been implemented in many reading comprehension studies, particularly sentence comprehension. Researchers investigate whether there are observable differences in individuals' reading behavior of different linguistic phenomena (e.g. garden path sentences, ambiguities, agreement relations, etc.). Of particular interest to this thesis, prior studies have investigated agreement relations through anomalies (Keating, 2009; Mancini et al., 2014; Sagarra and Herschensohn, 2010). I employ a similar experiment design for the current study, using eye tracking to observe learners' reading behavior of agreement anomalies.

Research Questions

By monitoring eye behavior, I observe how English learners of Spanish process grammatical information on Spanish verbs. I measure learners' sensitivity to different types of grammatical information presented on Spanish verbs through the use of a grammatical agreement violation paradigm. I also measure how learner sensitivity to grammatical errors is affected by language level and working memory capacity. To this end, this thesis addresses the following research questions:

1. Do L2 learners demonstrate similar sensitivity to agreement errors as native speakers?
2. Do L2 learners of different university class levels demonstrate similar sensitivity to agreement errors?
3. What effect does working memory capacity have on individuals' sensitivity to agreement errors?
4. Do features expressed by cumulative exponence elicit similar sensitivity from learners during processing of agreement violations?

Hypotheses

In response to the research questions presented above, I hypothesize learners will demonstrate quantitative and qualitative similarities to native speakers while processing agreement errors. In particular, advanced learners' sensitivity will follow similar patterns to native speakers' sensitivity. I also expect language level and WMC will display measurable effects in learners' ability to process verbal agreement errors. Due to their greater metalinguistic knowledge of Spanish and more experience with language use, learners in higher language levels will show quicker reading times and fewer regressive behaviors indicating less perturbation to agreement anomalies. I expect that learners with higher WMC will display quicker reading times and fewer regressive behaviors than learners with lower WMC. Learners with greater cognitive ability like WMC will have an advantage in performing cognitive processes like verbal agreement. This advantage will display itself in early reading measures like first fixation duration and gaze duration, indicating an advantage in lexical access. Because of their increased storage and processing capacity in early reading behavior, learners with higher WMC will experience less perturbation when processing agreement anomalies. Lastly, I hypothesize that there will be a dissociation between the processing of features expressed by cumulative exponence. Learners' sensitivity will show disparate reading responses to person, number, and tense agreement errors. In particular, I expect learners will exhibit greater perturbation to tense anomalies than to person or number anomalies. This prediction is based on the fact that Spanish verbal tense agreement has a different controller constituent for tense than for person and number. A temporal perspective disruption will also cause learners greater perturbation than person or number disruptions.

Chapter 3

Methodology

In order to accomplish the aims of this thesis and investigate the four research questions presented above, I administered a language background survey and two separate tasks—a reading span task and an eye tracking task—to each participant. Participants completed the survey and both tasks in the Humanities eye tracking lab on Brigham Young University (BYU) campus. Participants came to the lab one at a time and spent approximately 30 minutes in total completing the tasks stated above. Each participant was compensated for her or his time.

Participants

The participants involved in this thesis were native English speakers learning Spanish as a second language. All participants were required to be students enrolled in a Spanish language course at BYU during the semester in which the study was conducted (Winter semester 2018). Students were recruited using in-class announcements, recruitment posters, and email advertisements. Overall, 65 participants were recruited. The demographics of the participants can be reviewed in Table 3.1. I recruited participants from three different university class levels: 100 level classes, 200 level classes, and 300 level classes. I further separated the participants in the 300 level courses into two groups: participants with extended residency abroad and participants with no extended residency abroad. Time spent abroad in a foreign language community has been shown to contribute to learners' proficiency and understanding of the language (Isabelli-García et al., 2018; Sagarra and LaBrozzi, 2018; Yang, 2016). I expected to see a similar effect in my thesis. For this purpose, I determined that students with three or more months residency in a Spanish-speaking community abroad were placed in a separate group from students with no extended residency in a foreign Spanish-speaking community. Three months is the typical length of a study abroad program at BYU. I refer to the group with no foreign residency as NRes 300, and the group with foreign residency as Res 300. The mean number of years of Spanish study for each class level is shown in Table 3.1. As is expected, participants in the 100 level classes had little prior experience studying Spanish (i.e. 1–4 years). Participants in 200 level courses had 2–4 years' experience studying Spanish, and participants in 300 level courses (both with and without residency abroad) had 3+ years' experience studying Spanish.

Fifteen native Spanish speakers were also recruited to serve as the experimental control group. All of the native speakers chosen for the study were recruited from the English Language Center (ELC) at Brigham Young University. All native speakers were currently enrolled in English language courses at the ELC and had been taking courses there for less than a year. However, some native speakers had several years of prior exposure to English. Regardless of the their time spent learning English, all native speakers had enough English proficiency to read and understand the consent form and study instructions. In all, I recruited 80 research participants for the five separate language levels. The mean number of participants recruited from each language level was 15, which is approximately the same sample size as participants recruited for similar L2 agreement processing studies (Keating, 2009; Tokowicz and Warren, 2010).

Table 3.1: Participant Demographics

	Level 100	Level 200	NRes 300	Res 300	Native
# of participants	17	20	14	14	15
ratio female : male	14F : 3M	13F : 7M	10F : 4M	9F : 5M	10F : 5M
mean age	19.8	20.8	20.5	21.3	22.8
mean yrs of Spanish study	2.9	3.6	4.6	4.6	———

All participants completed a questionnaire prior to the experimental procedures. In the questionnaire, participants detailed their language experiences with Spanish and any other languages, and they provided details regarding residency abroad in a Spanish-speaking community (Tokowicz et al., 2004). They also provided a self-rated proficiency (Tuninetti et al., 2015), current Spanish class enrollment, years studying Spanish, and the contexts (e.g. high school, college, study abroad) in which they studied Spanish. The purpose of the language background survey was to identify factors that participants exhibit which could have an effect on the study. Therefore, if certain participants identified factors unfit for this thesis (e.g. not enrolled in proper university Spanish course or first language is not English for L2 learners or Spanish for native speakers), they were excluded from participation.

Reading Span Task

I created a reading span task to calculate a score for working memory capacity (WMC) for each participant. Several prior authors have noted that reading span tasks are a valid method to measure individuals' WMC (or verbal working memory) (Baddeley, 2003; Coughlin and Tremblay, 2013; Sagarra and Herschensohn, 2010; Waters and Caplan, 1996). With a WMC score, I would be able to include WMC as a fixed effect in a linear mixed model and identify its significance as a main effect on agreement processing and if it had any significant interactions with other factors in the model (e.g. language level or type of feature anomaly).

Stimuli. The reading span task was administered in English. Stimuli for the task were taken from an online English frequency list[1], which orders words by lemma frequency. This frequency list draws frequency data from words listed in the Corpus of Contemporary American English (COCA)[2] developed at BYU. In order to establish a plausible list of stimuli, I chose to use only nouns as stimuli. Only selecting stimuli from one lexical category allowed me easier control of lemma and word form frequency and semantic weight. In this way, I avoided problems with regularity and markedness of verb inflections and derivations as well as

[1]See `https://www.wordandphrase.info/`
[2]See `https://corpus.byu.edu/coca/`.

lower semantic weight from other content word types (i.e. adjectives and adverbs). Function words were wholly disregarded as stimuli items.

I obtained a list of the 500 most frequent nouns with their frequencies and register distribution from the online frequency list. In order to have a less varied distribution of frequencies I only used nouns from the last 250 that appeared on the list of 500. From this list, I next eliminated nouns that were found numerically more in one register than the other registers. COCA categorizes word frequencies by five registers: one spoken register and four written (e.g fiction, magazine, newspaper, and academic). Nouns that only appeared in the spoken register and not fiction or newspaper were excluded from selection. Some lemma frequencies appeared numerically more in one written register than the other written registers. For example, nouns such as *knowledge*, *factor*, and *analysis* were found significantly more often in the academic register than any of the other registers. Nouns of this type were excluded from selection.

Next, because length of noun may have an effect on working memory capacity, I only selected nouns that were 6-7 characters long, as this was the mean length of nouns from the developed list. From the remaining nouns, I identified each one as having a concrete or abstract sense and removed the abstract nouns. I selected only concrete nouns as stimuli because there are differences between individuals' recall of concrete and abstract nouns (Walker and Hulme, 1999). After completing the selection process detailed above, I randomly selected 40 nouns that would serve as stimuli.

I separated the 40 stimuli into different sets that would be used for the reading span task. Ten sets were created and each set was assigned a certain number of trials (i.e. sentences). Of the ten sets, two sets each had a number of trials ranging from two to six (i.e. two sets had two trials, two sets had three trials, etc.). To further control for lemma frequency, which can affect a participant's likelihood of remembering certain nouns (nouns with higher frequency are more likely to be remembered) (Hulme et al., 1997), the mean frequency of each set of nouns was calculated. The frequencies from each set were put through a one-way

ANOVA and differences between mean frequency were tested. The results of the ANOVAs and post-hoc tests (Tukey HSD and Bonferroni) indicated that there were no significant differences between mean frequencies of any of the sets. The nouns were then positioned into sentences for presentation to the participants. All the experimental nouns were positioned as the last word of the sentence. All sentences were approximately 70 characters in length and contained one dependent clause.

Procedure. The reading span task was administered to each participant using Qualtrics, a research and survey development software. Participants completed the task after the language questionnaire and before completing the eye tracking task. Each participant completed the task on a computer in the eye tracking lab. Participants were given two practice sets before completing the actual task. Each practice set consisted of two trials. The entire task consisted of 40 trials separated into 10 sets as explained above. Each trial was a separate English sentence. The sets were presented in random order and there were an approximately equal number of grammatical and ungrammatical sentences in each set.

For each trial, a complete sentence appeared on the screen with an arrow button indicating "continue" at the bottom. Participants advanced through the survey by clicking the arrow button. Participants were instructed to read the entire sentence out loud once. Having the participants read out loud serves as a distraction with the purpose of preventing repetition of the target word. Participants were then instructed to click the next arrow as soon as they finished reading and advance to the next sentence. Once they clicked the arrow, the next sentence automatically appeared in the same position on the screen as the previous sentence. Participants were instructed to remember the last word of each sentence they read, and that at certain periods throughout the survey they would be asked to recall those words. They were told that they should try to recall the words from each set in the order in which they were presented. However, with the scoring method I used they were not penalized for recalling words in a different order.

Each sentence remained on the screen for six seconds, giving the participants enough time to read the entire sentence once. After six seconds, the sentence disappeared. However, participants were told to read the sentence entirely once and press the arrow button on the screen when they finished. After the participants pressed the arrow button, a new screen appeared prompting them to indicate whether the sentence they just finished reading was grammatical or not. They were given choices “Yes” and “No” to indicate their response. They were instructed to give their response as quickly and accurately as possible. Once they provided their response, they clicked the arrow button at the bottom of the screen to move on to the next sentence.

After the participants read through each sentence in a set, a new screen appeared asking to participants to type in all the last words of the sentences they had read in that set. They were encouraged to type any words they remembered even if unsure it was a last word in a trial. After they typed in the words they could remember, they hit the arrow on the screen and a new screen appeared instructing them that a new set was about to begin. They hit the arrow again to begin the new set. Participants followed this procedure for each set in the task.

Analysis and scoring. The responses for the reading span task were recorded on the Qualtrics site. For each participant, the grammaticality judgment response was recorded for each trial and the words they recalled were recorded for each set. All the responses were exported as a single .csv file from the Qualtrics site. To eliminate any participants that did not attend appropriately to the task, accuracy scores for the grammatical judgment were calculated. Those participants that had an overall accuracy of less than 85% were eliminated from the results of the task. No participants were eliminated as all had an accuracy score higher than 85%.

I used a partial-credit unit scoring method to assign a score for each participant’s working memory capacity. This scoring technique has been used in several other working memory span studies (Conway et al., 2005; Foster et al., 2015; Friedman and Miyake, 2005;

Redick et al., 2012) and has been shown to be a valid and reliable measure of scoring working memory span tasks.

Using the partial-credit scoring technique, a mean percentage is calculated for each participant. This percentage is found by calculating the percentage of correctly recalled items for each set in the reading span task and then calculating the mean percent correct of all the sets. In the case of my reading span task, all the last words from each trial were the test items, with a possible total of 40 test items. For each participant, I calculated the percent of correctly recalled words for each set (i.e. for a set of 4 trials, if the participant recalled 3 words correctly, their percent correct is 75%). Once the percentages for each set were calculated, I summed these percentages together and divided by the total number of sets (i.e. 10), in order to find the overall mean percentage of correctly recalled words for the task. I followed this scoring procedure for each participant. The percentages I gathered were the scores I assigned to each participant for the span task. The scores were a continuous variable I used as a fixed effect along with its interactions with other fixed effects in the statistics I describe in the Results section of this thesis.

Eye Tracking Task

Stimuli. To investigate learners' sensitivity to different types of Spanish verbal agreement anomalies, I created an eye tracking task in the grammatical violation paradigm. To test learners' sensitivity to different verbal feature anomalies, I created four different experimental conditions for subject-verb agreement in Spanish. I created one concordant agreement (correct agreement) condition as a control and three types of discordant agreement according to three Spanish verbal features (i.e. number, person, and tense). The discordant conditions were created in such a way as to isolate one verbal feature in discordant agreement, leaving the other two features in concordant agreement.

The syntactic structure for each of these conditions I created was the same. Every sentence began with a temporal adverb followed by a subject pronoun. The main verb of each sentence followed the subject pronoun. After the main verb either a theme or patient object followed. In certain sentences (approximately half), a prepositional phrase followed the object. In the concordant agreement condition, the temporal adverb and subject pronoun agreed with the main verb in person, number, and tense information. In the person discordant condition, the temporal adverb agreed with the verb in tense information; the subject pronoun agreed with the verb in number information but disagreed in person information. In the number discordant condition, the temporal adverb agreed with the verb in tense information; the subject pronoun agreed with the verb in person information but disagreed in number information. And in the tense discordant condition, the temporal adverb disagreed with the main verb in tense information while the subject pronoun agreed with the verb in both person and number information. A sample of the experimental stimuli for the Spanish verb *dejar* ('to leave') is given in Table 3.2.

Table 3.2: Test Stimuli for the Spanish Verb *dejar*

Correct					
Ayer,	yo	dej-é	mi	libro	en la clase.
yesterday,	1SG.SUBJ	leave-1SG.PAST	1SG.POSS	book	in the classroom.
'Yesterday, I left my book in the classroom.'					
Number discord					
*Ayer,	nosotros	dej-é	mi	libro	en la clase.
yesterday,	1PL. SUBJ	leave-1SG.PAST	1SG.POSS	book	in the classroom.
'Yesterday, we (I left) my book in the classroom.'					
Person discord					
*Ayer,	tú	dejé	mi	libro	en la clase.
yesterday,	2SG.SUBJ	leave-1SG.PAST	1SG.POSS	book	in the classroom.
'Yesterday, you (I left) my book in the classroom.'					
Tense discord					
*Mañana,	yo	dejé	mi	libro	en la clase.
tomorrow,	1SG.SUBJ	leave-1SG.PAST	1SG.POSS	book	in the classroom.
'Tomorrow, I left my book in the classroom.'					

The test stimuli for the eye tracking task were Spanish verbs in the first person singular simple past (preterite) form. The verbs used in the task were taken from three different sources: a frequency dictionary of Spanish (Davies, 2006), a historical and genre-based corpus of Spanish[3], and a web-based corpus of Spanish[4]. The 200 most frequent verbs from each list were compared across the three sources. The comparison of these lists showed little variation (i.e. no uniqueness in keyword lists) between the sources. I gathered the 200 most frequent verbs that were shared by the three sources into one list. I then removed any irregular verbs from this list. I wanted to keep the length of my stimuli within a certain medium range, so verbs with fewer than five characters and greater than eight characters were removed from the list. It is important to control for word length in eye tracking because short words receive fewer fixations, less reading durations, and have greater probability of being skipped (Clifton et al., 2007; Rayner, 2009).

From the remaining verbs, I considered both lemma and word form frequency. Both types of frequencies were taken from the Corpus del Español: Web/Dialects[5]. To find the lemma frequency, I made a query that produced a list of the most frequent verbs. I selected the 250 most frequent verbs from this list and reduced it down to the common list I made from the three sources, with the eliminations of the irregular verbs and those of inappropriate length. To find the word form frequency, I individually searched for each word form (i.e. 1SG.PAST forms) of each verb on the list and recorded its frequency. Any verbs that had an outlying frequency in either lemma or word form were removed. That is, if the lemma frequency was high, but the word form frequency was extremely low, the verb was removed. Or if the lemma and word form frequency were too similar (indicating that the verb rarely appears as other word forms in the conjugation paradigm), the verb was removed.

With the frequencies recorded, I chose 40 verbs to use as test stimuli and 80 verbs to use as fillers. Spanish has three conjugation classes according to infinitive verbal endings: -*ar* verbs, -*er* verbs, and -*ir* verbs. I selected an approximately equal number of verbs from the different conjugation classes. However, because verbs in -*er* and -*ir* classes conjugate

[3] See `http://www.corpusdelespanol.org/hist-gen/`.
[4] See `http://www.corpusdelespanol.org/web-dial/`.
[5] See `http://www.corpusdelespanol.org/web-dial/`.

similarly for first person singular in the past tense (i.e. *í*), I considered these as a single class. In this way, the test verbs included approximately 20 *-ar* verbs and 20 *-er*/*-ir* verbs.

I used a Latin square design for this experiment in order to control for unknown variability in the measured responses. From the 40 test verbs I selected, four groups were created with each group containing ten verbs. I established these four groups because of my four different experimental conditions. There was an approximately equal number of the different verb classes in each group. Both the mean lemma and word form frequencies were calculated for each group. I ran a series of independent t-tests between each of the groups. No significant differences between the mean lemma frequency or the mean word form frequency were found between the four groups. I also ran an independent t-test between the lemma and word form frequencies of the test stimuli and the fillers. No significant differences between the lemma or the word form frequency were found between these two groups.

In order to reduce the effects of any unpredicted factors, I designed the experiment in such a way that different participants would see the same test stimuli in different experimental conditions. Four different experimental tests were created using the EyeLink Experiment Builder software. Each experimental test placed the 40 test stimuli in a different agreement condition.

Each of the participants from the different class levels received one of the four experimental tests depending on the order they arrived at the lab. That is, the first participant in each of the class levels was assigned experimental test 1; the second participant received experimental test 2, and so on until the fifth participant who received experimental test 1, and the pattern repeated itself every four participants.

With the test and filler verbs selected and separated into the appropriate experimental groups, I created 120 Spanish sentences, each sentence using a different verb from those selected. First, the Spanish sentences were all written in the simple past with an appropriate temporal adverb (e.g. *ayer* 'yesterday' or *anteayer* 'day before yesterday') and subject pronoun (e.g. *yo* 'I') that agreed with the conjugated verb in the first person singular past (i.e. 1SG.PAST) form. The sentences were given to three different native Spanish speakers. The native speakers judged each sentence on its grammaticality and meaningfulness. They also corrected any colloquial or dialectal words uncommon across most varieties of Spanish. They provided feedback and corrections. Those corrections were then made and given back to the native speakers to judge again. Once the native speakers approved all of the sentences, I manipulated the sentences to create the necessary test conditions for the experiment. That is, for the 40 test sentences, I created a concordant agreement, person discord, number discord, and a tense discordant condition for each sentence. For the 80 filler sentences I created 50 concordant agreement conditions and 30 discordant agreement conditions (10 sentences for each of the three types of discordant agreement). In this way, of the 120 experimental sentences, I had an equal number of sentences with concordant agreement and discordant agreement.

Other Spanish words used in the sentences were also of high frequency and are learned by first year students. Third person and second person plural forms were not used in any of the test or filler sentences because they did not vary with first person singular in either number or person information.

Procedure. The eye tracking task was administered to each participant after she or he completed the reading span task. For the eye tracking task, subjects were seated in front of a computer screen and positioned into a tower mount with a chin rest. A camera attached to the tower mount recorded their eye movements during the experiment. Eye movement data was acquired using the SR research EyeLink 1000 tower mount and Experiment Builder software. The eye tracker was set up to track monocular eye movements at a sampling frequency of 1000 hz. The movements of each participant's right eye were recorded for the experiment.

A computer screen in front of the participants was positioned approximately 70 cm from the tower mount. This computer was the display screen, where participants would read through the instructions of the experiment and the experimental stimuli would be presented.

A separate computer was used by the experimenter to set up the experiment and monitor the participants' progress throughout the experiment. Both computers were set up in a certain position so that the experimenter was not in the visual field of the participants. This setup allowed the experimenter to limit visual distractions during the experiment.

Once participants were comfortably positioned into the tower mount, the camera on the tower mount was set up to recognize participants' eyes and record their eye movements. This involved adjusting the camera image so that the right eye was approximately centered in the camera image. The focus of the camera was also adjusted appropriately. Both pupil and corneal reflections were used to estimate participants' gaze position. The pupil threshold and corneal threshold were set at appropriate limits for better calibration. Participants' right eyes were then calibrated using a 9-point grid. Once calibration was evaluated as good, a validation was run. Calibration and validation were run until both performances achieved a positive evaulation and the camera could accurately estimate gaze position.

After the participants' eyes were calibrated and validated within the approved guidelines, participants read through the experiment instructions on the display screen. Participants were instructed they were to read through several sentences in Spanish, during which time their eye movements would be recorded. They were told to read each sentence silently to themselves as each sentence appeared and that after reading each sentence they would be asked to judge whether the sentence was grammatical or not. After receiving the instructions, participants were encouraged to ask the experimenter any questions about the procedure before continuing to the practice sentences.

Four practice sentences were used to help participants familiarize themselves with the procedure of the task. Here, they were introduced to the procedure of the actual task. For each trial in the task, participants pressed the spacebar on the keyboard in front of them and their eyes were directed to a fixation circle on the left-center of the display screen. They were instructed to fixate on the center of the circle and press the spacebar. Once the camera registered their gaze position on the circle, the trial sentence appeared on the screen with the beginning of the sentence appearing in the position of the fixation circle. Participants read through the sentence silently and pressed the spacebar when they finished. The sentence disappeared and a new screen instructed them to press "f" on the keyboard if the sentence was grammatical and "j" if the sentence was ungrammatical. Once participants gave their grammaticality response, the fixation circle appeared for them to continue to the next trial sentence. After completing the practice trials, participants were given an opportunity to ask any questions. Their gaze position was checked for offset and drift by the experimenter and if all looked well the participants continued to the actual experiment following the same procedure as described above.

The actual experiment consisted of 120 trials. After 60 trials, participants were given a rest break. At this point the experimenter calibrated and validated the participants' gaze position again and let them continue with the rest of the trials. During the entire experimental procedure, the experimenter monitored the offset and drift of participants' gaze position. If at any point the offset or drift were too far, the experimenter recalibrated and revalidated the gaze position and let the participant continue where they left off in the trials.

Analysis and scoring. Data from the eye tracker were measured and recorded according to the areas of interest (AOI). For the purpose of this thesis, the primary AOI was the main verb of each sentence. Only the verb was contained in this AOI. Other AOIs were created around the verb. The syntactic constituents before the verb (i.e. the temporal adverb and subject pronoun) consisted of one AOI and the syntactic constituents following the verb consisted of one or two AOIs depending on if a prepositional phrase was present. Therefore every sentence had three or four AOIs from which I could record and analyze my data. Some examples of the experimental AOIs are shown below.

Table 3.3: Examples of AOIs from Experimental Sentences

AOI			
1	2	3	4
Ayer, yo	metí	dos goles	en el partido.
Ayer, yo	comí	un pastel grande.	

Dependent variables for the eye tracking task were taken from the reading time data that was recorded at the AOIs selected. My primary interest was the AOI covering the main verb of each sentence. All reading time data were rounded to the nearest millisecond for clearer reading and interpretation. Six measures of eye movement were drawn from the reading time data: *first fixation duration* (the duration of the first fixation within the AOI during the first pass reading), *first run dwell time* i.e. *gaze duration* (the duration of time in the AOI from first entering to first exit), *regression path duration* i.e. *go-past time* (the total duration of fixations and saccades within the AOI until a forward exit, including the reading times of the prior AOI that was regressed into), *dwell time* i.e. *total time* (the total duration in the AOI including when it was regressed into), fixation count (the total number of fixations in the AOI including when it was regressed into), and *regression out* (proportion of trials where a regression was made out of the AOI prior to leaving the AOI in a forward direction) (see Holmqvist et al., 2011; Lim and Christianson, 2015; Rayner, 2009, for discussion and use of these metrics).

Grammaticality judgment accuracy was used as a cutoff measure in order to exclude the data from participants who did not properly engage (attend to) the eye tracking task. Participants whose accuracy performance was below 80% had their data removed from the analysis. Six participants had lower than 80% accuracy; however, all six participants were in a 100-level class and had higher than 70% accuracy, so no participant data were removed due to low grammaticality judgment accuracy. Only a single participant's data were excluded from the analysis due to bad eye tracking recordings. Despite attempts to achieve calibration, the eye tracking data was too noisy and had to be removed. In cleaning up the eye tracking data, fixations shorter than 80 ms and greater than 3 standard deviations from the mean were removed. I also removed any trials from the data where an individual gave an incorrect

Table 3.4: Class Level Mean Accuracy on Grammaticality Judgment

	Level 100	Level 200	NRes 300	Res 300	Native
mean accuracy	86%	91%	94%	94%	95%

grammaticality response as these trials contain erroneous responses and may have an effect on individuals' correct reading behavior. The mean accuracy scores for each class level's grammaticality response are given in Table 3.4. In all, 8% of the data was excluded from the analysis.

Chapter 4

Results

In this section, I present the results from the eye tracking task. Eye movement data were collected at all the established areas of interest; however, only two areas are of concern in this analysis—the main verb of each sentence and the area immediately following the verb. The various reading time measures described above were drawn primarily from these two critical regions. The main verb of each sentence comprised the primary interest of the study. In each sentence, the controlled syntactic constituents participating in verbal agreement (i.e. temporal adverb and subject) were presented before the main verb. Participants, therefore, were expected to process concordant or discordant agreement with the verb before encountering the verbal critical region. The second region of interest following the verb was important in the case of observing any spill-over effects or effects in very late processing.

To analyze the reading data from these two areas, I developed two separate linear mixed effects models (LMM) using the lme4 package available in R (Bates et al., 2015; R Core Team, 2018). Using a linear mixed effects model allowed me to incorporate all of my data in the analysis and to use data that is difficult to control—data that is non-independent or hierarchical (e.g. individual differences between participants and experimental stimuli). I justify using two separate models in order to answer my separate research questions. One model was used to analyze whether L2 learners demonstrate similar agreement sensitivity to native speakers and included class level (also the native group) as a fixed effect. Because native speakers were not included in my question of whether first language WMC has an effect on verbal agreement processing, I needed a second model removing the native group and adding working memory capacity as a fixed effect. The WMC scores are based only on the L2 learners, not on native speakers. I therefore could not include WMC and class level as fixed effects in a single model because WMC scores would not accurately represent an interaction with class level.

For both models, only the predictor variables and interactions that were significant or approached significance were retained in the model. Predictor variables and interactions were significant at the $p < .05$ level and marginally significant at the $p < .10$ level. Both initial models had random intercepts and random slopes for both participants and items; however, the inclusion of random slopes for both effects did not change the significance of the results. Therefore, the final model removed random slopes and only random intercepts for both participants and items are reported below.

Before presenting the results of each model, I show the results of the raw reading measures. Table 4.1 and Table 4.2 show the means of the raw reading measures at the verb by class level and agreement condition respectively. For class level, each reading measure decreases as class level increases. Learners in lower class levels (i.e. 100 and 200 level classes) demonstrated longer reading times and higher probability of regression than learners in higher class levels (300 level classes) and native speakers. For agreement condition, the means of most reading measures are measurably lower for number anomalies than either person or tense anomalies. However, the exception for this is seen in Figure 4.1. This chart shows the means of the first fixations durations by class level for each agreement condition. It is clear that first fixations were measurably longer for number discord than for the other conditions of agreement. This is true across all class levels indicating that when at all readers' first encounter with the main verb, they were sensitive to number agreement errors but not to person or tense errors. Figure 4.2 and 4.3 show the mean total reading time and fixation count respectively. Notice for these charts that both measures have dropped lower for number errors and have risen for both person and tense errors. This suggests that learners across all levels and natives demonstrated early sensitivity to number errors and were able to address and repair those errors quickly allowing them to spend less total time on the main verb. However, person and tense errors were not noticed until later measures and required individuals to spend more time in repair and reanalysis.

Table 4.1: Reading Time Measures by Class Level

	First fixation (ms)	Gaze duration (ms)	Go-past time (ms)	Total time (ms)	Fixation count	% Regression out
100	295	484	848	1071	3.6	34
200	275	448	803	1019	3.5	37
300 NR	292	430	624	777	2.7	28
300 Res	250	357	517	670	2.5	26
Native	244	315	430	580	2.1	17

Table 4.2: Reading Time Measures by Agreement Condition

	First fixation (ms)	Gaze duration (ms)	Go-past time (ms)	Total time (ms)	Fixation count	% Regression out
Correct	263	602	980	1321	4.9	35
Number	289	489	726	757	2.6	30
Person	275	528	877	1025	3.6	36
Tense	270	499	797	1086	3.9	31

Class Level Model

The first model included the interaction between agreement condition and class level as fixed effects with random intercepts for both participants and items. The residuals for the reading time measures showed a strongly positively skewed distribution. In order to improve the best fit for the model, all reading time measures were log transformed. By taking the natural logarithm of each variable I was able to alter the scale and make the variable approach a normal distribution, thereby creating a better fit for the model. The binary variable, regression out, was analyzed using a logit mixed model. The results of the model are displayed in separate tables below. Each table shows the results of the LMM for each separate reading measure. The far right column of each table shows the adjusted p value

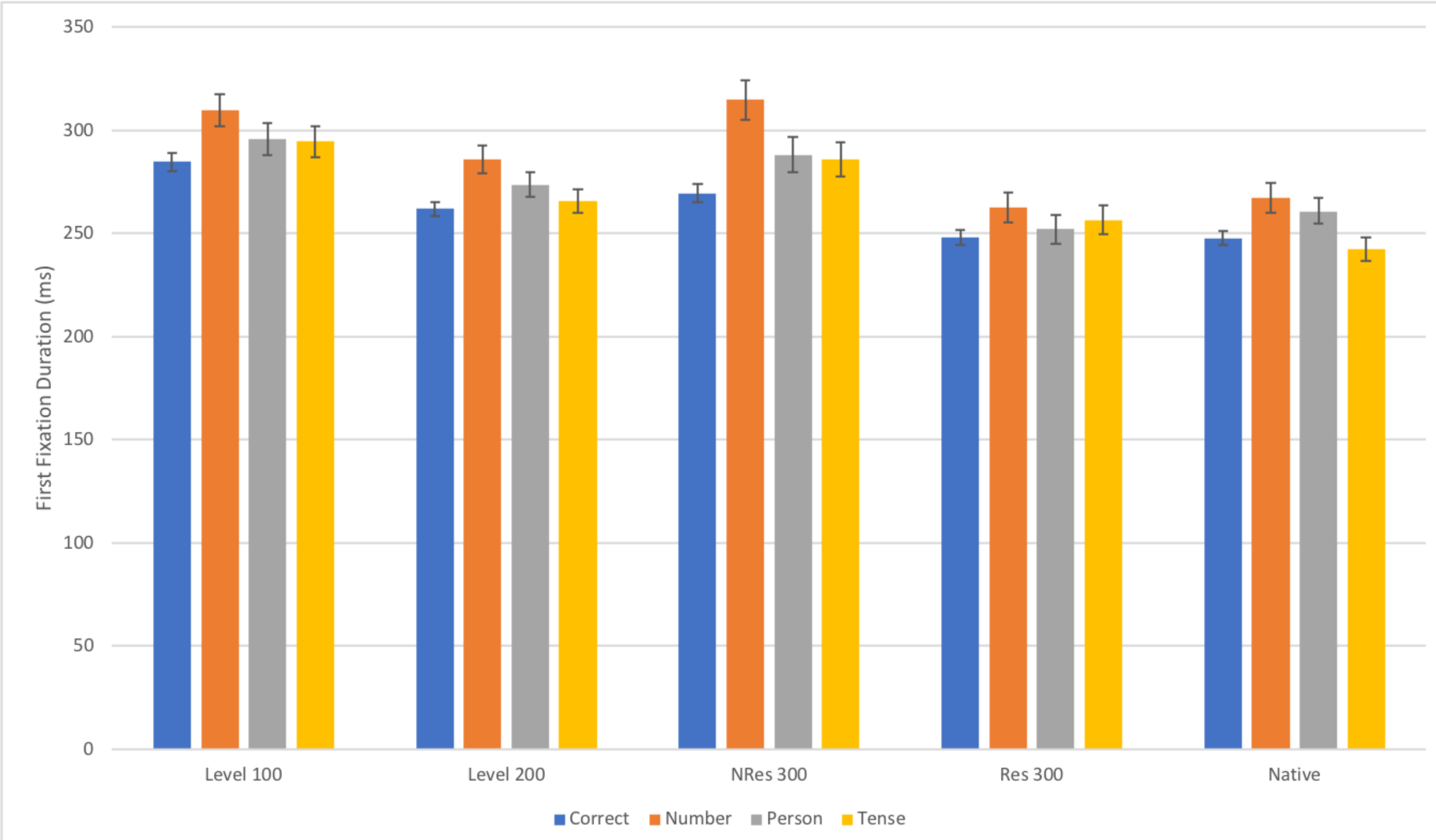

Figure 4.1: First Fixation Duration by Class Level and Agreement Condition

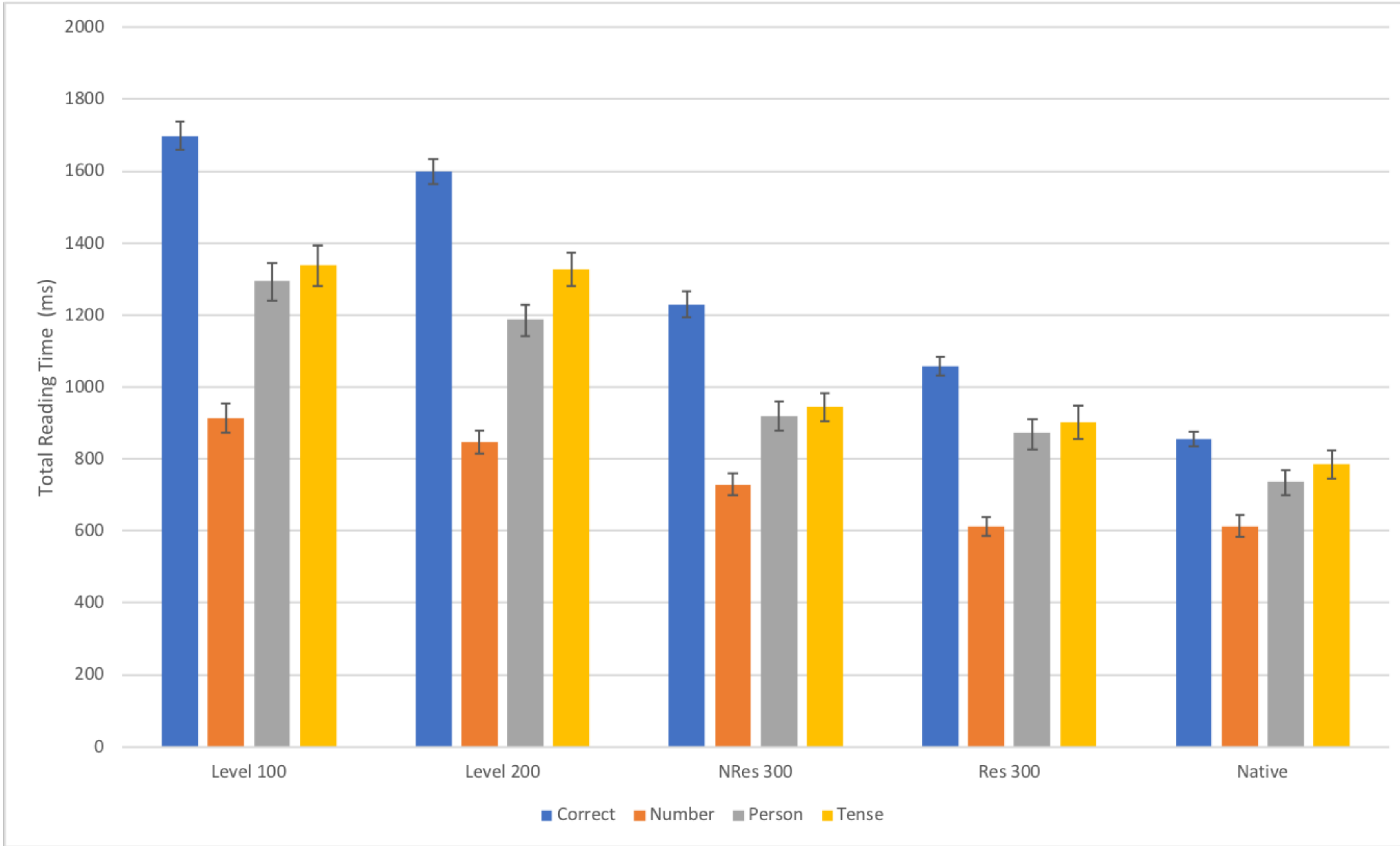

Figure 4.2: Total Reading Time by Class Level and Agreement Condition

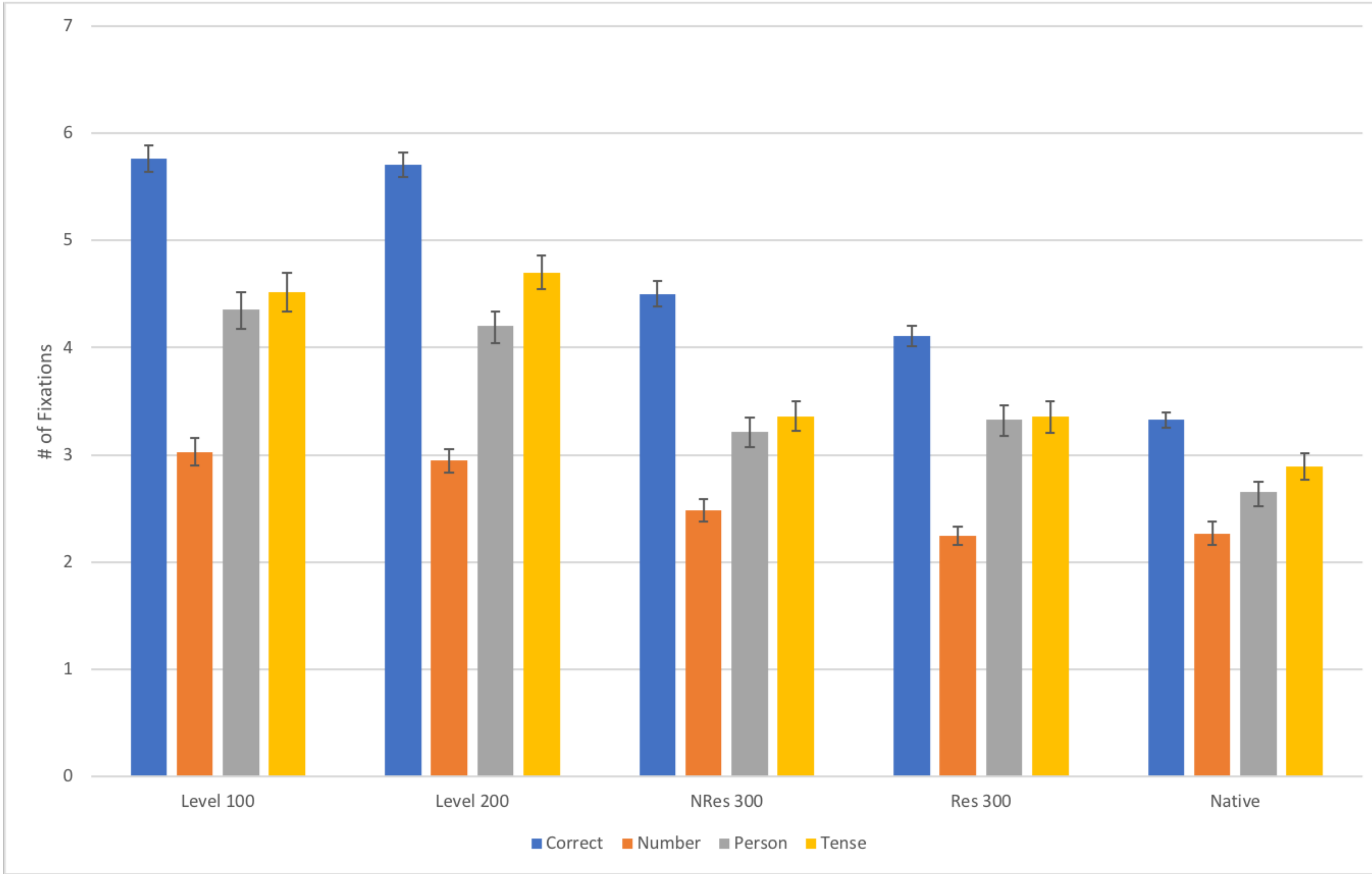

Figure 4.3: Fixation Count by Class Level and Agreement Condition

for the fixed effects and interactions of each reading measure. For several of the measures including first fixation (Table 4.3), total time (Table 4.6), and fixation count (Table 4.7), number mismatches elicited significantly different responses from the correct condition. All pariticipants had measurably longer first fixations for number mismatch than concordant agreement and measurably quicker total time and higher fixation counts for number mismatch than concordant agreement. This is true across all levels of participants, indicating that all individuals show greater sensitivity to number mismatches in early measures (i.e. first fixation duration) and less sensitivity in late measures (i.e. total time and fixation count). This suggests that all learners are quicker to recognize and repair number mismatchs than person or tense mismatches. It is also important to note that for each reading measure there is a lack of significant interaction between the native group and Res 300 group for some of the agreement conditions, indicating some advanced learners do demonstrate similar sensitivity to agreement errors as native speakers.

A multi-comparison analysis of the least square means differences was run for class level and agreement condition for each reading measure. The results of the analysis for each predictor and differences is presented for each reading measure in turn.

First fixation duration. At the first critical region, the verb, both of the predictors (agreement condition and class level) showed a significant effect with condition ($F[3, 237] = 3.08$, $p = 0.0280$) and class level ($F[4, 8902] = 7.44$, $p < .0001$). Each of the discordant agreement conditions showed measurably longer first fixation times than concordant agreement, but only one discordant condition had a main effect. Number discordant condition had significantly longer first fixation durations than concordant agreement ($t[237] = -2.95$, $p = 0.0180$). From the comparison of means differences between each of the class levels, significant differences were found. Notably, the level 100 group reported significantly longer first fixation durations than both the Res 300 group ($t[8902] = 4.38$, $p < .0001$) and the native group ($t[8902] = 4.35$, $p < .0001$). The level 100 group also had marginal significance with slightly longer first fixation times than the level 200 group ($t[8902] = 2.72$, $p = .0510$). The NRes 300

Table 4.3: LMM Results First Fixation Duration - Condition*Class Level

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	5.4340	———	———	———	
Number	0.0640	0.0281	2.276	0.0229	*
Person	0.0507	0.0275	1.841	0.0657	.
Tense	-0.1371	0.0282	-0.486	0.6267	
Level 100	0.1248	0.0561	2.226	0.0287	*
Level 200	0.0470	0.0540	0.871	0.3862	
NRes 300	0.0658	0.0588	1.119	0.2663	
Res 300	0.0071	0.0588	0.122	0.9035	
Number: Level 100	0.0160	0.0369	0.435	0.6637	
Person: Level 100	-0.0269	0.0369	-0.729	0.4660	
Tense: Level 100	0.0409	0.0403	1.015	0.3102	
Number: Level 200	0.0122	0.0355	0.343	0.7315	
Person: Level 200	-0.0002	0.0352	-0.007	0.9945	
Tense: Level 200	0.0146	0.0365	0.401	0.6887	
Number: NRes 300	0.0682	0.0383	1.778	0.0755	.
Person: NRes 300	0.0082	0.0385	0.212	0.8323	
Tense: NRes 300	0.0483	0.0390	1.237	0.2160	
Number: Res 300	-0.0238	0.0387	-0.6140	0.5390	
Person: Res 300	-0.0394	0.0383	-1.0290	0.3036	
Tense: Res 300	0.0351	0.0392	0.8960	0.3701	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

group reported significantly longer first fixation times than two groups: the Res 300 group ($t[8902] = 3.11$, $p = .0159$) and the native group ($t[8902] = -3.06$, $p = 0.0189$).

For the first fixation duration at the verb, participants had longer fixations for number anomalies than for person or tense anomalies. Also, participants in lower class levels and participants without foreign residency fixated on the verb longer than participants with residency and native speakers.

Gaze duration (first run dwell time). The next measure, gaze duration, showed a significant effect for condition ($F[3, 237] = 3.29$, $p = .0214$) and class level ($F[4, 8902] = 27.61$, $p < .0001$) at the main verb. While all discordant conditions showed numerically shorter gaze duration times than the concordant condition, only number discord was marginally significant ($t[237] = 2.46$, $p = .0695$). The comparison of group means also showed significant

Table 4.4: LMM Results Gaze Duration - Condition*Class Level

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	5.7975	———	———	———	
Number	0.0517	0.0436	1.186	0.2359	
Person	0.0734	0.0429	1.710	0.0874	.
Tense	0.0270	0.0440	0.612	0.5403	
Level 100	0.5044	0.0912	5.528	$< .0001$	***
Level 200	0.4128	0.0879	4.696	$< .0001$	***
NRes 300	0.3102	0.0956	3.244	0.0017	**
Res 300	0.1644	0.0956	1.719	0.0894	.
Number: Level 100	-0.0976	0.0526	-1.855	0.0636	.
Person: Level 100	-0.1136	0.0526	-2.158	0.0310	*
Tense: Level 100	-0.0838	0.0576	-1.457	0.1453	
Number: Level 200	-0.0688	0.0507	-1.357	0.1748	
Person: Level 200	-0.0939	0.0503	-1.867	0.0620	.
Tense: Level 200	-0.0915	0.0522	-1.754	0.0795	.
Number: NRes 300	-0.0174	0.0548	-0.318	0.7507	
Person: NRes 300	0.0546	0.0550	0.993	0.3208	
Tense: NRes 300	-0.0022	0.0558	-0.040	0.9683	
Number: Res 300	-0.0121	0.0552	-0.219	0.8268	
Person: Res 300	-0.0274	0.0547	-0.501	0.6166	
Tense: Res 300	0.0114	0.0559	0.204	0.8387	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

differences with the level 100 group showing significantly longer gaze durations than the Res 300 group ($t[8902] = 5.80$, $p < .0001$) and the native group ($t[8902] = 9.30$, $p < .0001$). The level 200 group reported significantly longer gaze durations than the Res 300 group ($t[8902] = 4.34$, $p < .0001$) and the native group ($t[8902] = 7.95$, $p < .0001$). The NRes 300 group reported significantly longer gaze durations than the Res 300 group ($t[8902] = 3.29$, $p = .0089$) and the native group ($t[8902] = -6.58$, $p < .0001$). Lastly, the Res 300 group demonstrated significantly longer gaze durations than the native group ($t[8902] = -3.23$, $p = .0108$).

For gaze duration on the verb, participants did not demonstrate significantly different durations between agreement condition, but there were differences between class level. Low-

Table 4.5: LMM Results Go-past Time - Condition*Class Level

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	6.0080	———	———	———	
Number	0.0697	0.0457	1.526	0.1271	
Person	0.2502	0.0449	5.569	< .0001	***
Tense	-0.0070	0.0461	-0.151	0.8799	
Level 100	0.8525	0.1199	7.108	< .0001	***
Level 200	0.7550	0.1156	6.532	< .0001	***
NRes 300	0.4970	0.1257	3.952	0.0002	***
Res 300	0.3039	0.1257	2.417	0.0179	*
Number: Level 100	-0.3304	0.0548	-6.032	< .0001	***
Person: Level 100	-0.2491	0.0548	-4.543	< .0001	***
Tense: Level 100	-0.1241	0.0599	-2.071	0.0384	*
Number: Level 200	-0.2528	0.0528	-4.789	< .0001	***
Person: Level 200	-0.2389	0.0524	-4.560	< .0001	***
Tense: Level 200	-0.0760	0.0544	-1.398	0.1621	
Number: NRes 300	-0.1129	0.0570	-1.980	0.0477	*
Person: NRes 300	-0.1868	0.0573	-3.261	0.0011	**
Tense: NRes 300	-0.0245	0.0581	-0.422	0.6733	
Number: Res 300	-0.0474	0.0575	-0.825	0.4096	
Person: Res 300	-0.1501	0.0569	-2.637	0.0084	**
Tense: Res 300	-0.0266	0.0583	-0.457	0.6475	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

level learners had longer gaze durations than high-level learners. Even with high-level learners, those without foreign residency had longer durations than learners with residency.

Total time (dwell time). Total reading time at the main verb showed a significant effect for condition ($F[3, 237] = 16.66$, $p < .0001$) and class level ($F[4, 8902] = 22.01$, $p < .0001$). Two discordant conditions reported significantly shorter total reading times than the concordant condition: number discord ($t[237] = 7.04$, $p < .0001$) and tense discord ($t[237] = 2.92$, $p = .0202$). Person discord showed marginally shorter total reading times than the concordant condition ($t[237] = 2.47$, $p = .0675$). The number discordant condition also reported significantly shorter total reading times than the other types of discord: person discord ($t[237] = -4.27$, $p = .0002$) and tense discord ($t[237] = -3.85$, $p = .0009$). The comparison of group means also reported significant differences with the level 100 group

Table 4.6: LMM Results Total Time - Condition*Class Level

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	6.4750	———	———	———	
Number	-0.1811	0.0436	-4.151	< .0001	***
Person	-0.0165	0.0429	-0.383	0.7015	
Tense	-0.0018	0.0440	-0.040	0.9680	
Level 100	0.6625	0.1247	5.312	< .0001	***
Level 200	0.6222	0.1202	5.176	< .0001	***
NRes 300	0.2939	0.1308	2.247	0.0274	*
Res 300	0.1829	0.1308	1.399	0.1658	
Number: Level 100	-0.2805	0.0524	-5.352	< .0001	***
Person: Level 100	-0.0909	0.0525	-1.732	0.0833	.
Tense: Level 100	-0.1575	0.0574	-2.746	0.0060	**
Number: Level 200	-0.2683	0.0505	-5.311	< .0001	***
Person: Level 200	-0.1326	0.0501	-2.645	0.0082	**
Tense: Level 200	-0.0596	0.0520	-1.146	0.2519	
Number: NRes 300	-0.1118	0.0546	-2.050	0.0404	*
Person: NRes 300	-0.0630	0.0548	-1.150	0.2503	
Tense: NRes 300	-0.1195	0.0556	-2.151	0.0315	*
Number: Res 300	-0.1397	0.0550	-2.539	0.0111	*
Person: Res 200	0.0022	0.0545	0.040	0.9684	
Tense: Res 300	-0.0828	0.0557	-1.485	0.1376	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

showing significantly longer total reading times than the NRes 300 group ($t[8902] = 3.53$, $p = .0038$), the Res 300 group ($t[8902] = 5.46$, $p < .0001$), and the native group ($t[8902] = 7.26$, $p < .0001$). The level 200 group reported significantly longer total reading times than the NRes 300 group ($t[8902] = 3.65$, $p = .0025$), the Res 300 group ($t[8902] = 5.65$, $p < .0001$), and the native group ($t[8902] = 7.52$, $p < .0001$). The NRes 300 group reported significantly longer total reading times than the native group ($t[8902] = -3.49$, $p = .0044$).

For total time spent on the verb, learners spent more time reading with sentences that had person and tense anomalies than number anomalies. With class level, a similar trend emerges indicating low-level learners spent more total time on the verb than high-level learners.

Table 4.7: LMM Results Fixation Count - Condition*Class Level

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	3.0592	———	———	———	
Number	-0.5105	0.1846	-2.766	0.0057	**
Person	-0.1019	0.1840	-0.554	0.5798	
Tense	-0.1123	0.1871	-0.600	0.5484	
Level 100	2.3993	0.4749	5.052	< .0001	***
Level 200	2.3323	0.4578	5.095	< .0001	***
NRes 300	1.1135	0.4979	2.236	0.0282	*
Res 300	0.7368	0.4979	1.480	0.1429	
Number: Level 100	-1.6396	0.2237	-7.328	< .0001	***
Person: Level 100	-0.7825	0.2265	-3.455	0.0006	***
Tense: Level 100	-0.8283	0.2468	-3.357	0.0008	***
Number: Level 200	-1.6172	0.2156	-7.502	< .0001	***
Person: Level 200	-0.8446	0.2168	-3.896	0.0001	***
Tense: Level 200	-0.3330	0.2240	-1.486	0.1372	
Number: NRes 300	-0.8697	0.2340	-3.716	0.0002	***
Person: NRes 300	-0.5562	0.2360	-2.357	0.0185	*
Tense: NRes 300	-0.5775	0.2388	-2.418	0.0156	*
Number: Res 300	-0.7277	0.2337	-3.114	0.0019	**
Person: Res 300	-0.1005	0.2356	-0.427	0.6695	
Tense: Res 300	-0.2130	0.2389	-0.892	0.3727	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Go-past time (regression path duration). Go-past time at the main verb showed a significant effect for condition ($F[3, 237] = 5.36$, $p = .0014$) and class level ($F[4, 8902] = 39.52$, $p < .0001$). Two discordant conditions reported significantly shorter go-past reading times than the concordant condition: number discord ($t[237] = 3.45$, $p = .0036$) and tense discord ($t[237] = 3.06$, $p = .0129$). The comparison of group means also reported significant differences with the level 100 group showing significantly longer go-past reading times than the NRes 300 group ($t[8902] = 4.00$, $p = .0006$), the Res 300 group ($t[8902] = 6.70$, $p < .0001$), and the native group ($t[8902] = 10.85$, $p < .0001$). The level 200 group reported significantly longer go-past reading times than the NRes 300 group ($t[8902] = 3.08$, $p = .0175$), the Res 300 group ($t[8902] = 5.88$, $p < .0001$), and the native group ($t[8902] = 10.17$, $p < .0001$). The NRes 300 group reported significantly longer go-past reading times than

Table 4.8: Logit Mixed Model for Regression Out - Condition*Class Level

Fixed effects	Estimate	Std. Error	z value	$Pr(>\|z\|)$	
(Intercept)	-1.9361	———	———	———	
Number	0.2065	0.1957	1.055	0.2912	
Person	0.8544	0.1769	4.831	$< .0001$	***
Tense	-0.1610	0.2095	-0.769	0.4421	
Level 100	1.5836	0.3590	4.411	$< .0001$	***
Level 200	1.8396	0.3454	5.325	$< .0001$	***
NRes 300	1.1130	0.3762	2.959	0.0031	**
Res 300	0.9076	0.3758	2.415	0.0157	*
Number: Level 100	-0.7565	0.2423	-3.122	0.0018	**
Person: Level 100	-0.8975	0.2271	-3.953	0.0001	***
Tense: Level 100	-0.0571	0.2680	-0.213	0.8313	
Number Level 200	-0.7549	0.2308	-3.271	0.0011	**
Person: Level 200	-0.9803	0.2142	-4.577	$< .0001$	***
Tense: Level 200	-0.1447	0.2455	-0.589	0.5557	
Number: NRes 300	-0.4664	0.2523	-1.849	0.0645	.
Person: NRes 300	-1.0555	0.2417	-4.368	$< .0001$	***
Tense: NRes 300	-0.0405	0.2654	-0.153	0.8787	
Number: Res 300	-0.2339	0.2506	-0.933	0.3506	
Person: Res 300	-0.6302	0.2339	-2.694	0.0071	**
Tense: Res 300	0.0549	0.2648	0.207	0.8357	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

the native group ($t[8902]$ = -6.46, $p < .0001$). The Res 300 group also reported significantly longer go-past reading times than the native group ($t[8902]$ = -3.84, p = .0012).

Whereas agreement conditions did not show significantly different go-past times on the verb, class level did demonstrate measurable differences. Again, low-level learners showed markedly longer go-past times than high-level learners.

Fixation count. Fixation count at the main verb showed a significant effect for condition ($F[3, 237] = 23.78$, $p < .0001$) and class level ($F[4, 8902] = 21.14$, $p < .0001$). All discordant conditions reported fewer fixations on the main verb than the concordant condition: number discord ($t[237] = 8.44$, $p < .0001$), person discord ($t[237] = 3.79$, $p = .0011$), and tense discord ($t[237] = 3.64$, $p = .0019$). Number discord also reported significantly fewer fixations than person discord ($t[237]$ = -4.37, $p < .0001$) and tense discord ($t[237]$ = -4.51, p

$< .0001$). The comparison of group means also reported significant differences with the level 100 group showing significantly more fixations on the main verb than the NRes 300 group ($t[8902] = 4.28$, $p = .0002$), the Res 300 group ($t[8902] = 4.68$, $p < .0001$), and the native group ($t[8902] = 7.11$, $p < .0001$). The level 200 group reported significantly more fixations on the main verb than the NRes 300 group ($t[8902] = 4.58$, $p < .0001$), the Res 300 group ($t[8902] = 5.00$, $p < .0001$), and the native group ($t[8902] = 7.53$, $p < .0001$). The NRes 300 group reported marginally significant differences from the native group ($t[8902] = -2.62$, $p = 0.0663$).

Individuals across all levels made significantly more fixations on the verb for sentences with person and tense anomalies than for sentences with number anomalies. Also, low-level learners made numerically fewer fixations on the verb than high-level learners.

Regression out. The logit mixed model for the binary variable regression out reported significance for both fixed effects—agreement condition and class level. Number discord and tense discord demonstrated significantly fewer regressions out of main verb region than the concordant condition: number ($z = -3.43$, $p = .0005$) and tense ($z = -3.20$, $p = .0014$). Both the level 100 group and level 200 group demonstrated significant differences from the native group: level 100 ($z = -3.76$, $p = .0016$) and level 200 ($z = -4.49$, $p < .0001$).

Overall, the comparison of means for agreement condition and class level indicates that both fixed effects are important predictors for most reading measures. This means that the type of agreement anomaly and the class level a learner is in are important factors in determining an individual's sensitivity to agreement anomalies. In particular, person and tense mismatches elicit greater sensitivity from learners than number mismatches. While learners in all class levels did demonstrate sensitivity to agreement anomalies, learners in low-class levels appeared to have greater perturbation than learners in high-class levels as indicated by longer reading times.

WMC Model

The second model included condition and working memory capacity and their interaction as fixed effects with random intercepts for both participants and items. All reading time measures were log transformed and the only binary variable was analyzed using a logit mixed model. While this model does report the F statistic and p values for the condition fixed effect and least square mean differences between conditions, I do not report those values here as they demonstrate a similar significance pattern to the values reported in the first model. The results of the model are shown in the tables below. Each table shows the results of a separate reading measure. The far right column of each table presents the adjusted p value of the model. It is important to notice that in Table 4.9 and Table 4.10, which show the results of first fixation duration ($F[1, 7320] = 3.40$, $p = 0.9936$), and gaze duration ($F[1, 7320] = 2.23$, $p = 0.2732$) respectively, WM Score did not show a significant effect. This indicates that for early reading measures, WMC did not matter as much regarding individuals' sensitivity to different agreement errors. However, Table 4.11, Table 4.12, and Table 4.13 demonstrate that WMC did show a main effect for late measures (i.e. for go-past time ($F[1, 7320] = 11.06$, $p = .0177$), total time ($F[1, 7320] = 12.55$, $p = .0205$), and fixation count ($F[1, 7320] = 9.53$, $p = .0327$), which indicates that higher WMC was facilitatory for individuals during late measures requiring less cognitive effort to identify and repair agreement errors. In regards to specific types of agreement errors, the lack of significant differences between person and tense anomalies with concordant agreement suggests that WMC was more important in identifing and repairing those agreement errors than number agreement errors.

Figure 4.4 and Figure 4.5 suggest that WMC had a greater effect for late measures and for person and tense mismatches compared to early measures and number mismatches. It is important to note in both figures that as WMC increases the difference between number mismatches and concordant agreement gets smaller, whereas the differences for person and tense mismatches remains the same. This indicates that WMC matters for person and tense

Table 4.9: LMM Results First Fixation Duration - Condition*Class Level*WM Score

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	5.5360	———	———	———	
Number	-0.0739	0.0282	-2.624	0.0087	**
Person	0.0137	0.0270	0.506	0.6129	
Tense	-0.0131	0.0263	-0.498	0.6185	
Level 200	-0.0598	0.0489	-1.224	0.2254	
NRes 300	-0.0566	0.0520	-1.088	0.2805	
Res 300	-0.0993	0.0512	-1.938	0.0570	.
WM Score	-0.0003	0.0399	-0.008	0.9936	
Number: Level 200	0.0891	0.0376	2.369	0.0179	*
Person: Level 200	0.0163	0.0358	0.455	0.6488	
Tense: Level 200	0.0480	0.0348	1.380	0.1675	
Number: NRes 300	0.0136	0.0395	0.344	0.7308	
Person: NRes 300	-0.0356	0.0382	-0.932	0.3516	
Tense: NRes 300	-0.0127	0.0372	-0.342	0.7323	
Number: Res 300	0.0696	0.0381	1.828	0.0675	.
Person: Res 300	0.0053	0.0370	0.143	0.8862	
Tense: Res 300	0.0253	0.0361	0.701	0.48302	
Number: WM Score	-0.0221	0.0295	-0.750	0.4530	
Person: WM Score	0.0010	0.0289	0.034	0.9731	
Tense: WM Score	-0.0082	0.0280	-0.293	0.7699	
Level 200: WM Score	-0.0075	0.0498	-0.151	0.8805	
NRes 300: WM Score	-0.0496	0.0508	-0.975	0.3331	
Res 300: WM Score	-0.0548	0.0643	-0.852	0.3972	
Number: Level 200: WM Score	-0.0056	0.0375	-0.149	0.8819	
Person: Level 200: WM Score	-0.0053	0.0365	-0.144	0.8855	
Tense: Level 200: WM Score	0.0197	0.0353	0.559	0.5759	
Number: NRes 300: WM Score	0.0115	0.0379	0.302	0.7624	
Person: NRes 300: WM Score	0.0255	0.0371	0.686	0.4930	*
Tense: NRes 300: WM Score	0.0754	0.0363	2.079	0.0377	
Number: Res 300: WM Score	0.0665	0.0464	1.435	0.1514	
Person: Res 300: WM Score	0.0454	0.0460	0.988	0.3233	
Tense: Res 300: WM Score	0.0504	0.0453	1.113	0.2656	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 4.10: LMM Results Gaze Duration - Condition*Class Level*WM Score

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	6.3330	———	———	———	
Number	-0.0718	0.0445	-1.615	0.1064	
Person	-0.0632	0.0444	-1.423	0.1548	
Tense	-0.0825	0.0443	-1.862	0.0626	.
Level 200	-0.1055	0.0845	-1.249	0.2163	
NRes 300	-0.2185	0.0899	-2.430	0.0179	*
Res 300	-0.3636	0.0886	-4.104	$< .0001$	***
WM Score	-0.0762	0.0690	-1.105	0.2732	
Number: Level 200	0.0507	0.0519	0.977	0.3288	
Person: Level 200	0.0202	0.0517	0.391	0.6957	
Tense: Level 200	0.0256	0.0516	0.497	0.6192	
Number: NRes 300	0.1095	0.0551	1.986	0.0471	*
Person: NRes 300	0.1717	0.0553	3.105	0.0019	**
Tense: NRes 300	0.1041	0.0551	1.889	0.0589	.
Number: Res 300	0.1009	0.0547	1.844	0.0652	.
Person: Res 300	0.0980	0.0543	1.806	0.0710	.
Tense: Res 300	0.0971	0.0544	1.785	0.0743	.
Number: WM Score	0.0894	0.0430	2.082	0.0373	*
Person: WM Score	0.0145	0.0425	0.342	0.7327	
Tense: WM Score	0.0174	0.0426	0.408	0.6834	
Level 200: WM Score	0.1120	0.0860	1.303	0.1974	
NRes 300: WM Score	-0.0013	0.0879	-0.015	0.9880	
Res 300: WM Score	0.0834	0.1112	0.750	0.4562	
Number: Level 200: WM Score	-0.0924	0.0537	-1.721	0.0853	.
Person: Level 200: WM Score	-0.0605	0.0530	-1.143	0.2533	
Tense: Level 200: WM Score	-0.0003	0.0531	-0.005	0.9959	
Number: NRes 300: WM Score	-0.1554	0.0544	-2.857	0.0043	**
Person: NRes 300: WM Score	-0.0334	0.0543	-0.615	0.5386	
Tense: NRes 300: WM Score	-0.0169	0.0542	-0.312	0.7547	
Number: Res 300: WM Score	-0.0301	0.0701	-0.429	0.6678	
Person: Res 300: WM Score	-0.0007	0.0687	-0.010	0.9922	
Tense: Res 300: WM Score	-0.0316	0.0687	-0.460	0.6456	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 4.11: LMM Results Go-Past Time - Condition*Class Level*WM Score

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	6.9360	———	———	———	
Number	-0.2842	0.0458	-6.211	$< .0001$	***
Person	-0.0183	0.0457	-0.400	0.6892	
Tense	-0.1527	0.0456	-3.348	0.0008	***
Level 200	-0.1964	0.1133	-1.734	0.0880	.
NRes 300	-0.4226	0.1206	-3.503	0.0009	***
Res 300	-0.6406	0.1189	-5.389	$< .0001$	***
WM Score	-0.2253	0.0925	-2.437	0.0177	*
Number: Level 200	0.1309	0.0530	2.469	0.0136	*
Person: Level 200	0.0084	0.0529	0.158	0.8741	
Tense: Level 200	0.0475	0.0527	0.900	0.3680	
Number: NRes 300	0.2368	0.0564	4.202	$< .0001$	***
Person: NRes 300	0.0818	0.0565	1.448	0.1477	
Tense: NRes 300	0.0953	0.0563	1.692	0.0907	.
Number: Res 300	0.2951	0.0559	5.277	$< .0001$	***
Person: Res 300	0.1206	0.0555	2.173	0.0298	*
Tense: Res 300	0.0947	0.0556	1.703	0.0886	.
Number: WM Score	0.1036	0.0439	2.361	0.0182	*
Person: WM Score	-0.0020	0.0434	-0.045	0.9641	
Tense: WM Score	-0.0185	0.0435	-0.425	0.6711	
Level 200: WM Score	0.1465	0.1153	1.270	0.2088	
NRes 300: WM Score	0.1763	0.1179	1.495	0.1400	
Res 300: WM Score	0.2769	0.1491	1.856	0.0682	.
Number: Level 200: WM Score	-0.0094	0.0548	-0.172	0.8635	
Person: Level 200: WM Score	-0.0337	0.0541	-0.623	0.5334	
Tense: Level 200: WM Score	0.0256	0.0542	0.472	0.6369	
Number: NRes 300: WM Score	-0.1272	0.0556	-2.289	0.0221	*
Person: NRes 300: WM Score	-0.0340	0.0555	-0.612	0.5404	
Tense: NRes 300: WM Score	-0.0168	0.0554	-0.303	0.7622	
Number: Res 300: WM Score	-0.0811	0.0716	-1.132	0.2575	
Person: Res 300: WM Score	0.0675	0.0702	0.962	0.3359	
Tense: Res 300: WM Score	0.0737	0.0702	1.049	0.2943	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 4.12: LMM Results Total Time - Condition*Class Level*WM Score

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	7.2210	———	———	———	
Number	-0.5153	0.0430	-11.987	$< .0001$	***
Person	-0.1055	0.0429	-2.456	0.0141	*
Tense	-0.1566	0.0429	-3.655	0.0003	***
Level 200	-0.1472	0.1248	-1.180	0.2426	
NRes 300	-0.4439	0.1329	-3.341	0.0014	**
Res 300	-0.5639	0.1309	-4.307	0.0001	***
WM Score	-0.2424	0.1018	-2.380	0.0205	*
Number: Level 200	0.0801	0.0500	1.603	0.1089	
Person: Level 200	-0.0461	0.0498	-0.926	0.3546	
Tense: Level 200	0.0560	0.0497	1.127	0.2598	
Number: NRes 300	0.2153	0.0531	4.054	0.0001	***
Person: NRes 300	0.0245	0.0533	0.460	0.6457	
Tense: NRes 300	0.0383	0.0531	0.720	0.4714	
Number: Res 300	0.1665	0.0527	3.158	0.0016	**
Person: Res 300	0.0800	0.0523	1.528	0.1264	
Tense: Res 300	0.0596	0.0524	1.137	0.2558	
Number: WM Score	0.1740	0.0414	4.208	$< .0001$	***
Person: WM Score	0.0267	0.0409	0.653	0.5138	
Tense: WM Score	0.0391	0.0410	0.952	0.3410	
Level 200: WM Score	0.1468	0.1270	1.156	0.2522	
NRes 300: WM Score	0.1935	0.1298	1.490	0.1414	
Res 300: WM Score	0.1981	0.1643	1.206	0.2326	
Number: Level 200: WM Score	-0.0950	0.0517	-1.839	0.0660	.
Person: Level 200: WM Score	-0.0452	0.0510	-0.886	0.3758	
Tense: Level 200: WM Score	-0.0726	0.0511	-1.421	0.1554	
Number: NRes 300: WM Score	-0.2334	0.0524	-4.455	$< .0001$	***
Person: NRes 300: WM Score	-0.0756	0.0523	-1.446	0.1482	
Tense: NRes 300: WM Score	-0.0245	0.0522	-0.469	0.6390	
Number: Res 300: WM Score	-0.0338	0.0675	-0.501	0.6163	
Person: Res 300: WM Score	-0.0143	0.0661	-0.217	0.8285	
Tense: Res 300: WM Score	0.0036	0.0662	0.055	0.9565	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 4.13: LMM Results Fixation Count - Condition*Class Level*WM Score

Fixed Effects	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	5.7676	———	———	———	
Number	-2.3349	0.2003	-11.655	$< .0001$	***
Person	-0.8577	0.2004	-4.280	$< .0001$	***
Tense	-0.9009	0.2005	-4.494	$< .0001$	***
Level 200	-0.5104	0.5026	-1.015	0.3139	
NRes 300	-1.5386	0.5351	-2.875	0.0055	**
Res 300	-1.9673	0.5272	-3.732	0.0004	***
WM Score	-0.8965	0.4104	-2.185	0.0327	*
Number: Level 200	0.3240	0.2355	1.376	0.1689	
Person: Level 200	-0.0465	0.2356	-0.197	0.8436	
Tense: Level 200	0.2489	0.2355	1.057	0.2906	
Number: NRes 300	0.9129	0.2510	3.637	0.0003	***
Person: NRes 300	0.1826	0.2510	0.728	0.4669	
Tense: NRes 300	0.1592	0.2509	0.634	0.5258	
Number: Res 300	1.0739	0.2471	4.347	$< .0001$	***
Person: Res 300	0.6762	0.2471	2.736	0.0062	**
Tense: Res 300	0.5379	0.2471	2.177	0.0295	*
Number: WM Score	0.6963	0.1935	3.598	0.0003	***
Person: WM Score	0.2076	0.1932	1.075	0.2826	
Tense: WM Score	0.1824	0.1931	0.945	0.3448	
Level 200: WM Score	0.3749	0.5118	0.733	0.4666	
NRes 300: WM Score	1.0874	0.5232	2.078	0.0419	*
Res 300: WM Score	0.6131	0.6621	0.926	0.3581	
Number: Level 200: WM Score	-0.2885	0.2422	-1.191	0.2336	
Person: Level 200: WM Score	-0.2018	0.2414	-0.836	0.4031	
Tense: Level 200: WM Score	-0.3703	0.2413	-1.535	0.1249	
Number: NRes 300: WM Score	-0.7825	0.2464	-3.175	0.0015	**
Person: NRes 300: WM Score	-0.3092	0.2460	-1.257	0.2089	
Tense: NRes 300: WM Score	-0.2114	0.2460	-0.859	0.3902	
Number: Res 300: WM Score	-0.3554	0.3154	-1.127	0.2598	
Person: Res 300: WM Score	-0.2585	0.3136	-0.824	0.4097	
Tense: Res 300: WM Score	0.1694	0.3135	0.540	0.5890	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 4.14: Logit Mixed Model for Regression Out - Condition*Class Level*WM Score

Fixed effects	Estimate	Std. Error	z value	$Pr(>\|z\|)$	
(Intercept)	-0.1778	———	———	———	
Number	-0.4938	0.1601	-3.084	0.0021	**
Person	-0.0929	0.1573	-0.591	0.5546	
Tense	-0.2474	0.1577	-1.569	0.1166	
Level 200	-0.0222	0.3283	-0.068	0.9460	
NRes 300	-0.6422	0.3516	-1.827	0.0678	.
Res 300	-1.0329	0.3480	-2.968	0.0030	**
WM Score	-0.5207	0.2689	-1.937	0.0528	.
Number: Level 200	0.0667	0.2025	0.329	0.7420	
Person: Level 200	-0.0173	0.1981	-0.087	0.9304	
Tense: Level 200	-0.2089	0.2005	-1.042	0.2974	
Number: NRes 300	0.2233	0.2248	0.993	0.3206	
Person: NRes 300	-0.0439	0.2216	-0.198	0.8431	
Tense: NRes 300	-0.0762	0.2232	-0.341	0.7330	
Number: Res 300	0.5573	0.2248	2.479	0.0132	*
Person: Res 300	0.3797	0.2212	1.716	0.0861	.
Tense: Res 300	0.0894	0.2296	0.389	0.6970	
Number: WM Score	-0.0875	0.1813	-0.483	0.6294	
Person: WM Score	0.0184	0.1687	0.109	0.9134	
Tense: WM Score	-0.1134	0.1729	-0.655	0.5122	
Level 200: WM Score	0.2647	0.3346	0.791	0.4289	
NRes 300: WM Score	0.5634	0.3435	1.640	0.1009	
Res 300: WM Score	1.1192	0.4419	2.533	0.0113	*
Number: Level 200: WM Score	0.3401	0.2178	1.561	0.1185	
Person: Level 200: WM Score	0.0175	0.2060	0.085	0.9322	
Tense: Level 200: WM Score	0.0295	0.2107	0.140	0.8887	
Number: NRes 300: WM Score	0.2236	0.2299	0.973	0.3308	
Person: NRes 300: WM Score	-0.1005	0.2200	-0.457	0.6477	
Tense: NRes 300: WM Score	-0.1945	0.2248	-0.865	0.3869	
Number: Res 300: WM Score	-0.3555	0.3156	-1.126	0.2601	
Person: Res 300: WM Score	0.1267	0.2976	0.426	0.6703	
Tense: Res 300: WM Score	0.4658	0.3115	1.496	0.1348	

Significance values: . $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

agreement errors but not for number agreement errors. Individuals with higher WMC have an advantage in identifying and repairing person and tense errors quicker than individuals with lower WMC. However, this advantage is not seen with number agreement errors.

Spill-over Effects

In order to account for very late processing effects, or spill-over effects (Jiang, 2007; Jiang et al., 2011), I created two more linear mixed effects models to anaylyze the same reading behavior variables on the critical region following the main verb. These two models were similar to the previous two explained (i.e. the class level model and WMC model). The first model included condition and class level and their interaction as fixed effects with participants and items having random intercepts. The second model included condition and working memory capacity and their interaction as fixed effects with participants and items having random intercepts. Similar to the models above, the initial models for the spill-over area had random intercepts and random slopes for both participants and items; however, the inclusion of random slopes for both effects did not change the significance of the results. Therefore, the final models for the spill-over area removed random slopes and only random intercepts for both participants and items were included in the models .For both models, all reading time measures were log transformed and a logit analysis was conducted for the binary variable.

Class level reported a main effect on the critical region following the verb for several variables. For first fixation duration $F[4, 8368] = 3.19$, $p = .0126$, as class level increased, first fixations times decreased. For gaze duration ($F[4, 8368] = 2.47$, $p = .0427$), as class level increased, gaze durations decreased. For go-past reading time ($F[4, 8368] = 12.15$, $p < .0001$), as class level increased, go-past reading times decreased. For regression out ($F[4, 8368] = 3.36$, $p = .0093$), as class level increased, the likelihood of regressing out decreased. Each of these main effects demonstrate significance in very late processing. Participants in

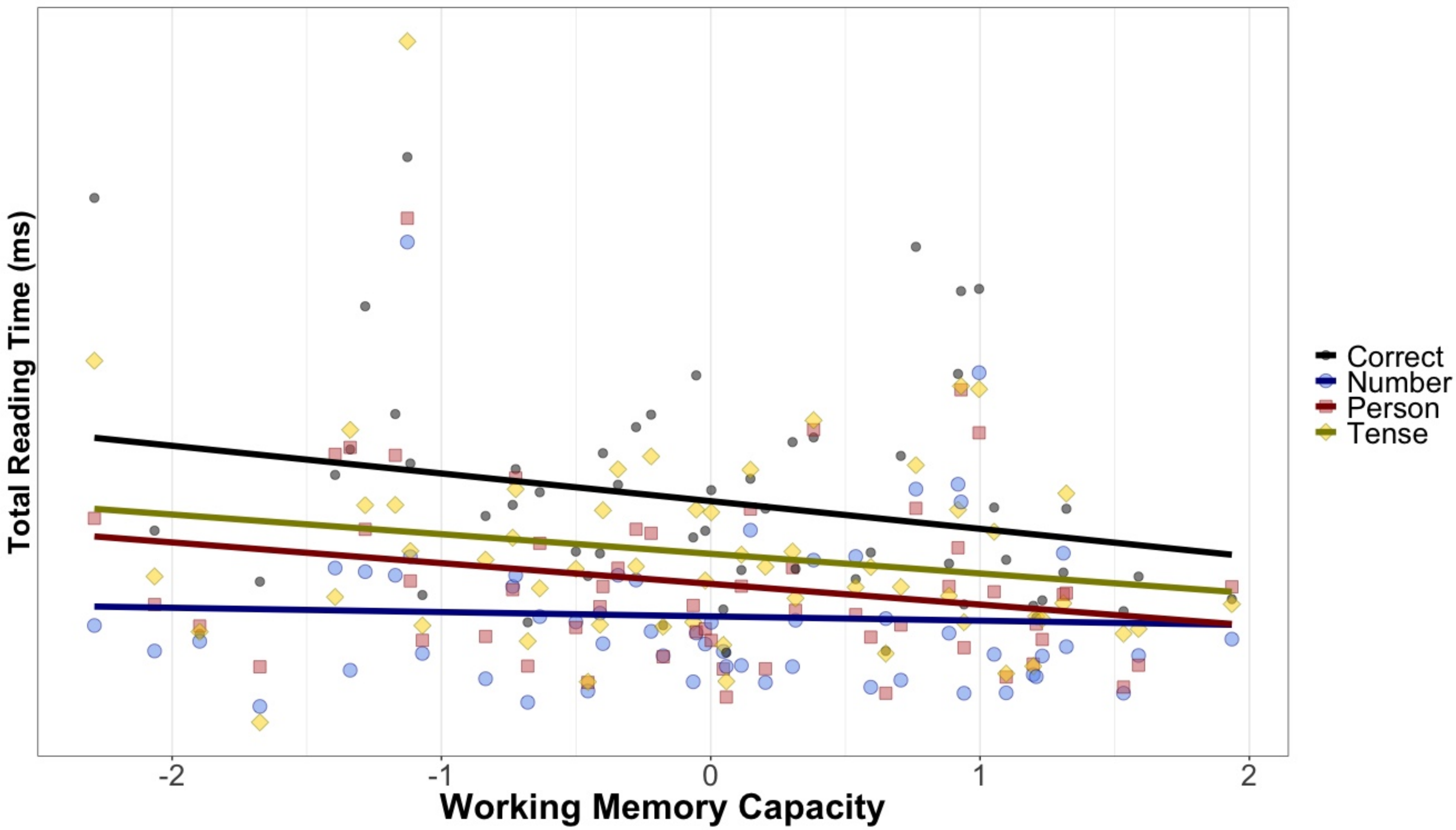

Figure 4.4: Total Reading Time by Working Memory Span

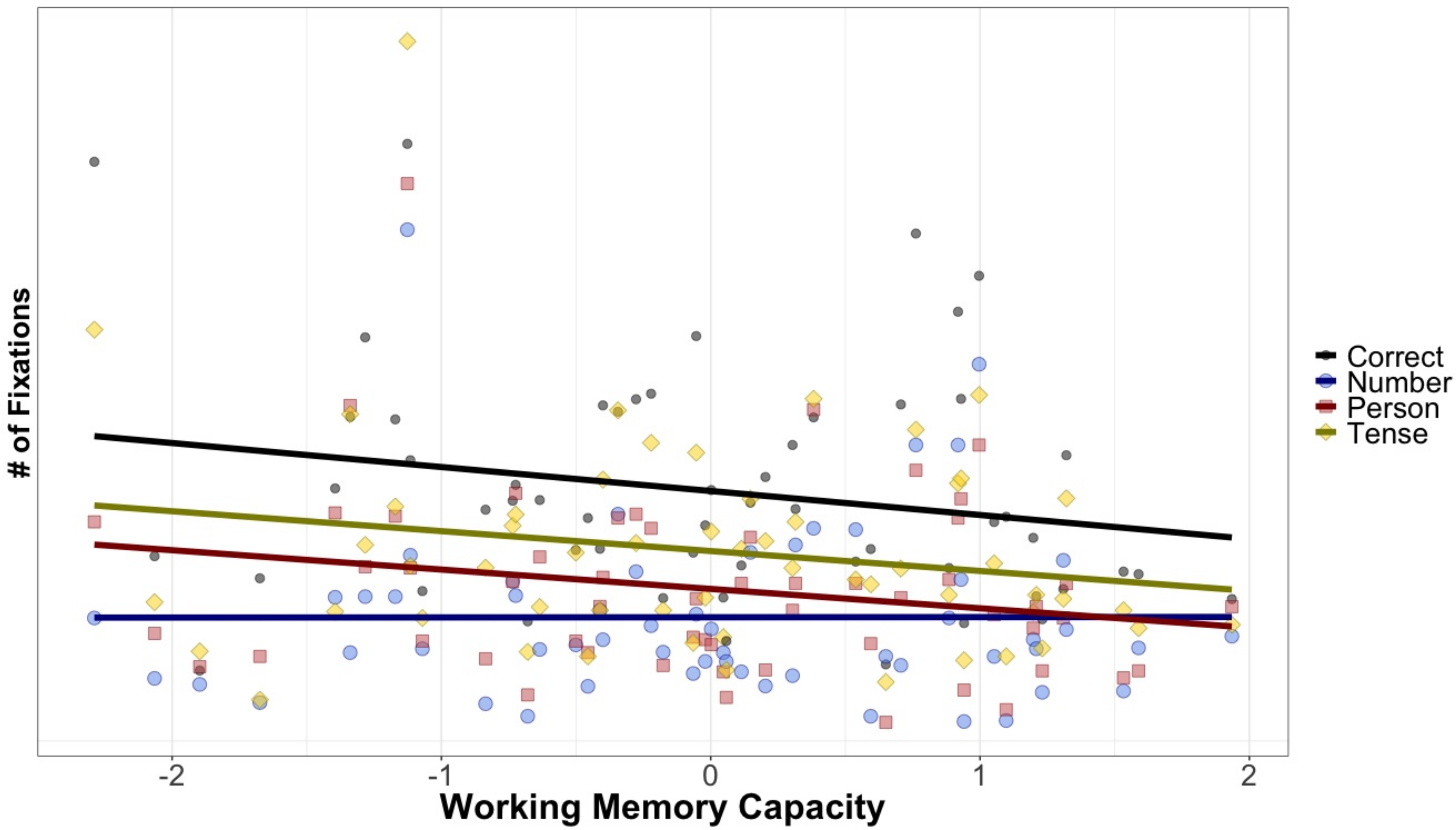

Figure 4.5: Fixation Count by Working Memory Span

lower class levels (primarily level 100) spend more time rereading and regressing to the main verb than higher class levels.

Working memory did have significant effects in the critical region following the verb which may also indicate very late processing. For only two of the variables did working memory capacity show a main effect; go-past reading times ($F[1, 6816] = 8.78$, $p = .0031$) and regression out ($F[1, 6816] = 11.42$, $p = .0007$). These effects indicate that as working memory capacity increased, time spent rereading and likelihood of regression to the main verb decreased. Similar to the WMC model presented earlier, WMC did not appear to matter as much in regards to recognized and repair number mismatches. Several reading measures did not change in response to number anomalies for individuals with different capacities of working memory, whereas those reading measures did change in response to person and tense anomalies. That is, as WMC increased, individuals across all levels of WMC exhibited similar reading times and fixations in response to number mismatches, but individuals with higher WMC demonstrated increasingly quicker reading times and fewer fixations in response to person and tense mismatches. Similar to what was stated above, this indicates that WMC may be more important in recognizing and repairing person and tense agreement violations than number agreement violations.

Chapter 5

Discussion

Given the prior research in second language agreement processing, this thesis has investigated four separate research questions. Each will be discussed in turn below.

Native-like Processing

Clahsen and Felser (2006b) introduced a question in their seminal paper: How native-like is non-native processing? I addressed the same question in this thesis, but more specifically: Do L2 learners of Spanish demonstrate similar sensitivity to agreement errors as native Spanish speakers? Prior studies have responded extensively to this question and have found disparate results. I expected learners at the advanced levels of Spanish learning to be able to show native-like patterns of sensitivity to agreement anomalies.

As expected, the results of the mixed model and comparison of means suggests L2 learners do exhibit similar sensitivity to verbal agreement errors as native speakers. The lack of significant interactions from the class level model indicates that low and high-level learners respond to agreement violations in a similar pattern to native Spanish speakers, with longer first fixations and quicker total time for number agreement violations compared to person and tense agreement violations. While the reading measures for low-level learners indicate longer reading times overall as compared to the native group, the lack of significant interactions between the learner groups, native speakers, and agreement condition indicate similar sensitivity to the different types of agreement anomalies. Such a finding provides contrary evidence to earlier studies which found that L2 learners do not demonstrate native-like patterns of agreement sensitivity (Jiang, 2004; Jiang et al., 2011; Keating, 2009). Unlike these prior studies, this thesis suggests that L2 learners can attain native-like behaviors in agreement processing, meaning they do integrate L2 morphological knowledge. Because only the advanced level group with residency (i.e. Res 300) showed consistent patterns of native-like sensitivity, the thesis also provides subtle support to the observation that foreign residency affects learners' ability to process morphological phenomena in a second language.

Class Level Differences

The second question this thesis addressed was whether class level made a difference in a learner's sensitivity to agreement errors. This question has similarly been addressed by other studies, but more particularly those studies question proficiency and not just class level. Here I do not assume class level and proficiency are similar and do not treat the results as if proficiency were the factor in the mixed model. Instead, I only address class level as the factor and make loose correlations to other studies that have addressed proficiency.

As expected with reading time, learners in higher university levels of Spanish showed dissimilar patterns of sensitivity to agreement errors versus learners in lower levels. The results show that individuals in every level of Spanish did show sensitivity to agreement errors, as seen with the significantly different reading times between conditions with correct agreement and those with an agreement anomaly (particularly number agreement). However, low-level learners showed quantitatively different reading measures than high-level learners. Overall, high-level learners exhibited quicker reading time measures and fewer regressive behaviors than low-level learners and each of these measures decreased as class level increased. This was expected and converges with the results of other L2 studies that as class level or proficiency increase, the ability to read and process second language increases as seen in quicker reading measures.

While high-level learners read through the sentences significantly quicker than low-level learners, the interaction between these groups and agreement condition was significant, but only in late measures (i.e. total time and fixation count). Low-level learners exhibited significantly first fixation durations for number agreement violations, similar to high-level learners. This indicates that learners at all levels detect number errors similarly during early measures, but respond to the errors differently during late measures. As stated above, all learner levels responded to agreement errors in a similar pattern. What is important to note is the effect of strategic task processes in this experiment. All class levels, including natives, demonstrated behavior consistent with the objective of the task—to correctly respond to the grammatical judgment of each sentence. Therefore the results indicate measurably longer reading times for concordant conditions than for any of the mismatch conditions. This suggests that when individuals did not see an agreement error in grammatical sentences, they would reread or fixate longer on the verb to make sure there was no agreement error. Learners in the 100 and 200 class levels were particularly susceptible to this strategic task process, and this is where a significant interaction is observed between the groups. The difference in total reading time between the correct condition and number mismatch becomes less significant as class level increases, suggesting that low-level learners are less confident in identifying errors compared to the high-level learners. This may also suggest that in late reading measures, low-level learners take longer to recognize and repair agreement violations of each feature type compared to high-level learners.

Working Memory Capacity

The third question addressed on the research agenda was what effect working memory had on L2 agreement processing. Whereas prior studies have largely investigated effects of such individual differences as WMC on certain L2 cognitive tasks (e.g. sentence and morphological processing), this thesis examined more specifically the role of WMC during the time course of reading of L2 verbal agreement. I was particularly interested in whether WMC played a greater role in early processing than late processing of agreement. Eye tracking methodology provided a closer examination of such early/late measures.

I expected WMC would show an overall effect on agreement processing and particularly with early measures (i.e. first fixation and gaze duration). Early measures indicate lexical access, including access of grammatical or morphosyntactic information. While WMC did show an overall measurable effect on agreement processing, the effect was only found with late measures of reading, opposite of what was initially expected.

The participants with greater WMC in this experiment exhibited quicker reading time measures overall, and less perturbation to agreement errors. Early measures of online processing like first fixation or gaze duration did not appear to matter as much in terms of WMC being an advantage during processing. However, higher WMC did appear to be an advantage during late processing as indicated by total reading time and fixation counts. The less time spent reading in the critical regions and the lower probability for regressing off the verb and area after the verb indicate that participants with higher WMC spent less time rereading and had a lower probability to do so. This can be interpreted that those with higher WMC incur less processing costs and repairs when processing agreement violations.

Regarding the interaction between WMC and type of agreement violation, the results indicated WMC matters more for tense and person violations than it does for number violations. Regardless of WMC, individuals demonstrated similar reading behaviors to number violations—quicker overall reading times and less probability of regression. However, the difference between number violations and person and tense violations is measurably significant when comparing across WMC. Individuals with lower WMC demonstrated significant differences between number violations and person and tense violations, whereas individuals with higher WMC did not exhibit such significant differences. This indicates that WMC is facilatatory in detecting person and tense violations but not so much in detecting number violations. Individuals with higher WMC have an advantage in detecting and repairing person and tense violations more quickly compared individuals with lower WMC.

Furthermore, although the model included class level instead of proficiency as a fixed effect, the lack of a significant interaction between WMC and class level indicates WMC may not significantly covary with class level. It is not necessarily the case that WMC increases with class level. This loosely suggests that Coughlin and Tremblay's (2013) interpretation that WMC only shows an effect for high-proficiency learners is incomplete.

Feature Differences

The final question addressed in this thesis concerned the effect of feature differences on individuals' sensitivity to agreement errors. Using three Spanish verbal agreement features, the study examined whether learners exhibited different eye behavior in response to agreement errors of features expressed by a single exponent (i.e. cumulative exponence). The results of the LMM indicated a main effect of feature for several reading variables, suggesting different feature errors elicit different sensitivity from learners. Overall, the model showed that learners were less sensitive to number agreement errors than to person or tense agreement errors. In most every reading measure, learners spent less time reading and rereading sentences with number anomalies than either other feature anomaly. Such results suggest that learners were less perturbed by number anomalies than by person or tense anomalies.

All levels, including native speakers, demonstrated similar sensitivity to the types of agreement violations, where greater sensitivity was elicited by person and tense violations than by number violations. For first fixation duration there was a significant interaction between class level and number disagreement, but there were no significant interactions between any agreement condition and class level for the other reading measures. It is interesting to note that number errors elicited longer first fixations but shorter total reading time from all individuals, regardless of class level. This may indicate that at very early reading measures, number errors are more salient and cause greater perturbation than tense or person errors. The quicker reading times for number during late stages of processing suggest that the intital sensitivity to number errors is repaired quickly, and over the time course of reading number errors require less repair and reanalysis costs.

Whereas these results show disparate effects from some studies of feature differences in agreement processing (Barber and Carreiras, 2005; Romanova and Gor, 2017), they do converge with Mancini et al.'s (2014; 2011) results, particularly with the person/number distinction in agreement processing. I adopt a similar explanation to Mancini et al.'s (2014) in this thesis. Person violations in sentences generally cause learners greater perturbation than number because person violations disrupt the perspective of the sentence, whereas number violations only disrupt the number represented by the subject.

What is of particular importance from this thesis is the finding that tense agreement violations elicit similar responses from individuals as person violations. As of the writing of this thesis, no other study has compared individuals' sensitivity to tense violations in Spanish agreement, specifically when comparing person and number violations. The results indicated that tense violations elicit more sensitivity than number violations. Using a similar explanation as person violations, I attribute this difference between tense and number to the disruption of the temporal perspective of the sentence. Violating the temporal perspective of the sentence causes greater disruption than violating number agreement with the subject. In this regard, tense and person violations were costlier and required greater repair as seen in longer reading times and greater probability of regression.

Conclusion

This thesis of L2 processing of Spanish shows that advanced English learners of Spanish do show quantitatively similar patterns of sensitivity to Spanish verbal agreement errors as native Spanish speakers. Similar to other studies (Foote, 2011; Sagarra and Herschensohn, 2010), this suggests that L2 speakers can attain native-like levels of processing and do integrate L2 morphological knowledge during reading comprehension at least.

The study also shows that class level differences in agreement processing do exist and that there is a pattern of emergence of sensitivity to agreement errors. While beginners do recognize errors qualitatively similar to intermediate and advanced learners (i.e. are more perturbed by person and tense errors than by number errors), their eye behaviors indicate significantly slower reading times. This may indicate greater processing and repair costs for beginners and the repair costs lessen as learners become more advanced in the language.

In regards to WMC, the study showed that WMC becomes important for late processing. Individuals with higher WMC appear less perturbed by agreement errors and spend less time reading due to lower repair and reanalysis costs.

Converging with the results of Mancini et al. (2014), feature errors expressed by cumulative exponence do elicit different sensitivity from learners, specifically, person and tense anomalies cause learners greater perturbation, which results in greater repair costs and reanalysis.

Although this thesis provides more evidence in emerging areas of agreement processing research, it demonstrates the need for further investigation into feature type and exponence roles in L2 processing. The study also highlights the need for further research on the effects that individual variables (e.g. WMC) have on L2 agreement processing.

References

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10):829–839.

Baddeley, A. (2007). *Working memory, thought, and action.* Oxford University Press, New York, NY.

Barber, H. and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1):137–153.

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

Bates, E., Devescovi, A., Hernandez, A., and Pizzamilgio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, 58(7):992–1004.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56(1):9–49.

Bond, K., Gabriele, A., Fiorentino, R., and Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In Tanner, D. and Herschensohn, J., editors, *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference*, pages 17–29, Somerville, MA. Cascadilla Proceedings Project.

Bosch, S. and Clahsen, H. (2016). Accessing morphosyntax in L1 and L2 word recognition: A priming study of inflected German adjectives. *The Mental Lexicon*, 11(1):26–54.

Clahsen, H. and Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(1):107–126.

Clahsen, H. and Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1):3–42.

Clifton, C. J., Staub, A., and Rayner, K. (2007). Eye movements in reading words and sentences. In van Gompel, R. P. G., Fischer, M. H., Murray, W. S., and Hill, R. L., editors, *Eye movements: A window on mind and brain*, pages 341–371. Elsevier, Amsterdam.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, 12(5):769–786.

Corbett, G. G. (1998). Morphology and agreement. In Spencer, A. and Zwicky, A. M., editors, *The Handbook of Morphology*, pages 191–205. Blackwell Publishers Ltd., UK.

Coughlin, C. E. and Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, 34(3):615–646.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, 20(4):659–678.

Dai, Y. (2015). Working memory in L2 sentence processing: The case with relative clause attachment. In Wen, Z., Mota, M. B., and McNeill, A., editors, *Working memory in second language acquisition and processing*, pages 105–124. Multilingual Matters, Bristol, UK.

Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. Routledge, New York, NY.

DeKeyser, R. (2005). What makes learning a second-language grammar difficult? A review of issues. *Language Learning*, 55:1–25.

DeKeyser, R. and Larson-Hall, J. (2005). What does the critical period really mean? In Kroll, J. and de Groot, A. M. B., editors, *Handbook of bilingualism*, pages 88–108. Oxford University Press, Oxford, UK.

Felser, C. and Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23(1):9–36.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, 32(1):187–220.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., and Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory and Cognition*, 43(2):226–236.

Foucart, A. and Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 66(1):226–248.

Fraga, I., Padrón, I., Acuña Fariña, C., and Días-Lago, M. (2017). Processing gender agreement and word emotionality: New electrophysiological and behavioural evidence. *Journal of Neurolinguistics*, 44(1):203–222.

Friedman, N. P. and Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, 37(4):581–590.

Gabriele, A., Fiorentino, R., and Alemán Bañó, J. (2013). Examining second language development using event-related potentials: A cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism*, 3(2):213–232.

Gillón-Dowens, M., Barber, H., Guo, J., Guo, T., and Carreiras, M. (2008). ERP correlates of automatic syntactic processing in highly proficient L1 Chinese–L2 Spanish late learners. In *Paper presented at the Cognitive Neuroscience Society Annual Meeting*, University of California, Davis.

Grosjean, F., Dommergues, J., Cornu, E., Guillelmon, D., and Besson, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics*, 56(5):590–598.

Guillelmon, D. and Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition*, 29(3):503–511.

Haspelmath, M. and Sims, A. D. (2010). *Understanding morphology.* Hodder Education, London, UK.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., and Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, 59(1):73–112.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford University Press, Oxford, UK.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22(3):369–397.

Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing.* PhD thesis, Rijksuniversiteit Groningen.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between nonnative and native speakers. *Lingua*, 120(4):901–931.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., and Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5):1217–1232.

Isabelli-García, C., Bown, J., Plews, J. L., and Dewey, D. P. (2018). Language learning and study abroad. *Language Teaching*, 51(4):439–484.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25(4):603–634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57(1):1–33.

Jiang, N., Novokshanova, E., Masuda, K., and Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61(3):940–967.

Juffs, A. (2015). Working memory and sentence processing: A commentary. In Wen, Z., Mota, M. B., and McNeill, A., editors, *Working memory in second language acquisition and processing*, pages 125–135. Multilingual Matters, Bristol, UK.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59(3):503–535.

Lim, J. H. and Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(6):1283–1315.

Lopez Prego, B. (2012). *Examining the nature of variability in gender and number agreement in native and non-native Spanish*. PhD thesis, University of Kansas.

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: Discourse involvement in subject-verb agreement computation. *Brain Research*, 1410:64–76.

Mancini, S., Postiglione, F., Laudanna, A., and Rizzi, L. (2014). On the person-number distinction: Subject-verb agreement processing in Italian. *Lingua*, 146(1):28–38.

Mancini, S., Quiñones, I., Molinaro, N., Hernandez-Cabrera, J. A., and Carreiras, M. (2017). Disentangling meaning in the brain: Left temporal involvement in agreement processing. *Cortex*, 86:140–155.

Marull, C. (2017). *Processing (in)efficiency in the second language: Linguistic experience and cognitive effects on morphosyntactic predictions*. PhD thesis, Rutgers University.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3):381–401.

Molinaro, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47:908–930.

Molinaro, N., Vespignani, F., and Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, 1228:161–176.

Pickering, M. (1999). Sentence comprehension. In Garrod, S. and Pickering, M., editors, *Language processing*, pages 123–153. Psychology Press, Hove, East Sussex.

Qing, C., Lassiter, D., and Degen, J. (2018). What do eye movements in the visual world reflect? A case study of adjectives. In Rogers, T., Rau, M., Zhu, X., and Kalish, C. W., editors, *Proceedings of CogSci*, pages 2297–2302, Austin, TX. Cognitive Science Society.

R Core Team (2018). R: A language and environment for statistical computing. Available at https://www.R-project.org/.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8):1457–1506.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., and Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28(3):164–171.

Reichle, R., Tremblay, A., and Coughlin, C. (2016). Working memory capacity in L2 processing. *Probus*, 28(1):29–55.

Romanova, N. and Gor, K. (2017). Processing of gender and number agreement in Russian as a second language. *Studies in Second Language Acquisition*, 39(1):97–128.

Sagarra, N. and Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120(8):2022–2039.

Sagarra, N. and Herschensohn, J. (2011). Asymmetries in gender and number agreement processing in late bilinguals. In *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, pages 169–177.

Sagarra, N. and LaBrozzi, R. (2018). Benefits of study abroad and working memory on L2 morphosyntactic processing. In Sanz, C. and Morales-Front, A., editors, *The Routledge handbook of study abroad research and practice*, pages 149–164. Taylor and Francis, New York, NY.

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

Tagarelli, K. M., Mota, M. B., and Rebuschat, P. (2015). Working memory, learning conditions and the acquisition of L2 syntax. In Wen, Z., Mota, M. B., and McNeill, A., editors, *Working memory in second language acquisition and processing*, pages 224–247. Multilingual Matters, Bristol, UK.

Tokowicz, N., Michael, E. B., and Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, 7(3):255–272.

Tokowicz, N. and Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, 22(7):1092–1106.

Tuninetti, A., Warren, T., and Tokowicz, N. (2015). Cue strength in second-language processing: An eye-tracking study. *The Quarterly Journal of Experimental Psychology*, 68(3):568–584.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30(1):37–69.

Walker, I. and Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5):1256–1271.

Waters, G. S. and Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology: Section A*, 49(1):51–79.

Yang, J. S. (2016). The effectiveness of study-abroad on second language learning: A meta-analysis. *Canadian Modern Language Review*, 71(1):66–94.

Zeyrek, D. and Acarturk, C. (2014). The distinction between unaccustative and unergative verbs in Turkish: An offline and an eye tracking study of split ergativity. In Bello, P., Guarini, M., McShane, M., and Scassellati, B., editors, *Proceedings of CogSci*, pages 1832–1837, Austin, TX. Cognitive Science Society.