GRAMMATICAL GENDER AGREEMENT IN L2 SPANISH:
THE ROLE OF SYNTACTIC CONTEXT

By

Le Anne L. Spino-Seijas

A DISSERTATION

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

Second Language Studies—Doctor of Philosophy

2017

ABSTRACT

GRAMMATICAL GENDER AGREEMENT IN L2 SPANISH:
THE ROLE OF SYNTACTIC CONTEXT

By

Le Anne L. Spino-Seijas

A pervasive question in second language (L2) research is whether L2 learners can acquire parameterized functional features that are not instantiated in their first language (L1). While some researchers have argued for a representational deficit (e.g., Clahsen & Muysken, 1989; Hawkins & Chan, 1997), claiming that L2 learners' competence is fundamentally deficient, others have argued that learners can indeed acquire features that are not instantiated in their L1 (e.g., Prévost & White, 2000; Schwartz & Sprouse, 1996), and ascribe any optionality to communication pressure or other external factors. In this dissertation, grammatical gender agreement was used as a test case to determine if L2 Spanish learners (L1 = English) can indeed acquire a parametrized feature not present in their L1.

Many researchers that investigate grammatical gender agreement do not have a principled reason for investigating a particular type of agreement (e.g., determiner-noun, noun-adjective). It may be the case, though, that not all types of agreement are equally difficult for L2 learners. In studies that investigate whether L2 learners have a representational deficit for grammatical gender agreement, it is therefore impossible to conclude whether learners truly have a representational deficit, or whether they are performing poorly because of the type of agreement under investigation. Therefore, this dissertation tests grammatical gender agreement in three different syntactic contexts that are commonly used in this type of research: determiner-noun (DET-N), noun-adjective (N-ADJ) and null nominal (N-DROP). These syntactic contexts were

hypothesized to differ in difficulty for L2 learners, with DET-N being the easiest and N-DROP the most difficult.

Native Spanish speakers and L2 learners read a series of sentences embedded with violations of these three different types of grammatical gender agreement while their eye-movements were recorded with an eye-tracker. Participants' sensitivity was measured both via reading times and self-reports on a post-reading questionnaire. Linear mixed-effects models indicated that native Spanish speakers were sensitive to all three types of grammatical gender agreement, as evidenced by longer reading times on ungrammatical relative to grammatical areas of interest, but L2 Spanish learners were sensitive only to DET-N agreement, and not N-ADJ and N-DROP agreement. The self-reports paralleled these findings, with L2 learners reporting a higher instance of seeing DET-N agreement violations than N-ADJ and N-DROP violations in the experimental stimuli. These results indicate that the L2 learners likely do not have a representational deficit for grammatical gender agreement, and that the type of grammatical gender agreement under investigation matters, as the syntactic context of the agreement may affect performance. Results are discussed in terms of the types of knowledge L2 learners use during online processing in studies that detect sensitivity to grammatical violations.

In loving memory of Philip P. Spino
best of fathers and best of men

ACKNOWLEDGEMENTS

I am indebted to so many people for their assistance and guidance as I navigated my way through this degree. The following is my humble attempt to express my gratitude for those who have helped me along the way.

First, I'd like to thank Bill VanPatten for being such a limitless source of knowledge. To say that his outlook on language research and instruction has been instrumental in my career is an understatement. Among other things, I am thankful for all the time he has spent going over my research with me, for his patience in showing me how to most effectively create class activities, and for the many, many times he has made me laugh. I am also greatly indebted to Aline Godfroid for long and productive conversations about research, academia, and life in general. My interactions with Aline have undoubtedly made me a better researcher, a better writer, and a more thoughtful person. I have so many fond memories of our time together in Michigan, and I am privileged to count her as a friend. I also extend my thanks to Patti Spinner and Paula Winke for agreeing to being part of this project. They have both provided me with friendly conversation, thoughtful feedback, and words of wisdom over the years.

In my estimation, the Second Language Studies program at Michigan State University has professors that are not only incredibly talented, but that also care deeply about their students and go out of their way to help them succeed. I am thankful to not only my committee members, but also to many other professors who have taken time from their own research to plan interesting classes, help students, and answer their queries.

So many of my Michigan friends are also very deserving of my gratitude. A special thank you to Megan Smith for being my café work buddy, sounding board, and official baker; to Daniel Trego for always lending me a helping hand and being the other half of our "dream team"; to Luca Giupponi for being my photographer and craft beer connoisseur; to Angelika Kraemer for always keeping me entertained; and to Walter Hopkins for making every day brighter with his infectious laugh. My time in Michigan wouldn't have been the same without you all.

I am grateful for my colleagues at Princeton University as well. I am particularly thankful to my *compinche*, Anais Holgado Lage, for giving me a place to stay and keeping me company, especially during long commutes. I would also like to thank the rest of the *Aprendo* team: Catalina Méndez Vallejo, Adriana Merino, and Sylvia Zetterstrand.

Finally, I'd like to extend a heartfelt thank you to the people I love most – my family. They have all rallied around me during difficult times, and for that I am incredibly grateful. Thanks especially to my mother and late father for their patience with me and unwavering support. My brother and I hit the lottery as far as parents go, and we know it. Thanks also to Kaitlin Mignella for her longstanding, tried and true friendship. She's not technically family, but she might as well be.

Finally, I would like to thank my top-notch husband, Jorge Méndez Seijas. Jorge has worked diligently with me over the years to make sure that I remain optimistic, independent, and well-traveled. I am very thankful for his role in shaping me into the person I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION AND BACKGROUND

The Nature of L2 Competence

A perennial question in second language acquisition (SLA) is whether non-native speakers of a language can acquire native-like competence in a second language (L2). The crux of the debate centers around whether Universal Grammar (UG), which is assumed to guide first language (L1) acquisition, is still accessible for late L2 learners. While all normally developing children do ultimately converge on a target-like grammar for their L1, adults learning a second language often create a steady state grammar that diverges in certain areas from that of adult L1 speakers of the target language.

Representational vs. Non-Representational Deficit Accounts

To account for these differing outcomes, some researchers have argued that L2 learners' competence is characterized by a representational deficit. Over the years, there have been multiple proposals in the literature that argue for a representational deficit account, such as the No Parameter Setting Hypothesis (Clahsen & Muysken, 1989), the Failed Functional Features Hypothesis (Hawkins & Chan, 1997), and the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou, 2007). Although the details of these proposals differ, they all claim that L2 learners cannot acquire parameterized functional features that are not instantiated in their L1. For this reason, the L2 learners' competence is deficient, and is therefore fundamentally different from that of native speakers.

This representational deficit view contrasts with other accounts that claim L2 learners still have access to UG, and can therefore acquire functional features not instantiated in their L1. This view includes hypotheses such as the Full Transfer/Full Access Hypothesis (Schwartz & Sprouse, 1996) and the Missing Surface Inflection Hypothesis (MSIH) (Prévost & White, 2000).

While the details of these hypotheses also differ, the underlying argument is the same: L2 learners' underlying competence is not impaired. That is, because L2 learners have access to UG, they are able to reset parameters, and ultimately acquire functional features that are not instantiated in their L1. To the extent that non-targetlike performance occurs, researchers ascribe it to different task demands such as communication pressure (Prévost & White, 2000, p. 129) or processing constraints (Hopp, 2010; Keating, 2010).

Grammatical Gender: A Test Case

Research on grammatical gender, a parameterized functional feature, has especially informed questions examining L2 learners' competence. At the heart of the problem is whether learners whose L1 does not have grammatical gender agreement (e.g., English[1]) can develop a native-like gender agreement system in an L2 that does have grammatical gender (e.g., Spanish, French, Dutch, German). While some researchers studying the acquisition of grammatical gender agreement have found that these learners' underlying linguistic systems are not native-like (e.g., Franceschina, 2001; Franceschina, 2005; Sabourin, Stowe, & de Haan, 2006), others have found that they are (e.g., Hopp, 2012; Keating, 2009; White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004).

Gender agreement occurs between a 'trigger' (generally a noun) and multiple targets (e.g., articles, adjectives, pronouns, demonstratives). It may well be the case that not all these types of agreement are equally difficult for the L2 language learner. That is, learners may demonstrate differential levels of accuracy in comprehending and producing[2] agreement between

---

[1] While English does mark natural gender on pronouns, it has no agreement (Corbett, 1991).
[2] While the goal of these studies is to investigate the status of L2 learners' underlying linguistic system, researchers resort to production measures and/or online or offline comprehension measures to indirectly access the linguistic system. Therefore, even though these researchers are

the trigger and different categories of targets (e.g., articles vs. adjectives). If this is the case, then conclusions drawn by researchers supporting the representational deficit account may be due to at least two possible explanations, either: (a) L2 learners have a representational deficit for grammatical gender agreement or, (b) L2 learners do not have a representational deficit for grammatical gender agreement, but perform poorly because of the type of agreement under investigation.

The Present Dissertation

In this dissertation, I will examine L1 English learners' acquisition of grammatical gender agreement in L2 Spanish. I will compare three different types of grammatical gender agreement: Determiner-Noun (DET-N), Noun-Adjective (N-ADJ) and null nominal agreement (N-DROP). This study will use an online sentence comprehension task to measure participants' sensitivity to violations of grammatical gender agreement, which will be measured by recording participants' eye-movements with an eye tracker. If sensitivity to the violations of grammatical gender agreement is found, it will be taken as evidence that learners do not have a representational deficit for morphology. If no such sensitivity is found, it will be taken as evidence that either (a) the L2 learners have a representational deficit, or (b) the L2 learners do have target-like representations, but they could not be detected[3]. Of particular interest in the present dissertation is whether the three types of agreement previously stated are equally difficult

---

ultimately concerned with linguistic competence, they tend to examine linguistic competence via performance measures.

[3] Unfortunately, it is very difficult to adjudicate between these possibilities. While many researchers conducting these types of studies do take lack of sensitivity to violations of grammatical gender agreement as an indication of a representational deficit, there always exists the possibility that the participants do have target-like representations, but the data collection tool and experimental design employed were not sensitive enough to detect them. Another possibility is that learners did not perform like native speakers because of communication or processing pressures (White et al., 2004).

for the L2 learners, or whether certain types of agreement are less difficult than others. To further understand how L2 learners process sentences, I will also explore their sensitivity to violations of grammatical gender agreement in a novel way: through a reading questionnaire administered after the eye-tracking experiment. Of particular interest is whether the L2 learners report seeing agreement violations in the experimental stimuli, and if so, whether their self-reports mirror any evidenced sensitivity in reading times.

Overview of the Dissertation

The organization of this dissertation is as follows: in Chapter 2, I will provide an overview of the literature and will explain the motivation for this dissertation. I will explain the nature of gender assignment and agreement, how different methodologies can be used to investigate grammatical gender agreement, and why I have chosen to investigate how learners compute agreement in three different syntactic contexts with the eye-tracking methodology. In Chapter 3 I will describe the method of the study, including the results of a pilot study that was critical in constructing the stimuli for the primary data collection. In Chapter 4, I will present the results of the eye-tracking study in two ways: both by statistical tests performed on both participant groups and individually on a participant-by-participant basis. I will also triangulate the L2 learners' reading times with their responses on a reading questionnaire, designed to measure if they were aware of the agreement violations in the experimental stimuli. Finally, in Chapter 5, I will discuss the results and the limitations of the current study, as well as propose future lines of research.

Definition of Terms

Universal Grammar (UG): A construct within generative theory that argues language is innate, uniquely human, and different from all other types of cognition. UG consists of principles and parameters that constrain language acquisition.

Linguistic competence: The mental representation of language that is abstract, implicit and normally not describable in lay terms.

Gender assignment: An inherent lexical feature that, in Spanish, ascribes nouns to one of two classes: masculine or feminine.

Grammatical gender agreement: A syntactic process by which controller nouns (called "triggers") enter an agreement relationship with targets (e.g., articles, adjectives, etc.).

Functional features: Subcategories of functional categories that express information (e.g., gender, tense, finiteness).

Input processing: Term that describes how syntactic and grammatical computations are made during sentence comprehension.

CHAPTER 2: MOTIVATION OF THE CURRENT STUDY

In this chapter I will describe the motivation of the current study. I will begin by examining the nature of grammatical gender assignment and agreement and will explore the syntax of gender agreement. Then, I will explore common methodological tools that researchers have employed to assess language learners' ability to compute grammatical gender agreement and will weigh the relative benefits and drawbacks of each. Next, I will make the case that the syntactic context in which agreement takes place affects the order in which it is acquired both for L1 and L2 acquisition. Finally, I will explain the rationale for the current study and present the research questions.

Grammatical Gender Assignment and Agreement

Gender assignment is the term used to refer to an inherent lexical feature on nouns (Carroll, 1989). Spanish nouns are divided into two classes: masculine or feminine. By and large, the endings on nouns are transparent: those that end in /–o/ are masculine while those that end in /–a/ are feminine. However, gender assignment is not always so straightforward: there are counterexamples to transparent nouns (e.g., *problema*, 'problem,' m., *mano*, 'hand,' f.) and some nouns are opaque, meaning that the grammatical gender cannot be deduced from the nominal ending (e.g., *cristal*, 'cristal,' m., *nariz*, 'nose,' f.). It is generally agreed that gender is an interpretable φ-feature that is lexically determined on the noun (Chomsky, 1995). Because masculine is considered default and unmarked, Spanish nouns are marked with the gender feature [±feminine] (Carroll, 1989; Carstens, 2000).

Gender agreement, on the other hand, is manifested on lexical items in a different syntactic category—such as determiners and adjectives—that bear some relationship to the noun.

Unlike nouns, these lexical items do not carry intrinsic gender; instead, they are marked for gender based on an agreement relationship with the noun they modify (e.g., *el refresco frío,* 'the$_{MASC}$ drink$_{MASC}$ cold$_{MASC}$'; *la guitarra negra*, 'the$_{FEM}$ guitar$_{FEM}$ black$_{FEM,}$'). While gender assignment is determined lexically, gender agreement is the result of a syntactic operation whereby the interpretable features on the triggers and the uninterpretable features on the targets are matched and checked off in the course of derivation. Table 2.1 summarizes the most salient differences between grammatical gender assignment and agreement.

Table 2.1
Comparison of Gender Assignment and Agreement in L2 Spanish

Gender Assignment	Gender Agreement
- Interpretable feature on triggers (i.e., nouns) - Lexically determined - Masculine is the default and is unmarked (feminine is marked)	- Uninterpretable feature on targets (e.g., determiners, adjectives) - Result of syntactic feature-checking operation

In example (1) below, I have depicted a syntactic tree for the DP *la guitarra negra*. As previously stated, gender is an intrinsic, interpretable feature of Spanish nouns. In example (1), the determiner (*la*) and the adjective (*negra*) both have uninterpretable gender features. Number is considered a strong feature in Spanish, which triggers the movement of the noun shown below[4]. Gender agreement transpires due to feature-checking and gender is checked as a 'free-rider' to number agreement in Spanish, because it does not prompt the movement (Carstens, 2000). A more detailed description of this syntactic structure can be found in Franceschina (2001, 2005) and Carstens (2000)[5].

---

[4] In English, number is a weak feature, accounting for the difference in English adjective placement (i.e., 'The black guitar').
[5] In the syntactic structure depicted, gender does not project its own phrase; however, that is not the stance taken by all researchers. For example, Picallo (1991) proposes a GenP and Bernstein (1993) proposes a Word Marker Phrase (WMP).

(1)
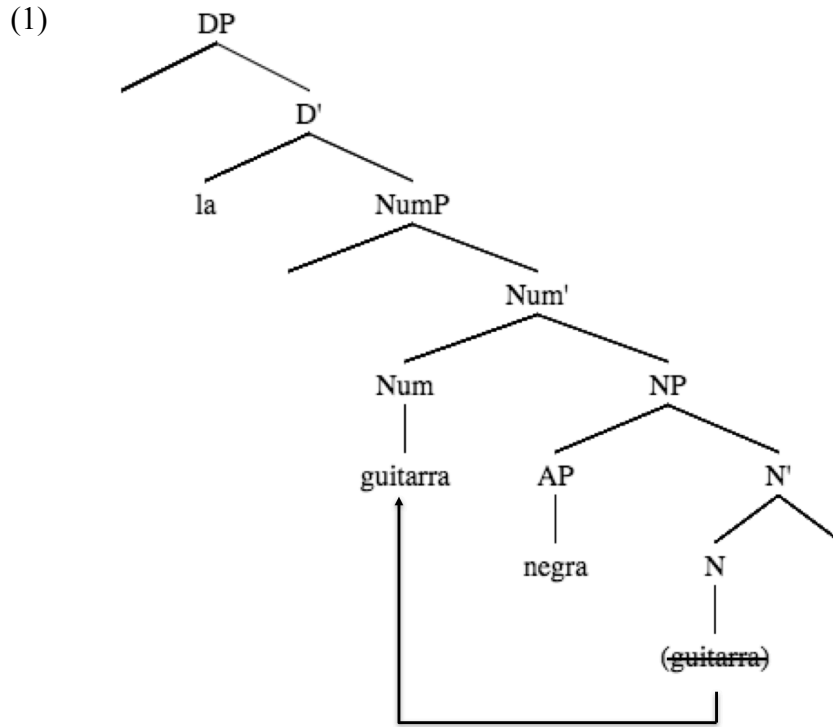
In research on L2 Spanish, it is often found that L2 learners exhibit a masculine default, meaning that masculine inflection occurs in feminine contexts (2a) more often than the reverse (2b).

(2) a. *la   manzana   delicioso

       the FEM   apple FEM   delicious MASC

       'the delicious apple'

   b. *el   cuarto   rosada

       the MASC   room MASC   pink FEM

       'the pink room'

Because masculine is considered the unmarked or underspecified gender in Spanish, (2a) is also referred to as an underspecification error. Gender agreement asymmetry has been widely documented in Spanish (e.g., Franceschina, 2001; McCarthy, 2007; Montrul, Foote, & Perpiñán,

2008; White et al., 2004), as well as other languages (e.g., a feminine default in German Spinner & Juffs, 2008).

Representational deficit accounts hypothesize that L2 learners cannot acquire uninterpretable gender features that are not instantiated in their L1. This means that L2 Spanish learners (L1 = English) are predicted to be able to assign the correct gender to nouns, but not acquire the uninterpretable features on determiners and adjectives (Franceschina, 2001, 2005). For that reason, gender agreement and not gender assignment is of primary interest to researchers investigating representational deficits.

That said, researchers cannot leave the question of gender assignment completely aside. To determine if L2 learners can successfully compute gender agreement, researchers must determine whether the language learner has assigned the correct gender to the noun in the first place. This represents a challenge for L2 researchers, as the interpretable gender features under investigation are not directly measurable. Researchers have used the agreement relationship between a determiner and noun as a reflection of gender assignment in both offline posttests (e.g., White et al., 2004), and online production measures (e.g., Hopp, 2012), on the assumption that determiners may be the most immediate reflection of a given lexical item's gender (Carroll, 1989); however, conflating agreement with assignment can make results difficult to interpret. What is more, researchers may not be consistent with their terminology within a single study. For example, Grüter, Lew-Williams and Fernald (2012) took participants' ability to use the determiner's gender as a predictive cue in a looking-while-listening task as their ability to compute agreement in real time; but in their production task, the researchers used DET-N agreement as evidence that "the speaker correctly classified the noun with regard to its gender class" (p. 201). Finally, researchers may also not be consistent in maintaining terminology

employed by the authors they cite. For example, Bruhn de Garavito and White (2002) investigated whether learners exhibited correct gender *agreement* in DPs involving a determiner and noun, yet Hopp (2012) interpreted their results as gender *assignment* (p. 35). In sum, Montrul et al. (2008) noted that whether a determiner's gender is related to gender agreement or assignment is "very hard to tease apart" (p. 510).

In the present dissertation, the relationship between determiners and nouns will be considered one of gender agreement and not assignment, as that is how the relationship is generally conceived of in the literature (Carroll, 1989; Carstens, 2000). Assignment will be measured in a posttest where learners must assign the correct gender (masculine or feminine) to the nouns under investigation.

Methodological Tools to Assess Gender Agreement

Because the linguistic system cannot be accessed directly, researchers must rely on participants' performance on different tasks to examine whether L2 learners have a representation for features that are not instantiated in their L1. Researchers can select from production or comprehension tasks, and the comprehension tasks may be either offline or online[6,7]. Minimizing task effects is critical in these types of studies, because they can potentially obscure the researchers' measurement of the linguistic system. I will review the potential task effects of production tasks, offline comprehension tasks and online comprehension tasks in turn.

---

[6] Online tasks are moment-by-moment measures of what participants are doing (e.g., eye tracking, SPR), while offline tasks are often paper-and-pencil tasks that do not collect moment-by-moment data (e.g., GJTs, cloze tests) (VanPatten & Benati, 2010).

[7] Production tasks are generally online by nature. It is difficult to imagine what an offline production task would look like, and, as Grüter et al. (2012) note, it would likely tap metalinguistic knowledge, thus making it an inappropriate method for examining the linguistic system.

When researchers administer production tasks (e.g., Bruhn de Garavito & White, 2002; McCarthy, 2008; Montrul et al., 2008; Prévost & White, 2000; White et al., 2004), they assume that participants' percent accuracy during performance is an indirect, yet relatively accurate, reflection of the state of their linguistic system. If participants' performance is highly accurate, then they are assumed not to have a representational deficit. The problem, however, is that if non-native speakers do not exhibit 100% accuracy on these tasks, the cause of the optionality is then up to the researcher's interpretation. Supporters of a representational account attribute any optionality to the inability to acquire grammatical gender agreement, however non-deficit accounts, such as the MSIH, interpret the optionality as being a by-product of communication pressure (Prévost & White, 2000, p. 129). Put differently, MSIH supporters would argue that even though learners still exhibit optionality, their underlying representations are target-like, and the optionality stems from task demands. The problem for researchers, then, is deciding exactly how much optionality can be explained away by communication pressure, and how much is indicative of a true representational problem. Since no such consensus exists, researchers could feasibly arrive at opposing conclusions when examining the same set of production data.

To obviate the communication pressure issue inherent in online production tasks, researchers have coupled these tasks with offline comprehension tasks such as multiple-choice tasks (e.g., Bruhn de Garavito, 2003), gender recognition tasks where learners must circle the agreeing determiner or adjective (Montrul et al., 2008), or a picture identification task (e.g., Grüter et al., 2012; McCarthy, 2008; Montrul et al., 2008; White et al., 2004). These offline tasks sidestep the communication pressure issue because participants are not under time pressure to respond. Consequently, performance accuracy on these offline measures is arguably more reflective of the linguistic system. The potential problem with these types of tasks, though, is that

their offline nature may invite explicit reasoning and reliance on metalinguistic knowledge. Competence, by definition, is implicit in nature, so any task that invites metalinguistic reasoning introduces an intervening variable.

For this reason, some researchers have turned to online comprehension tasks borrowed from processing and parsing research to tap the linguistic system. The benefit of these online comprehension tasks is that moment-by-moment data can be collected on how participants are processing language, which can then be used as an indirect measure of the linguistic system. There is a large range of methodologies to select from, including eye-tracking (e.g., Keating, 2009; Lim & Christianson, 2014; Sagarra & Ellis, 2013), self-paced reading (e.g., Jiang, 2004, 2007; Jiang, Novokshanova, Masuda, & Wang, 2011; VanPatten, Keating, & Leeser, 2012), self-paced listening (e.g., Marinis, 2007a, 2007b), ERPs (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2014; Barber & Carreiras, 2005; Dowens, Vergara, Barber, & Carreiras, 2010; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010) and neuroimaging (e.g., Christensen, Kizach, & Nyvad, 2013).

In many online comprehension studies, researchers attempt to tap participants' underlying competence for different linguistic phenomena by testing their sensitivity to grammatical violations (e.g., Kreiner, Garrod, & Sturt, 2013; Lim & Christianson, 2014; Sagarra & Herschensohn, 2010a). In these studies, researchers present participants with sentences embedded with grammatical violations of the linguistic phenomena under investigation as well as matched sentences that are properly formed. These target sentences are then distributed among distractors and/or fillers that may or may not also contain grammatical violations. After reading an experimental sentence, participants complete a secondary task, such as responding to a question about what they read. The researcher then compares the participants' reading times on

the grammatical and ungrammatical regions of the target sentences, and a relatively longer reading time on the ungrammatical region is taken as indication that the learner was indeed sensitive to the violation.

Jiang (2007) notes that one of the online methodologies discussed above, SPR, is particularly useful for testing competence because during production, L2 learners are more likely to monitor themselves and therefore use explicit knowledge. Because production is not necessary during SPR experiments, he notes that SPR "thus eliminates the motivation for applying explicit knowledge" (p. 12). Jiang does not conclude that SPR is a completely implicit measure; however, he notes that it elicits "little involvement of explicit knowledge" (p. 12). The same could be true for other online comprehension tasks (e.g., eye-tracking, SPL)[8]. Jiang is not alone in positing that these comprehension measures tap primarily implicit knowledge. In fact, researchers often pair a comprehension task with a separate grammaticality judgement task (GJT), on the premise that the former measures more implicit knowledge and the latter more explicit knowledge (e.g., Coughlin & Tremblay, 2013; Roberts & Liszka, 2013; Sagarra & Herschensohn, 2010b).

Yet the extent to which researchers avoid tapping explicit knowledge during these sentence comprehension experiments is still open to debate, and is likely dependent on a host of methodological decisions. One such decision is the type of secondary task participants engage in after reading each experimental sentence. Some options for secondary tasks include comprehension questions (e.g., Keating, 2009; VanPatten et al., 2012), plausibility judgements (e.g., Wen, Miyao, Takeda, Chu, & Schwartz, 2010), grammaticality judgements (Godfroid et

---

[8] It should be noted, though, that Jiang (2007) ultimately favors SPR to eye tracking because of the transient nature of the stimuli presentation in the former.

al., 2015), producing a translation of the experimental sentence (Lim & Christianson, 2014), and assessing the accuracy of a translation of the experimental sentence (Keating, 2009). It seems, though, that the type of task may affect how learners process the sentences and what type of knowledge they employ. For example, Leeser, Brandl and Weissglass (2011) tested L2 Spanish learners to violations of noun-adjective agreement and subject-verb inversion in wh-questions. Participants completed two SPR experiments: in one, they answered a yes/no comprehension question after each experimental sentence, and in the other they made grammaticality judgements. Leeser et al. (2011) found that task type affected participants' sensitivity to noun-adjective agreement violations: they were sensitive when asked to perform grammaticality judgements, but not when answering comprehension questions. The participants were not sensitive to violations of subject-verb inversion in either condition. The grammaticality judgement Leeser et al. (2011) employed was an untimed judgement, but another potential mediating factor is whether the GJT is timed or untimed, as the former is often posited to measure implicit knowledge and the latter explicit knowledge (Ellis, 2005; Godfroid et al., 2015).

In addition to the type of secondary task employed, another methodological decision that may affect the extent to which participants tap explicit knowledge is percentage of grammatical violations that appear in the experimental sentences, both in the targets and distractor sentences, with more grammatical violations likely increasing the likelihood that participants will tap explicit knowledge. Researchers vary greatly in the percentages of grammatical violations present in the sentences. For example, Coughlin and Tremblay (2013) report that half of their distractor sentences contained violations, yet Sagarra and Herschensohn (2010b) report that only about 13% of their sentences contain violations.

While it is very difficult to gauge whether participants are aware that the experimental sentences contain violations and are actively looking for them during these online comprehension tasks, posttest questionnaires could be used to determine how aware participants were of the violations during the experiment. In this dissertation, I will report findings from a reading questionnaire that aims to examine precisely this question.

For an online comprehension measure, I selected eye-tracking methodology to examine participants' sensitivity to violations of grammatical gender agreement (for reviews of this methodology see Dussias, 2010; Holmqvist et al., 2011; Keating, 2014; Roberts, 2012). One of the benefits of eye tracking is that the stimuli can be presented in their entirety rather than on a word-by-word or segment-by-segment basis. This allows participants to read more naturally in eye-tracking experiments than other online comprehension measures, such as self-paced reading (Dussias, 2010; Roberts, 2012; Witzel, Witzel, & Forster, 2012). Another benefit of the eye-tracking methodology is that participants do not need to perform a secondary task while reading[9] (e.g., pressing a button) which could potentially affect comprehension (Dussias, 2010). Eye tracking also provides a plethora of fine-grained data that can reflect early and late cognitive processing, whereas other online processing measures, such as self-paced reading, provide reaction time data, which are more unidimensional.

Gender Agreement in Different Syntactic Contexts

In Spanish, grammatical gender marking is very prevalent, and can occur between a trigger (generally a noun) and multiple different targets (e.g., articles, adjectives, pronouns, and demonstratives). This preponderance of agreement marking provides researchers with many

---

[9] A secondary task is usually performed *after* reading each sentence, though (e.g., answering a comprehension question).

types of agreement that could potentially be investigated (e.g., DET-N, N-ADJ) either during

oral and written production, or during online and offline receptive tasks (oral or written). While

many researchers do provide a theory-based justification for the types of tasks they employ (e.g.,

Grüter et al., 2012; McCarthy, 2008), a rationale for the type of agreement under investigation is

generally lacking.

For example, McCarthy (2008) investigated intermediate and advanced L2 Spanish

learners' (L1 = English) ability to correctly produce adjective and direct object clitic agreement

in a production task and comprehend direct object clitic agreement on an interpretation task.

McCarthy found that the learners' performance was variable on both tasks. Because the

optionality extended into comprehension, McCarthy interprets these results as evidence that the

L2 learners have a representational deficit for gender agreement. However, McCarthy also

noticed that many of the advanced learners' errors were underspecified, thus showing evidence

of a masculine default[10]. In her discussion, McCarthy cites White et al. (2004), who found that

L2 learners exhibited optionality for gender agreement, but also perfect accuracy on noun-

adjective word order (thus showing an unimpaired syntax). McCarthy therefore argues that the

deficit for her participants occurs in the morphology, not in the syntax.

It is unclear why McCarthy would decide to examine agreement on direct object clitics

during comprehension, because they are difficult to begin with for Spanish language learners.

Therefore, performance problems could easily be ascribed to the difficulty of the structure, rather

than agreement problems (cf. VanPatten et al., 2012, p. 112). An intriguing question is whether

---

[10] Supporters of the MSIH interpret these types of errors as evidence of the MSIH by arguing
that the pattern of the errors is evidence of a functioning system. The errors themselves are
ascribed to communication or processing pressures (Prévost & White, 2000; White et al., 2004).

McCarthy would have obtained the same results, had she tested gender agreement in a less difficult syntactic context.

To be sure, in studies that examine whether L2 learners can acquire uninterpretable gender features, the syntactic context question is tangential, since participants only need to show that they have acquired the features when tested on a single type of agreement. It may well be the case, though, that there is an interaction between the syntactic context where the agreement transpires and whether participants can compute the agreement. If this were the case, then studies that examine more difficult types of agreement may conclude that learners have not acquired the features, when in reality they have, just in another syntactic context. For this reason, it is essential to determine in which syntactic contexts agreement is the most and least difficult to compute. I will first explore the literature examining the acquisition of grammatical gender agreement in Spanish L1 acquisition in different syntactic contexts, and will then turn to L2 studies.

Acquisition of Agreement in L1 Spanish

L1 Spanish speaking children do exhibit optionality in grammatical gender agreement, although such optionality is relatively infrequent compared to other morphological errors, such as verbal morphology (Mariscal, 2009).

Agreement on the definite article seems to be acquired before agreement on the indefinite article, perhaps owing to the former's higher frequency in the input (Mariscal, 2009). Also DET-N agreement is mastered before N-ADJ agreement (Hernández Pina, 1984). López Ornat (1988) noted that L1 Spanish-speaking children correctly produce DET-N agreement from about 18 to 24 months. Then, at about 24 months, children exhibit variability in DET-N agreement, leading López Ornat to attribute the previous stage to the production of unanalyzed chunks. Around 30

months, learners begin to exhibit optionality in the marking of (DET-)N-ADJ agreement, and these errors continue until about roughly 36 months of age.

L1 Spanish speakers also acquire agreement over longer linear distances. In Spanish, nouns do not always have to be overtly realized. Null nominals (also known as N-drop or nominal ellipsis) are DPs in which the noun is dropped or omitted (Bernstein, 1993; Snyder, 1995). This occurs when the referent can be recovered through the gender or number on the determiner of the null nominal. This is shown in example (3):

(3)   *Gladys quiere la pelota blanca pero Maruja quiere la **e** roja.*

        Gladys wants the white ball but Maruja wants the red (one).

In example (3), the noun is overt in the first DP (*la pelota blanca*), however, in the second DP it has been omitted (*la **e** roja*). The gender and number features on the determiner allow the noun to be recovered from the context. In English, a similar interpretation is achieved by using the pronoun 'one.'

Snyder, Senghas and Inman (2001) investigated the acquisition of N-DROP by two Spanish-speaking children: María and Koki. María mastered agreement on determiners and adjectives at around 25 months, and began producing N-DROP around the same time as attributive adjectives. Koki, on the other hand, mastered agreement on determiners and adjectives long before producing N-DROP. Koki began producing determiners at 19 months, but she did not produce N-DROP utterances until 30 months. Liceras, Díaz and Mongeon (2000) also reported comparable findings.

Acquisition of Agreement in L2 Spanish

In L2 research, syntactic context may also play a role in acquisition. For example, in an elicited production task, Bruhn de Garavito and White (2002) found that L2 Spanish speakers

(L1 = French) were more accurate marking grammatical gender agreement on definite rather than indefinite DPs, and that agreement errors were more frequent on feminine than masculine nouns, thus showing evidence of a masculine default. Participants also exhibited more optionality on N-ADJ agreement (roughly 29-31% total errors, depending on proficiency group) than DET-N agreement (only 11-18.5% errors) (for similar results in L2 French see Dewaele & Véronique, 2001), and that optionality on N-ADJ agreement was more frequent with feminine than masculine nouns.

In a Spanish production task, White et al. (2004) also found that their participants fared slightly worse on (DET-)N-ADJ agreement than DET-N agreement, especially at lower levels of proficiency. Although White et al. examined DET-N and (DET-)N-ADJ agreement during production, for their receptive task, they elected to examine N-DROP. During this task, participants read a short dialogue that ended with a null nominal, that is, a DP with only a determiner and an adjective (no noun). They were then presented with images of three nouns that differed in gender and/or number and had to indicate to what picture the null nominal referred. If they had acquired gender and number, then it was hypothesized that they would select the correct picture. If they had not, then they should select pictures randomly. Participants with an advanced level of Spanish performed quite well on this measure in terms of gender agreement (over 90% accuracy), but participants at intermediate and beginner proficiency levels performed worse (about 80-85% accuracy for intermediate learners and 55-65% for beginners[11]). It is unclear whether the intermediate and beginner learners performed worse because they did not understand the syntactic structure of N-DROP, or because they could not compute the syntactic agreement. Also, since the researchers used agreement in different syntactic contexts to assess L2 learners'

---

[11] Chance level was 33% accuracy.

19

ability to compute grammatical gender agreement, it is impossible to disentangle the difficultly of the grammatical structure from participants' ability to compute the agreement.

Grüter et al. (2012) also investigated advanced L2 Spanish (L1 = English) learners' acquisition of grammatical gender agreement by comparing the learners' performance on an offline measure of comprehension, an online measure of comprehension and a production task. For the offline measure of comprehension, the researchers employed the picture-identification task created by White et al. (2004). Both native Spanish speakers and Spanish learners performed at ceiling on this task. To measure online comprehension, the authors employed the looking-while listening procedure (Fernald, Zangl, Portillo, & Marchman, 2008). Two pictures were presented to the participants on a screen and they listened to a sentence naming one of the images. The pictures showed two nouns that were either of the same or different genders, and the researchers investigated whether participants could use gender as a cue to select the correct picture. In their oral production task, Grüter et al. elicited DET-N-ADJ and null nominal structures (DET-ADJ), but collapsed the two for analysis, making it impossible to determine if participants were equally accurate on the two types of agreement. Grüter et al. found that in the production task, native and near-native speakers did not differ in gender agreement, but near-native speakers were statistically less accurate in gender assignment. This is likely because many of the targeted nouns were not transparent (for a description of the same study materials see Montrul et al., 2008). Importantly, assignment for the production task was operationalized as DET-N agreement, "on the assumption that determiner choice is the most immediate reflection of a noun's lexical gender" (p. 201, cf. Carroll, 1989). It should be noted though, that for the looking-while listening task, determiners were used as a predictive cue indicating ability to compute gender agreement, not assignment.

Unlike the previous studies, Franceschina (2001) investigated gender agreement in multiple syntactic contexts in a single task. In this study, oral production data was collected from Martin, a L2 Spanish learner (L1 = English) who lived in a Spanish immersion context for 24 years. Martin's accuracy for gender agreement depended heavily on syntactic context: demonstrative (100%), pronoun (98%), article (94%), adjective (77%). He also showed evidence of masculine defaults in his production. This non-native-like performance led Franceschina to conclude that Martin evidenced a syntactic deficit.

In a more complex and comprehensive study, Franceschina (2005) investigated the performance of near-native L2 Spanish learners from L1s with and without grammatical gender (henceforth +gen and –gen, respectively). Franceschina collected naturalistic production data as well as results from six experimental tasks: five that tested gender agreement and one that tested assignment. In the naturalistic production data, Franceschina found that the native speakers and L1 +gen participants performed at ceiling for gender agreement on determiners, adjectives and pronouns, but the L1 –gen participants fared worse, with 93% accuracy on determiners, 90% accuracy on adjectives and 87% percent accuracy on pronouns. The five experimental tasks that tested agreement were: (1) a guessing game, (2) a missing word task, (3) a cloze/multiple choice task, (4) a GJT, and (5) a novel word task. Test 1 investigated nouns, adjectives and pronouns, test 2 investigated clitic pronouns, test 3 investigated nouns and adjectives, test 4 investigated all categories marked for gender and test 5 investigated agreement on adjectives and pronouns. In the battery of tests, the –gen participants consistently performed worse than the +gen participants, leading Franceschina to conclude that they had a representational deficit for gender agreement. Tests 4 and 5 were production tasks, tests 1 and 3 were comprehension tests, and test 2 required both production and comprehension; however, it is worth noting that all the

comprehension tests were offline tests. Even though the L2 learners had naturalistic exposure to the L2, they still may have used their knowledge of grammatical rules to complete the offline comprehension tasks.

Summary of L1 and L2 Research

In sum, it seems that there are some parallels between the acquisition of grammatical gender agreement in L1 and L2 Spanish. First, definite articles seem to exhibit less optionality than indefinite articles for both groups. Second, DET-N agreement seems to be easier to compute than N-ADJ agreement. It also seems that L1 Spanish learners acquire agreement in overt constructions before being able to produce N-DROP.

In many of the L2 studies mentioned, syntactic context is a confounding variable. For example, White et al. (2004) tested DET-N and N-ADJ agreement during production, but N-DROP during an offline comprehension task with a picture-identification task. While this task should be praised for its ingenuity, it is unclear why the researchers decided to examine different types of agreement in the productive and receptive mode[12]. Even though Franceschina did investigate agreement of varying syntactic contexts in a single task, these tasks were either production tasks (2001, 2005) or offline compression tasks (2005), leaving open the question of how L2 learners would perform on an online comprehension task.

<div align="center">

Motivation of the Present Dissertation

</div>

Researchers investigating L2 grammatical gender agreement oftentimes do not seem to have a principled reason for the types of grammatical gender agreement that they investigate, and

---

[12] Although N-DROP is like DET-N and N-ADJ agreement in the sense that a target adjective must agree with a trigger noun, they are different in that the noun is not overtly realized within the null nominal. Instead, the speaker/listener must rely on discourse factors to recover the gender and number features of the null noun.

many seem to assume that agreement in all syntactic contexts is the same in difficulty for L2 learners; however, given the results of studies that investigate L1 acquisition, and the work of Franceschina (2001, 2005), that is likely not the case. In many studies that investigate whether L2 learners have a representational deficit for grammatical gender agreement, it is impossible, then, to conclude whether learners truly do have a representational deficit, or whether they are performing poorly because of the agreement type under investigation. Therefore, to tease apart the effects of the syntactic context of agreement, various types of agreement must be tested in a single task.

To this end, I assess three types of agreement to determine if they are equally difficult for L2 Spanish learners. The three types of agreement I selected are DET-N, N-ADJ and N-DROP for comparability purposes, as they seem to be the three frequent types of agreement studied in this line of research (e.g., Grüter et al., 2012; White et al., 2004). Judging by the L1 and L2 research previously reviewed, it seems that in L1 Spanish learners acquire agreement in the following order:

1. DET-N agreement

2. N-ADJ agreement

3. N-DROP agreement

L2 learners seem to generally be more accurate on DET-N agreement than N-ADJ agreement (Bruhn de Garavito & White, 2002; Franceschina, 2001). Although task type or the coding of data area often confounding variables in L2 studies, if L2 learners acquire agreement as L1 learners do (Snyder et al., 2001), then they will likely be more accurate on N-ADJ agreement than N-DROP.

In this dissertation, participants' sensitivity to violations of grammatical gender agreement in each of the three conditions was tested in a reading comprehension task. As they read, participants' eye movements were recorded with eye-tracking methodology. Longer reading times on violations of gender agreement relative to matched grammatical regions was assumed to reflect a processing cost, which was taken as indirect evidence that the participants did not have a representational deficit for grammatical gender agreement. Participants were also given a reading questionnaire after finishing the eye-tracking experiment, to determine if they were aware of the gender agreement violations. This task was employed to determine what type of strategies L2 learners are using when processing sentences that contain violations.

Research Questions and Predictions

The following research questions will guide this dissertation:

1. Are native Spanish speakers and L2 learners sensitive to violations of DET-N, N-ADJ and N-DROP gender agreement, as evidenced by their reading times during an online processing task?

2. Is sensitivity to the grammatical violations contingent on the type of agreement under investigation?

3. Do native Spanish speakers and L2 learners report sensitivity to the violations in the experimental stimuli?

I predict that native Spanish speakers will be sensitive to violations of grammatical gender agreement in all three conditions. I also predict that the L2 learners will be sensitive to violations of grammatical gender agreement, but not necessarily in all three conditions. This could mean that the L2 learners would not have a representational deficit for gender agreement (since they are indeed sensitive to violations in at least one condition), but that their sensitivity to

the violation is mediated by the agreement condition under consideration. Give the previous literature in L1 and L2 acquisition of grammatical gender agreement, I predict that L2 learners will show higher rates of sensitivity to violations of DET-N agreement, followed by N-ADJ agreement and then N-DROP agreement. In the analyses, I will investigate whether this is true both for the L2 learner group as a whole, and also for each individual participant. As for research question 3, I predict that native Spanish speakers will report sensitivity to the violations of grammatical gender agreement, but the L2 learners will likely only report such sensitivity if they also evidence sensitivity in the reading time data.

CHAPTER 3: METHOD

This chapter describes the method used to answer the research questions presented in Chapter 2. Below, I provide details on the research design, participants, materials, procedures, and analyses employed in the present dissertation.

Research Design

Eye-tracking methodology was used to determine if participants were sensitive to violations of grammatical gender agreement. Sensitivity to these violations was operationalized as a relatively longer reading time on ungrammatical regions involving grammatical gender agreement relative to matched grammatical regions. This increased reading time was assumed to reflect a processing cost, and was taken as indirect evidence that the participants did not have a representational deficit for grammatical gender agreement, provided that an increased reading time was also found in a control group of native Spanish speakers (cf. Sagarra & Herschensohn, 2010a; VanPatten et al., 2012).

In the present dissertation, there were three independent variables and four dependent variables. The independent variables were group (native Spanish speaker, L2 Spanish learner), condition (DET-N, N-ADJ, N-DROP), and grammaticality (grammatical, ungrammatical). The dependent variables were first fixation duration, first-pass time, go past-time, and total time. A description of the dependent variables is located under **analysis**.

Participants

Sixty people participated in the current study. Participants were either L2 Spanish learners or native Spanish speakers. All participants were recruited from the Michigan State University (MSU) community.

L2 Spanish Learners

The L2 Spanish learners ($n = 29$) were notified of the opportunity to participate in the study via three possible routes: (1) an email sent by their undergraduate advisor, (2) an announcement made in one of their classes, or (3) word of mouth. Only Spanish majors that were close to completing their degree at MSU were invited to participate in the study[13].

In order to be retained for analysis, the L2 Spanish learners had to: (a) be at least 18 years old, (b) be a native speaker of English, (c) be a Spanish major at MSU, (d) be close to completing their major (i.e., have no more than three classes left), (e) have studied no other language with grammatical gender agreement for more than two years, (f) demonstrate they understood the experimental sentences by scoring at least 80% on the comprehension questions[14], (g) have normal or corrected vision, and (h) consistently pass calibration during the eye-tracking experiment.

A total of four L2 learners were eliminated from the analyses for having a native language other than English ($n = 1$), for studying another language with grammatical gender (French) for more than two years ($n = 1$), and for not being close enough to completing their Spanish major ($n = 2$). This yielded a total of 25 L2 Spanish learners retained for analysis.

The L2 learners were on average 21.56 years old ($SD = 1.23$). Nineteen were female and six were male. Some participants were majoring only in Spanish ($n = 7$), while others were dual majors in both Spanish and another subject ($n = 18$). Some had studied other languages (e.g., French, Japanese, Hindi), but no participant had studied any language with grammatical gender

---

[13] Most L2 learners were seniors, but some were juniors that had taken as many Spanish classes as seniors.

[14] A description of these comprehension questions is provided in the **experimental sentences** section below.

agreement for more than two years. Spanish majors at MSU must take a minimum of 12 Spanish classes (36 credits) that count towards the major. The majors that participated in this study still had an average of .94 classes left to take ($SD = 1.24$, range: 0 - 3) before completing the major. On average, participants began learning Spanish at age 12.32 ($SD = 3.11$) and reported actively studying Spanish for an average of 8.84 years ($SD = 3.12$). On a scale of 0 (none) to 10 (very good), participants rated their level of proficiency speaking ($M = 6.44$, $SD = 1.78$), understanding spoken language ($M = 7.20$, $SD = 1.78$), and reading ($M = 7.12$, $SD = 1.48$). Participants estimated that in a given week they spent 3.92 hours ($SD = 5.48$) speaking Spanish, 5.36 hours ($SD = 5.81$) listening to Spanish, 3.12 hours ($SD = 4.81$) writing in Spanish and 5.08 hours ($SD = 6.21$) reading in Spanish.

Many L2 learners had also studied abroad in a Spanish-speaking country ($n = 21$). Spanish majors at MSU are highly encouraged to study abroad for at least eight weeks. Spanish majors who do not study abroad for eight weeks must either (a) complete an internship experience in a Spanish-speaking environment, (b) complete a service learning experience in a Spanish-speaking environment, or (c) enroll in an additional class. In the current study, 15 participants had studied abroad, 3 studied abroad and completed an internship, 3 studied abroad and completed a service learning experience, 1 completed only an internship, 1 completed only a service learning experience and 2 had not done either of these activities[15]. Those that studied abroad studied in Spain, Ecuador or Peru for an average of 13.06 weeks ($SD = 4.53$, range: 8.00-21.70).

---

[15] These two students presumably took an additional class.

Native Spanish Speakers

The native Spanish speakers ($n = 31$) were notified of the opportunity to participate in the study via three possible routes: (1) an email sent to native Spanish speakers on the MSU campus, (2) flyers posted around campus, or (3) word of mouth.

To be retained for analysis, the native Spanish speakers had to: (a) be at least 18 years old, (b) be a native speaker of Spanish, (c) be born in a Spanish-speaking country, (d) immigrate to the United States at or after age 16, (e) identify as Spanish-dominant, (f) demonstrate they understood the experimental sentences by scoring at least 80% on the comprehension questions, (g) have normal or corrected vision, and (h) consistently pass calibration during the eye-tracking experiment.

A total of four native Spanish speakers were eliminated from the analyses for not reporting Spanish as their dominant language ($n = 2$), not having normal or corrected vision ($n = 1$), and scoring below 80% on the comprehension questions during the reading portion of the experiment ($n = 1$). This yielded a total of 27 native Spanish speakers that were retained for analysis.

The native Spanish speakers were on average 25.59 years old ($SD = 7.74$). Fourteen were female and thirteen were male. They were born in various Spanish-speaking countries[16]: Argentina, Colombia, Costa Rica, the Dominican Republic, Guatemala, Mexico, Panama, Peru, Puerto Rico, Spain, and Venezuela. On average, they immigrated to the United States at age 22.41 ($SD = 6.51$) and had spent a total of 3.02 years in the United States ($SD = 3.05$). All but

---

[16] Unlike some aspects of Spanish (e.g., subjunctive vs. indicative distribution), grammatical gender agreement is not variable across dialects.

two had attended at least some college, and nine were pursuing advanced degrees (either master's or doctoral degrees).

<div align="center">**Materials**</div>

Participants completed five tasks: they first read sentences while their eye movements were recorded with an eye-tracker, and then also completed a reading questionnaire, vocabulary posttest, background questionnaire, and proficiency test. The creation of the materials for each of these tasks is described below.

Eye-tracking Materials

The eye-tracking materials consisted of 60 critical experimental sentences that were used to investigate the research questions stated in Chapter 2. In this section I describe how I selected the target nouns for inclusion in the experimental sentences, how the experimental sentences were constructed, and the apparatus used to record participants' eye movements.

Section of target nouns: A pilot vocabulary test. The target nouns used in the current study were selected based on the results of a pilot vocabulary test. The pilot vocabulary test was administered to Spanish 310 students at MSU through Survey Gizmo, an online survey tool. Spanish 310 is a grammar class, and it is the "gateway" course that all students must take before taking courses in the major or minor sequence at MSU. I decided to collect the pilot vocabulary data from these participants because they are generally of a lower proficiency than the Spanish majors who participated in the eye-tracking experiment. In this way, the Spanish 310 students' performance on the vocabulary test yielded a conservative estimate of the vocabulary knowledge of the participants in the primary data collection. Students were made aware of the survey in their SPA 310 class, and took this survey outside of class, during their free time.

For this pilot vocabulary test, participants were given a list of 32 nouns: 16 masculine and 16 feminine. I judged these words to be nouns that are commonly taught in basic Spanish language textbooks and/or are common in classroom discourse (e.g., *ensayo* 'essay').

For each of the 32 words, participants were first asked to translate the word into English, and then rate their knowledge of the word with the following scale:

- 4 − I know this word very well; I translated it correctly and rapidly.

- 3 − I know this word somewhat well; I translated it correctly after some thought.

- 2 − I'm unsure of this word; I'm unsure if my translation is correct

- 1 − I don't think I know this word; I don't think my translation is correct.

- 0 − I definitely don't know this word; My translation is definitely incorrect.

This scale was selected because it is familiar to these participants, as students at MSU are graded on a scale such as this one. After the participants translated the 32 words and rated their knowledge, they were asked to identify the gender of each word. The 32 words were always presented in isolation (i.e., without a sentential context), but participants were told that the words were nouns. The presentation of the words was randomized for each participant, and the words in the translation and rating activity were broken up into four blocks of eight words, to reduce participant fatigue. The pilot vocabulary test is presented in Appendix A.

The design of the pilot vocabulary test ensured that the participants in the primary data collection were likely to (a) know the gender of the target nouns and, (b) be familiar with their translations. The former is important because learners' sensitivity to the agreement violations hinges on their ability to assign the correct gender to the target noun. The latter is important because word familiarity affects processing (Clifton, Staub, & Rayner, 2007; Rayner, 1998), which could potentially introduce a confounding variable into the reading time measures.

A total of 42 Spanish learners completed the pilot vocabulary test. Four participants reported a native language other than English and were removed. This left a total of 38 participants retained for analyses. The participants had taken an average of 2.21 semesters of Spanish classes at MSU ($SD = 1.07$) and had been actively studying Spanish for an average of 6.96 years ($SD = 3.71$).

I first calculated the average knowledge rating and percentage of correct gender assignment across participants for each noun. I then coded the translations for each noun into English as correct or incorrect. Because this coding required a small degree of interpretation, a native Spanish speaker who is also highly proficient in English coded the translations as well. Inter-rater agreement was high − 99.18%. All disagreements between raters were discussed and resolved. The results of the pilot vocabulary test are presented in Table 3.1.

Participants generally performed well on the gender assignment task, assigning the correct gender on average 99.84% of the time for masculine nouns ($SD = 1.01$; range: 93.75-100.00) and 98.5% of the time for feminine nouns ($SD = 3.96$; range: 81.25-100.00). Participants were less accurate translating the words correctly into English. On average, they correctly translated masculine nouns 89.47% of the time ($SD = 10.59$; range: 62.50-100.00), and feminine nouns 86.68% of the time ($SD = 11.64$; range: 56.25-100.00).

Table 3.1
Pilot Vocabulary Test Results

		English Equivalent	Correct Translation (%)	Knowledge Rating		Correct Gender Assignment (%)
				Mean	SD	
Masculine Nouns	zapato	'shoe'	100.0	3.95	0.23	97.4
	sombrero	'hat'	100.0	3.92	0.40	100.0
	almuerzo	'lunch'	100.0	3.89	0.39	100.0
	refresco	'drink'	100.0	3.61	0.68	100.0
	trabajo	'work'	97.4	3.95	0.23	100.0
	mercado	'market'	94.7	3.71	0.84	100.0
	ensayo	'essay'	92.1	3.76	0.59	100.0
	dibujo	'drawing'	92.1	3.63	0.91	100.0
	archivo	'file'	92.1	2.37	1.26	100.0
	regalo	'gift'	89.5	3.79	0.66	100.0
	vestido	'dress'	89.5	3.76	0.59	100.0
	partido	'game'	86.8	3.68	0.81	100.0
	espejo	'mirror'	84.2	3.16	1.22	100.0
	cuchillo	'knife'	78.9	3.13	1.02	100.0
	abrigo	'coat'	68.4	2.58	1.62	100.0
	cuaderno	'notebook'	65.8	3.42	1.03	100.0
Feminine Nouns	comida	'food'	100.0	3.97	0.16	100.0
	escuela	'school'	100.0	3.97	0.16	100.0
	ventana	'window'	100.0	3.95	0.23	100.0
	bebida	'drink'	100.0	3.82	0.61	100.0
	iglesia	'church'	97.4	3.97	0.16	97.4
	pregunta	'question'	97.4	3.97	0.16	97.4
	camisa	'shirt'	97.4	3.79	0.53	100.0
	manzana	'apple'	92.1	3.97	0.16	100.0
	cocina	'kitchen'	92.1	3.95	0.23	100.0
	piscina	'pool'	92.1	3.87	0.41	100.0
	mochila	'backpack'	92.1	3.63	1.03	94.7
	pintura	'painting'	78.9	3.55	1.01	97.4
	cerveza	'beer'	76.3	3.63	0.91	100.0
	corbata	'tie'	63.2	2.45	1.67	92.1
	revista	'magazine'	60.5	3.03	1.17	100.0
	maleta	'suitcase'	47.4	2.45	1.30	97.4

Note. Results are presented in descending order in terms of correct translation percentage and then mean knowledge rating.

Because the gender assignment scores were all high, the target words were selected based on the translation accuracy as well as the knowledge rating. The percentage of correct translations was used as the primary determining factor as to whether a word should be included as a target word in the primary data collection. The knowledge rating was the secondary determining factor. The goal of the pilot vocabulary knowledge study was to select the 10 masculine and 10 feminine nouns that were best known to the participants; however, since one of the words in the top 10 (*archivo,* 'file') had a relatively high translation score (92.1%) but a low average knowledge score (2.37), only the first eight masculine and feminine words were selected for inclusion in the main study. The masculine words were *zapato* 'shoe', *sombrero* 'hat', *almuerzo* 'lunch', *refresco* 'drink', *trabajo* 'work', *mercado* 'market', *ensayo* 'essay', and *dibujo* 'drawing'. The feminine words were *comida* 'food', *escuela* 'school', *ventana* 'window', *bebida* 'drink', *iglesia* 'church', *pregunta* 'question', *camisa* 'shirt', and *manzana* 'apple'.

Cognates were selected for the remaining 4 words (2 masculine and 2 feminine) to ensure that participants would be familiar with them. The masculine cognates were *momento* 'moment' and *proyecto* 'project', and the feminine cognates were *guitarra* 'guitar' and *cámara* 'camera'. Because lexical access is affected by cognate status (e.g., Dufour & Kroll, 1995; van Hell & de Groot, 1998), these cognates were distributed carefully across different lists (see **experimental sentences** section below) to wash out the potential influence cognate status may have on reading times.

In sum, a total of 20 nouns (10 masculine, 10 feminine) were selected for the primary data collection. All the nouns were three syllable nouns (6-8 letters in length) and had transparent /-o/ (masculine) or /-a/ (feminine) endings.

I conducted an a posteriori investigation of the frequency of the target nouns using
Davies' frequency dictionary (2006). This frequency dictionary includes the 5,000 most common
words in the Spanish language, drawn from a 20,000,000-word corpus. Approximately two-
thirds of this corpus is based on written texts (literary and non-literary) and one-third is based on
spoken Spanish.

A vocabulary test was used to select the target words instead of frequency rank
calculations because the latter may not accurately reflect the majority of the input L2 learners
receive. As a case in point, according to Davies (2006) the verbs *bañar* 'to bathe' and *cenar* 'to
eat dinner' are roughly as frequent (ranked 3224 and 3261, respectively) as the verbs *rozar* 'to
touch lightly' and *alentar* 'to encourage' (ranked 3226 and 3189, respectively). An intermediate
L2 Spanish learner, however, would likely know only the first two. The frequency ranks of the
target nouns in the current study are presented in ascending order in Table 3.2 below.

Table 3.2
Frequency Ranks for Target Words

Masculine Nouns	Frequency Rank	Feminine Nouns	Frequency Rank
momento	108	pregunta	481
trabajo	145	escuela	532
proyecto	604	comida	873
mercado	609	iglesia	1111
dibujo	1692	cámara	1172
ensayo	1835	ventana	1265
zapato	1932	camisa	2443
sombrero	2899	bebida	2828
almuerzo	3104	manzana	2853
refresco	---	guitarra	2885

Most of the target nouns were ranked within the 3,000 most frequent Spanish words. One
word, *refresco* 'drink' was not listed in the frequency dictionary, as its frequency did not fall
within the 5,000 most common words in the Spanish language. This word was retained as a

target word, however, because (a) 100% of the students translated it correctly on the pilot

vocabulary test, and (b) research with similar populations has also found it to be a well-known

word (Keating, 2005).

The overall average frequency rank for the selected target nouns (not including *refresco*)

was 1545.82 (*SD* = 1034.96). The average rank for the masculine nouns was 1436.44 (*SD* =

1129.16) and the average for the feminine nouns was 1644.30 (*SD* = 993.09).

Experimental sentences. I created 60 NPs, 20 for each condition. For the DET-N

condition, the NP consisted of a definite article and the target noun. For the N-ADJ condition,

the NP consisted of a definite article followed by a noun and then a modifying adjective. In the

N-DROP condition, the NP consisted of a definite article and modifying adjective. In all three

conditions, the target nouns were preceded by definite articles (*el* and *la*) rather than indefinite

articles (*un* and *una*) for two reasons. First, masculine and feminine definite articles contain the

same number of letters, which makes their reading times more comparable. Second, learners are

more accurate at providing correct agreement on definite than indefinite articles (Bruhn de

Garavito & White, 2002), so definite articles were used to 'bias learners for the best.' Each of the

20 target words appeared once in each of the three conditions (i.e., DET-N, N-ADJ, and N-

DROP) to lessen the likelihood that word knowledge could mediate sensitivity to violations of

grammatical gender agreement across conditions.

In the N-ADJ and null nominal conditions, adjectives with transparent /*-o*/ and /*-a*/

endings agreed with the target noun. These adjectives were all 2-3 syllables long, were

comprised of 4-7 letters, and were not cognates. Twenty unique adjectives were used. They are

barato 'cheap', *bello* 'lovely', *blanco* 'white', *bonito* 'pretty', *bueno* 'good', *caro* 'expensive',

corto 'short', *frío* 'cold', *largo* 'long', *limpio* 'clean', *lindo* 'beautiful', *malo* 'bad', *negro*

'black', *nuevo* 'new', *pequeño* 'small', *rojo* 'red', *rosado* 'pink', *sucio* 'dirty', *tonto* 'silly', and

viejo 'old'. According to Davies' (2006) frequency dictionary, all 20 of these adjectives were in

the top 5,000 most frequent Spanish words, with an average frequency rank of 1247.55 (*SD* =

1143.86; range: 99 - 4661). In the interest of uniformity, the same 20 adjectives were used twice:

once in the N-ADJ condition and again in the N-DROP condition; however, the adjective

modified different nouns in both conditions.

Examples of DET-N (4), N-ADJ (5), and N-DROP (6) experimental sentences are

presented below. A complete list of target sentences for both masculine and feminine nouns is

presented in Appendix B.

(4)   *El chico bebe el/*la refresco cuando ve la película en el cine con su familia.*

'The boy drinks the$_{MASC}$/*the$_{FEM}$ drink$_{MASC}$ when (he) watches the movie in the

theater with his family.'

(5)   *El atleta toma el refresco frío/*fría cuando termina de correr tres millas por la*

mañana.

'The athlete drinks the drink$_{MASC}$ cold$_{MASC}$/*cold$_{FEM}$ when (he) finishes running

three miles in the morning.'

(6)   *El jefe pide el refresco grande y su empleado pide el pequeño/*la pequeña*

cuando van a McDonald's.

The boss orders the$_{MASC}$ drink$_{MASC}$ big and his employee orders the$_{MASC}$

small$_{MASC}$/*the$_{FEM}$ small$_{FEM}$ (one) when (they) go to McDonald's.

All 60 experimental sentences began with a two-word singular NP (e.g., the child, the

mother), which was the subject of the sentence. Half of these subjects were masculine and half

were feminine. The subjects were followed by a 1 to 3-syllable singular verb in the present tense and then a direct object.

In the DET-N and N-ADJ conditions, the direct object was an NP with the determiner and target noun (and adjective in the case of the N-ADJ condition) described above. These NPs were followed by the words *cuando* 'when' or *durante* 'during' and then 6-10 more words to complete the sentence.

In the null nominal condition, the direct object was an NP with a definite article, the target noun and an opaque modifying adjective[17] (e.g., *el refresco grande* in example 5). The adjective was then followed by a coordinating conjunction (either *y* 'and' or *pero* 'but'), another two-word singular NP (half masculine, half feminine) that served as the subject of the coordinating clause, and another 1 to 3-syllable verb in the present tense. This verb was followed by the target NP (i.e., the null nominal, consisting of a determiner, covertly realized target noun and modifying adjective) which was followed by the words *cuando* 'when' or *durante* 'during' and then 1 to 3 more words to complete the sentence. The linear distance between the target noun (e.g., *el refresco* in example 5) and the null nominal (e.g., *el pequeño*) was always 5-6 words (between 10-12 syllables) long. All experimental sentences were between 13 and 16 words long.

Each of the 20 experimental sentences contained the same 10 masculine target nouns, and 10 feminine target nouns. There were two versions of each experimental sentence, one grammatical, and one ungrammatical. Participants only saw one version of each sentence. For each of the 3 conditions, participants read 10 grammatical sentences and 10 ungrammatical sentences. This distribution of target nouns is presented in Figure 3.1.

---

[17] An opaque adjective was used to ensure that the learners must process the gender of the target noun rather than the subsequent adjective to be able to match the features to the null nominal.

Determiner Noun (20 sentences)		Noun Adjective (20 sentences)		Null Nominal (20 sentences)	
Masculine Target Noun	Feminine Target Noun	Masculine Target Noun	Feminine Target Noun	Masculine Target Noun	Feminine Target Noun
1 G	1 G	1 G	1 G	1 G	1 G
2 G	2 G	2 G	2 G	2 G	2 G
3 G	3 G	3 G	3 G	3 G	3 G
4 G	4 G	4 G	4 G	4 G	4 G
5 G	5 G	5 G	5 G	5 G	5 G
6 U	6 U	6 U	6 U	6 U	6 U
7 U	7 U	7 U	7 U	7 U	7 U
8 U	8 U	8 U	8 U	8 U	8 U
9 U	9 U	9 U	9 U	9 U	9 U
10 U	10 U	10 U	10 U	10 U	10 U

Figure 3.1
Distribution of experimental sentences across conditions
G = grammatical sentences; U = ungrammatical sentences.

Because reading times on the grammatical and ungrammatical sentences were compared for analyses, when determining which grammatical and ungrammatical sentences each participant read, I paid special attention to the gender of the target noun, the cognate status of the target noun, and the number of letters in each target noun and adjective to ensure that the grammatical and ungrammatical sentences were as comparable as possible.

The experimental sentences were divided across two lists, so that participant A would see the grammatical versions of sentences 1-5 and the ungrammatical versions of sentences 6-10 and participant B would see the reverse. Figure 3.1 shows the distribution of sentences for a hypothetical participant A. The 60 critical items in both lists were interspersed among 88

distractors[18]. The two presentation lists were randomized twice for a total of four lists, so that participants read the experimental sentences in one of four orders. Participants were assigned randomly to one of the four lists.

As previously mentioned, target words were repeated throughout the experimental stimuli. That is, the 20 target nouns were recycled across all three conditions, and the 20 target adjectives were also recycled across the N-ADJ and N-DROP conditions. Because repetitions of these words may increase familiarity and ultimately affect processing (Reichle, Pollatsek, Fisher, & Rayner, 1998), the lists were also pseudo-randomized so that no two of the same target nouns or adjectives appeared within 15 experimental sentences of each other.

Each experimental sentence was followed by a comprehension question to ensure that participants were focused on the meaning of the sentences while reading. For example, the comprehension question to (4), reprinted below, is the following (7):

(7) a. *El chico bebe el/*la refresco cuando ve la película en el cine con su familia.*

'The boy drinks the$_{MASC}$/*the$_{FEM}$ drink$_{MASC}$ when (he) watches the movie in the theater with his family.'

b. *¿El chico está solo?*

A: *Sí*   B: *No*

'Is the boy alone?'

A: Yes   B: No

---

[18] The distractors tested linguistic phenomena that were not directly related to the present dissertation (adverb placement, tense morphology, subject-verb agreement, and adjective placement). Half the distractors the participants read contained ungrammaticalities.

None of the questions tested the participants' comprehension of the noun or adjective. Half of the comprehension question required "yes" answers and the other half required "no" answers.

Previous sentence processing research has not reached a consensus as to how many comprehension questions participants must respond to correctly to be retained for analysis. For example, researchers have set cutoffs of 60% (Sagarra & Herschensohn, 2010a, 2010b), 63% (Jiang, 2004), 75% (Jiang et al., 2011), 80% (Jiang, 2007), and 85% (Lim & Christianson, 2014). A relatively conservative cutoff point of 80% was selected for the current study.

Apparatus. The data were collected on an EyeLink 1000 eye-tracker with a desktop tower mount. Participants rested their head in between a chin rest and a forehead rest while reading. Participants were seated approximately 60 cm from a 20-inch computer screen while completing the reading portion of the experiment. The experimental sentences were presented on a single line of text in size 18 Calibri font. Sentences were presented in black upper- and lower-case letters on a white background. The stimuli were divided into 6 blocks (25 sentences in the first four blocks, 24 sentences in the last two), so that participants could take five breaks during the experiment, one after each block. A 9-point calibration was performed at the beginning of the experiment and after each break. Drift correction was performed before each experimental sentence. Participants progressed through the experimental sentences and responded to comprehension questions by clicking buttons on a hand-held controller. The experiment was written with Experiment Builder software (SR Research Ltd.).

Additional Experimental Materials

In addition to the eye-tracking task, participants also completed a background questionnaire, reading questionnaire, vocabulary posttest and proficiency test. I describe the creation of these materials below.

Background questionnaire. Both the L2 Spanish learners and the native Spanish speakers completed a background questionnaire. Both questionnaires were modified versions of the Language Experience and Proficiency Questionnaire (LEAP-Q), a language background questionnaire designed specifically for bilinguals (see Marian, Blumenfeld, & Kaushanskaya, 2007). The LEAP-Q is a validated survey used for examining the experience and proficiency of bilinguals by means of self-report data. I modified the LEAP-Q to include questions specific to the participants of this study (e.g., the Spanish classes the L2 learners had taken at Michigan State University) and to reduce its length. The L2 learners' background questionnaire is located in Appendix C, and the native Spanish speakers' background questionnaire is located in Appendix D.

Reading questionnaire. To determine whether participants were aware that some of the sentences in the reading experiment contained errors, they completed a short questionnaire immediately after finishing the reading portion of the experiment. This reading questionnaire contained five short questions to gauge participants' awareness of the errors. The questions are listed below:

1. Did you notice anything strange about the sentences you read during the eye-tracking experiment? If so, what?

2. Where there any grammatical errors in the sentences you read during the eye-tracking experiment?

   Options:

   a. Yes

   b. No

3. What types of grammatical errors did you notice? Please list all the errors you remember, and provide examples when possible.

4. Please check off all the types of errors you noticed in the sentences. If you are unsure as to what something is, please ask the researcher:

   Options:

   a. ADVERBS appeared in the wrong place in the sentence

   b. incorrect tense (present, past, etc.) was used

   c. incorrect gender agreement between articles (e.g., el, la) and nouns

   d. ADJECTIVES appeared in the wrong place in the sentence

   e. incorrect agreement between subjects and verbs

   f. subjunctive was used incorrectly

   g. por and para were used incorrectly

   h. ser and estar were used incorrectly

   i. incorrect gender agreement between nouns and adjectives

   j. incorrect gender agreement between nouns and null nominals (e.g., el rico, la cómica)

5. What percentage of the experimental sentences (EXCLUDING comprehension questions) do you think contained grammatical errors?

The participants completed the questionnaire on Survey Gizmo, an online survey tool. All participants answered questions 1 and 2; however, if a participant selected "No" for question 2, indicating that they had not seen any errors while reading, then the survey automatically skipped questions 3 through 5, which inquired as to the nature of those ungrammaticalities. To ensure that subsequent questions did not influence responses on previous questions, each question was presented on a different page online, and participants could not navigate backwards through the survey to return to previous questions.

Vocabulary posttest. To ensure that participants were familiar with all the target words, I administered a vocabulary posttest immediately following the reading questionnaire portion of the experiment. The participants were presented with the 20 target nouns from the experiment. They were asked to first translate each noun and then mark the gender of the noun by checking off if it was masculine or feminine. Next, participants were asked to translate two of the null nominal sentences into English to ensure that they could indeed interpret those sentences. The vocabulary posttest is presented in Appendix E.

Proficiency test. All participants completed the grammar portion of the *Diploma de Español como Lengua Extranjera* (Certificates of Spanish as a Foreign Language, DELE) for intermediate learners (Instituto Cervantes, 2008). The test consisted of a total of 20 multiple choice questions in which participants must select from 3 options to fill in a blank in a paragraph. Participants received 1 point for each correct response for a total of 20 possible points. The L2 Spanish learners scored an average of 10.84 points out of 20, (*Mdn* = 10.00; *SD* = 3.24, range = 4-18) while the native Spanish speakers scored an average of 18.30 points (*Mdn* = 18.00; *SD* = 1.24, range = 16-20).

Procedures

The data were collected on an individual basis in an eye-tracking lab. Data collection lasted approximately 1.5 hours, and participants were paid $20 each for their time. Participants first read and signed the consent form and completed the eye-tracking portion of the experiment. The eye-tracking portion of the experiment began with 5 practice sentences to familiarize participants with the procedure, followed by the 148 experimental sentences (60 critical, 88 distractors). Participants were not told that the sentences contained ungrammaticalities, and none of the practice sentences were ungrammatical. Each experimental sentence was followed by a comprehension question, to ensure that participants attended to meaning while reading.

Once participants completed the eye-tracking experiment, they completed the reading questionnaire, vocabulary posttest, the background questionnaire, and finally the proficiency test on Survey Gizmo, an online survey tool.

Analysis

Areas of Interest

The dependent variables in the current study were reading times on the areas of interest. The reading times on the grammatical regions were compared to the reading times on the ungrammatical regions for each of the three syntactic context conditions. Relatively longer reading times on ungrammatical regions were assumed to reflect a processing cost, which was taken as evidence that the participant was sensitive to the violation in grammatical gender agreement.

The area of interest was different for each of the three agreement conditions. In the

DET-N agreement condition, the critical region encompassed both the determiner and noun in

each target sentence, shown in (4) reprinted as (8) below:

(8) *El chico bebe <u>el/*la refresco</u> cuando ve la película en el cine con su familia.*

Both the determiner and noun combined to form a single area of interest because readers

often skip over short words (Brysbaert, Drieghe, & Vitu, 2005; Brysbaert & Vitu, 1998; Frenck-

Mestre, 2005; Vitu, O'Regan, Inhoff, & Topolski, 1995), making two-letter articles a challenge

to measure directly with eye tracking (cf. Spinner, Gass, & Behney, 2013).

The area of interest for the N-ADJ agreement condition encompassed just the adjective,

and the area of interest for the N-DROP condition was the full null nominal (determiner and

adjective). These areas of interest are shown in examples (9) and (10) below, reprinted from

examples (5) and (6), respectively.

(9) *El atleta toma el refresco <u>frío/*fría</u> cuando termina de correr tres millas por la mañana.*

(10) *El jefe pide el refresco grande y su empleado pide <u>el pequeño/*la pequeña</u> cuando van a*

McDonald's.

To control for spillover effects, the word following the critical regions in each condition

was always either *durante* 'during' or *cuando* 'when.'

Fixation Time Measures

I investigated four fixation time measures in the present dissertation:

- First fixation duration – The duration of the first time a participant fixates on the area
  of interest.

- First-pass time – The sum of all fixations when a participant first fixates on the area of interest before that participant's gaze leaves that area. This is also known as gaze duration when the fixation area is comprised of only one word[19].

- Go-past time (also known as regression path time) – Includes the first fixation and all subsequent fixations (both in and outside of the area of interest) until exiting the word to the right (in scripts that are read from left to right).

- Total time – Sum of all fixations on a single area of interest.

First fixation duration and first-pass time are early measures of processing, and are thought to reflect word identification processes. Total time is a late processing measure, which can often indicate processing difficulty (see Pickering, Frisson, McElree, & Traxler, 2004). Go-past time is often considered both an early and late processing measure because it includes both word integration (an early measure) and the time it takes to overcome any processing difficulties and move on in the sentence (Clifton et al., 2007).

Data Cleaning

The eye-tracking data were first cleaned manually. Any fixations that were slightly above or below the target region were moved vertically so that they fell into the target regions. Missing data due to skipped words or track loss accounted for 4.9% of the total data set. Trials in which the critical region was not fixated on for at least 80 ms were removed from the data set when the fixation reports were generated, and trials that were over or under 2 standard deviations away from each participant's mean for both conditions (grammatical and ungrammatical) were

---

[19] The areas of interest for the DET-N and N-DROP conditions are comprised of two words, while the area of interest for the N-ADJ condition is comprised of only one. In the interest of consistency, the term first-pass time will be used to refer to all three areas of interest in the present dissertation.

manually replaced with a value two standard deviations away from that mean (Keating, 2014; Keating & Jegerski, 2015).

If participants indicated on the vocabulary posttest that they were unfamiliar with the translation or gender of one of the target nouns, the experimental sentences containing that noun were excluded on a participant basis from the analyses of all three conditions (DET-N, N-ADJ and N-DROP). This is a common practice in studies that investigate grammatical gender agreement (e.g., White et al., 2004), and was done to ensure that the L2 learners' knowledge of the target words approximated that of the native Spanish speakers, at least at a declarative level. A total of eight L2 Spanish learners translated at least one of the target nouns incorrectly, which resulted in a loss of another 0.89% of the data set. Thirteen L2 Spanish learners marked the incorrect gender for at least one of the target nouns, which resulted in a loss of 1.43% of the data set.

All participants correctly translated the null nominal sentences from Spanish into English on the vocabulary posttest (see Appendix E); however, if they could not, their reading times in the N-DROP condition would have been removed from the analyses.

Statistical Analyses

These data were analyzed with linear mixed-effects models (Baayen, Davidson, & Bates, 2008). Linear mixed-effects models can be used for data with repeated measures, and for data that are hierarchical or clustered. They differ from means-based inferential statistics, such as ANOVAs, in many ways. First, instead of comparing the means of participants per condition, mixed-effects models consider each of the participants' observations separately. Second, mixed-effects models are considered to be more robust than ANOVAs because they allow for all the factors under investigation to be considered simultaneously (Cunnings, 2012; Cunnings &

Finlayson, 2015; Jaeger, 2008; Keating & Jegerski, 2015; Plonsky, 2013), thereby obviating the need for separate by-participants and by-items analyses (Cunnings, 2012; Cunnings & Finlayson, 2015). Another benefit of linear mixed-effects models is that, unlike ANOVAs, they are relatively robust against missing data (Quené & van den Bergh, 2004), which is not uncommon in eye-tracking research because of measurement error. They are also robust against unequal sample sizes per group (Quené & van den Bergh, 2004).

Linear mixed-effects models measure how well an outcome variable can be predicted by fixed and random effects. Fixed effects model how the independent variable(s) affect(s) the outcome variable, while random effects model variance that can be attributed to other factors inherent in the sampling of the study, such as subject or item variance (Cunnings, 2012). Whereas fixed effects are hypothesized to have systematic and predictable effects on the outcome variable, random effects are expected to be idiosyncratic and unpredictable (Cunnings, 2012). In this dissertation, the outcome variable was one of four different reading time measures (first fixation duration, first-pass time, go-past time and total time). The fixed effects were group (native Spanish speakers, L2 Spanish learners) and grammaticality (grammatical, ungrammatical). Subject and item were also included as crossed random effects. For each of the four fixation time measures, each participant yielded 20 observations per structure. Separate analyses were conducted for each of the three syntactic context conditions (DET-N, N-ADJ, N-DROP) because, for practical reasons, the areas of interest could not be controlled for length and frequency across those three conditions, thus making direct within-analysis comparisons impossible.

The data were first checked to make sure they met the assumptions of linear mixed-effects models. One assumption is that the data should be normally distributed. An initial

exploration into the data for this dissertation revealed that the reading times were, in fact, skewed to the right. I have showed this lack of normality in the data in Figure 3.2. Another assumption of linear mixed-effects models is the absence of heteroscedasticity[20], meaning that the variance in the residuals should be relatively equal across the range of the predicted values. The residual plot in Figure 3.3, created with 800 fictitious and random data points, shows residuals that are homoscedastic, as there is no obvious pattern in the plot. Figure 3.4 shows the residuals for one of the models run in this dissertation with total reading time in the DET-N condition as the outcome variable. The plot is cone-shaped, indicating that the larger the predicted means are, the larger the residuals are.
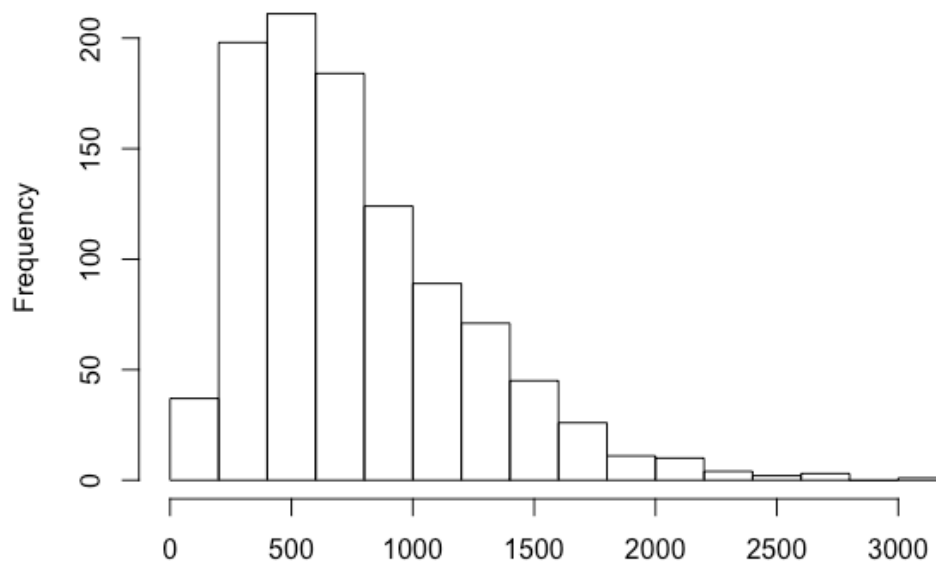
Figure 3.2
Histogram depicting total time data for the DET-N condition

---

[20] Some researchers vary in their opinion on how important this assumption is. For example, Winter (2013) notes that linear mixed-effects models cannot violate the assumption of absence of heteroscedasticity while Quené and van den Bergh (2004) claim they are still robust in the face of heteroscedasticity.
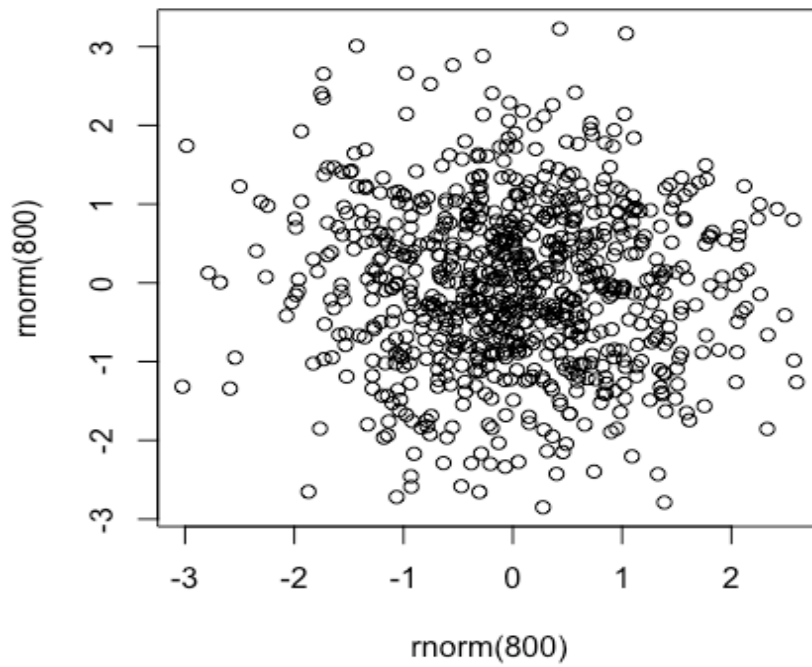
Figure 3.3
Residual plot from fictitious data that does not violate the assumption of absence of heteroscedasticity
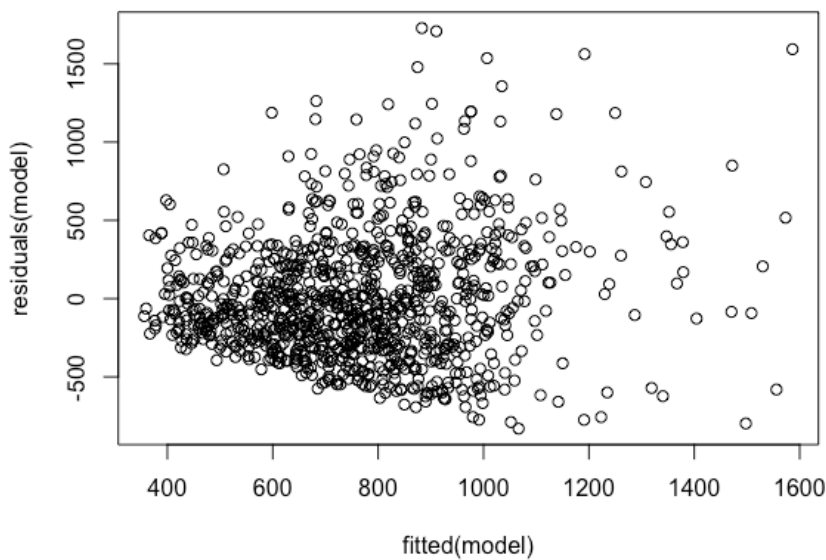
Figure 3.4
Residual plot depicting total time data for the DET-N condition

To account for these violations of the assumptions of the model, a log transformation was performed on all the outcome variables (i.e., the different reading time measures). This log

transformation corrected the problems of lack of normality and absence of homoscedasticity, as can be seen in Figures 3.5 and 3.6, respectively.
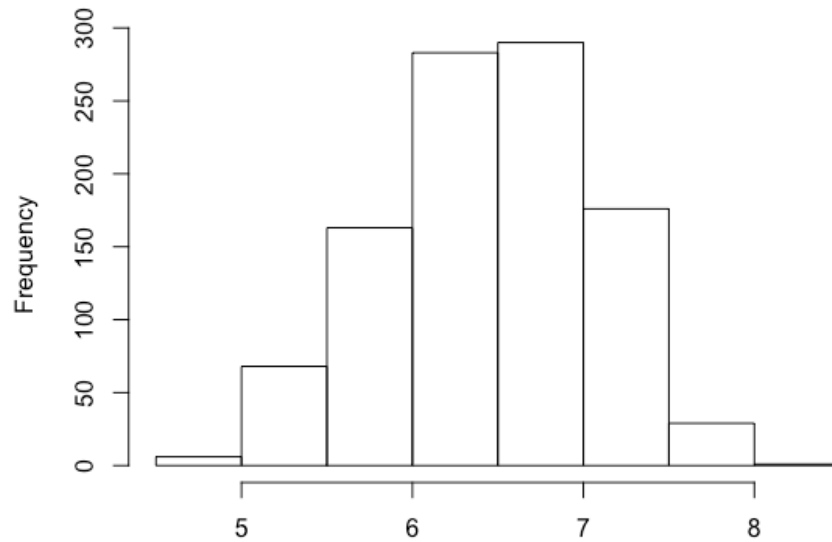
Figure 3.5
Histogram depicting the log-transformed total time data for the DET-N condition
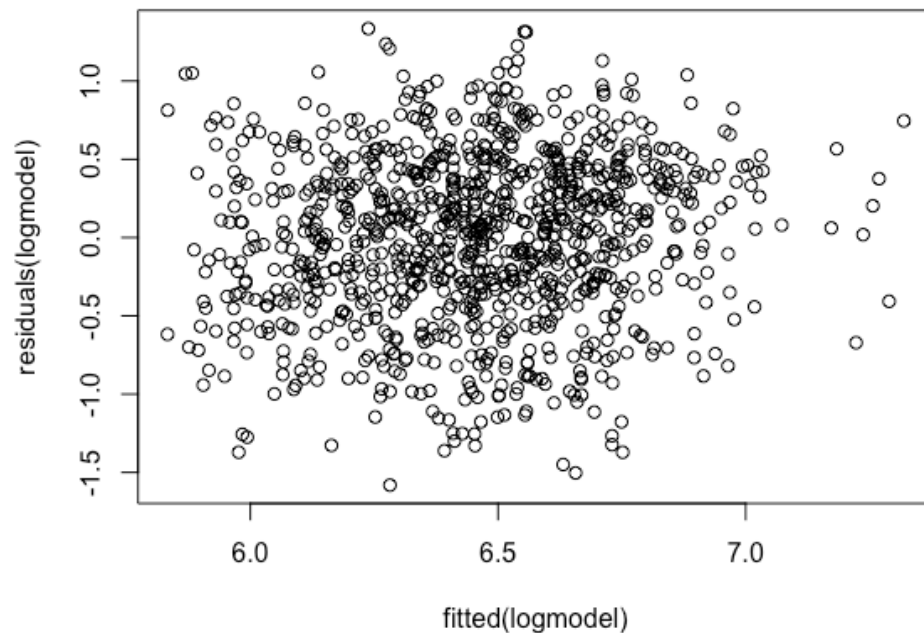
Figure 3.6
Residual plot depicting the log-transformed total time data for the DET-N condition

To summarize, I analyzed the data with linear mixed-effects models with reading times as the outcome variables, participant group and grammaticality as fixed effects, and subject and

item as random effects. All the outcome variables were log-transformed. Each participant

contributed 20 data points to each of the three conditions (DET-N, N-ADJ, N-DROP), 10 for

grammatical items and 10 for ungrammatical items. I conducted four tests within each condition,

one for each of the four reading time measures specified as outcome variables (first fixation

duration, first-pass time, go-past time and total time). The data were analyzed using models with

and without an interaction between group and grammaticality. Interactions are only reported

when they were significant. All statistics for this dissertation were computed in R (R Core Team,

2015) with the lme(4) package (Bates, Mächler, Bolker, & Walker, 2015). For the interested

reader, Cunnings (2012), Cunnings and Finlayson (2015), Gries (2015) and Winter (2013) all

provide clear explanations on how to compute linear mixed-effects models in R.

CHAPTER 4: RESULTS

This chapter presents the results of this dissertation and is divided into three primary sections. In the first section, I will present the statistical analyses detailing the native speaker and L2 learners' sensitivity to grammatical violations for each of the three conditions (DET-N, N-ADJ and N-DROP) separately. I will first present the descriptive statistics for each of the four outcome variables: first fixation duration, first-pass time, go-past time and total time. I will then present the inferential statistics for each of the four outcome variables.

In the second section, I will present the findings of additional analyses examining only the L2 learners' sensitivity to grammatical violations by the gender of the target nouns. The rationale for and description of these additional analyses will be presented at the beginning of this section. Then, I will present the descriptive and inferential statistics for each of the four outcome variables in all three conditions (some of these analyses can also be found in Appendix H).

In the third section, I will present results examining sensitivity to grammatical violations on an individual participant basis. In this section, I explore the results of the reading questionnaire, and also triangulate L2 learners' reading times in the eye-tracking experiment with their responses on the reading questionnaire.

Statistical Analyses: Native Speakers' and L2 Learners' Reading Times

on Grammatical and Ungrammatical Sentences

Determiner-Noun Agreement

Descriptive statistics. Table 4.1 depicts the mean fixation times for each of the four fixation measures for the DET-N condition. As is typical in this kind of research, the native Spanish speakers tended to have shorter reading times than the L2 Spanish learners across reading measures. Both native Spanish speakers and L2 Spanish learners have longer reading times for ungrammatical regions relative to grammatical ones for first-pass, go-past and total time.

Table 4.1
Mean (Standard Deviation) Fixation Times for DET-N Condition

	First Fix.	First-Pass	Go-Past	Total
Native Speakers				
G	242.00	341.94	461.32	592.19
	(98.91)	(187.12)	(308.41)	(385.18)
UG	244.08	411.23	591.72	754.86
	(115.52)	(257.11)	(429.39)	(431.26)
Difference	2.08	69.29	130.40	162.67
L2 Learners				
G	273.81	481.99	626.46	781.90
	(106.33)	(284.37)	(413.40)	(453.82)
UG	254.93	524.89	713.13	938.27
	(100.30)	(284.67)	(463.47)	(507.90)
Difference	-18.88	42.90	86.67	156.37

Note. First Fix, First Fixation Duration; First-Pass, First-Pass Time; Go-Past, Go-past Time; Total, Total Time; G, Grammatical; UG, Ungrammatical. All times are presented in milliseconds.

First fixation duration. The results of the linear mixed-effects model for first fixation duration in the DET-N condition are presented in Table 4.2. The native Spanish speakers' reading times for the grammatical condition were statistically different from zero ($t = 179.64$, $p < .001$). First fixation duration showed an effect of group ($t = 2.04$, $p = .046$) with L2 Spanish

learners reading 8%[21] more slowly than the native Spanish speakers. There was no main effect of grammaticality (t = -1.01, p < .311). The final model did not include an interaction term.

Table 4.2
Model Results for First Fixation Duration in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of being a native speaker for the grammatical condition	Intercept	5.43	0.03	179.64	< .001
Overall main effect of being an L2 learner	Group	0.08	0.04	2.04	.046
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.02	0.02	-1.01	.311
	Random Effects				
	Subject	0.014			
	Item	0.000			
	Residual	10.132			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

First-pass time. The results of the linear mixed-effects model for first-pass time in the DET-N condition are presented in Table 4.3. The native Spanish speakers' reading times for the grammatical condition were statically different from zero (t = 125.09, p < .001). First-pass time showed an effect of group (t = 5.26, p = < .001), with L2 Spanish learners reading 29% more slowly than the native speakers. There was also an effect of grammaticality (t = 3.56, p < .001), with participants reading in the ungrammatical condition 11% more slowly than the grammatical condition. The final model did not include an interaction term.

---

[21] Because a log transformation was performed on the outcome variable, and the two predictor variables are binary, the coefficients can be multiplied by 100 to calculate percent change from the reference category.

Table 4.3
Model Results for First-Pass Time in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of being a native speaker for the grammatical condition	Intercept	5.72	0.05	125.09	< .001
Overall main effect of being an L2 learner	Group	0.29	0.06	5.26	< .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.11	0.03	3.56	< .001
	Random Effects				
	Subject	0.027			
	Item	0.006			
	Residual	0.261			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Go-past time. The results of the linear mixed-effects model for go-past time in the DET-N condition are presented in Table 4.4. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero ($t = 113.82, p < .001$). Go-past time showed an effect of group ($t = 4.24, p = < .001$), with L2 Spanish learners reading in the grammatical condition 27% more slowly than the native speakers. There was also an effect of grammaticality ($t = 4.90, p < .001$), with participants reading in the ungrammatical condition 17% more slowly than the grammatical condition. The final model did not include an interaction term.

Table 4.4
Model Results for Go-Past Time in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of being a native speaker for the grammatical condition	Intercept	5.98	0.05	113.82	< .001
Overall main effect of being an L2 learner	Group	0.27	0.06	4.24	< .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.17	0.04	4.90	< .001
	Random Effects				
	Subject	0.038			
	Item	0.009			
	Residual	0.294			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Total time. The results of the linear mixed-effects model for total time in the DET-N condition are presented in Table 4.5. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero ($t = 114.63$, $p < .001$). Total time showed an effect of group ($t = 3.83$, $p = < .001$), with L2 Spanish learners reading in the grammatical condition 27% more slowly than the native speakers. There was also an effect of grammaticality ($t = 6.38$, $p < .001$), with participants reading the ungrammatical sentences 22% more slowly than the grammatical sentences. The final model did not include an interaction term.

Table 4.5
Model Results for Total Time in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of being a native speaker for the grammatical condition	Intercept	6.22	0.05	114.63	< .001
Overall main effect of being an L2 learner	Group	0.27	0.07	3.83	< .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.22	0.03	6.38	< .001
	Random Effects				
	Subject	0.049			
	Item	0.005			
	Residual	0.297			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Summary of results. All four of the models in the DET-N condition showed a main effect of group, with L2 learners reading more slowly than native Spanish speakers. For first-pass time, go-past time, and total time, there was also a main effect for grammaticality, with participants reading more slowly in the ungrammatical condition relative to the grammatical one. There were no interactions between group and grammaticality for any of the four outcome variables.

Noun-Adjective Agreement

Descriptive statistics. Table 4.6 depicts the mean fixation times for each of the four fixation measures for the N-ADJ condition. The native Spanish speakers generally tended to have shorter reading times than L2 Spanish learners. The native Spanish speakers evidenced increased reading times for ungrammatical regions for all four fixation measures with the largest increase showing up in total time. The L2 learners evidenced reading times that were roughly equivalent for first fixation, first-pass and go-past reading times, with a slight increase for the ungrammatical region in total time.

Table 4.6

Mean (Standard Deviation) Fixation Times for N-ADJ Condition

	First Fix.	First-Pass	Go-Past	Total
Native Speakers				
G	275.30	300.51	371.71	425.29
	(98.538)	(113.99)	(219.77)	(235.81)
UG	289.67	331.13	457.37	592.25
	(121.10)	(144.09)	(316.25)	(393.39)
Difference	14.37	30.62	85.66	166.96
L2 Learners				
G	298.56	365.64	461.38	531.04
	(114.49)	(144.04)	(293.81)	(288.88)
UG	290.81	352.58	464.71	570.40
	(106.58)	(183.99)	(328.88)	(358.88)
Difference	-7.75	-13.06	3.33	39.36

Note. First Fix, First Fixation Duration; First-Pass, First-Pass Time; Go-Past, Go-past Time; Total, Total Time; G, Grammatical; UG, Ungrammatical. All times are presented in milliseconds.

First fixation duration. The results of the linear mixed-effects model for first fixation duration in the N-ADJ condition are presented in Table 4.7. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero ($t = 195.52$, $p < .001$). However, first fixation duration did not show a main effect of group ($t = 1.41$, $p = .166$), or grammaticality ($t = 0.39$, $p = .694$). The final model did not include an interaction term.

Table 4.7

Model Results for First Fixation Duration in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of being a native speaker for the grammatical condition	Intercept	5.56	0.03	195.52	< .001
Overall main effect of being an L2 learner	Group	0.05	0.04	1.41	= .166
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.01	0.02	0.39	= .694
	Random Effects				
	Subject	0.008			
	Item	0.002			
	Residual	0.132			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

First-pass time. The results of the linear mixed-effects model for first-pass time in the N-ADJ condition are presented in Table 4.8. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero ($t = 141.79$, $p < .001$). First-pass time showed a main effect of group ($t = 4.34$, $p < .001$) and grammaticality ($t = 2.78$, $p <.001$), and a group*grammaticality interaction ($t = -3.15$, $p < .001$). The native Spanish speakers had reading times that were 9% longer in the ungrammatical condition relative to the grammatical one, and post hoc analyses confirmed that this difference was statistically significant ($β = -.10$, $t = -2.78$, $p = .029$). L2 Spanish learners had reading times that were 7%[22] shorter on the ungrammatical condition, but this difference was not statistically significant ($β = .06$, $t = 1.67$, $p = .343$).

---

[22] This percentage is computed by using the coefficients to first calculate the percentage change for the L2 learners' reading times in the grammatical condition relative to the reference category, (5.62+.20 = 5.82), then to calculate the percentage change for the L2 learners' reading times in the ungrammatical condition relative to the reference category (5.62 + .20 +.09 - .16 = 5.75) and then calculating the difference between the two and multiplying by 100.

Table 4.8
Model Results for First-Pass Time in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of being a native speaker for the grammatical condition	Intercept	5.62	0.04	141.79	< .001
Overall main effect of being an L2 learner	Group	0.20	0.05	4.34	< .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.09	0.04	2.78	< .001
Interaction between group and grammaticality	Group*Grammaticality	-0.16	0.05	-3.15	< .001
	Random Effects				
	Subject	0.011			
	Item	0.011			
	Residual	0.144			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Go-past time. The results of the linear mixed-effects model for go-past time in the N-ADJ condition are presented in Table 4.9. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero (t = 117.85, p < .001). Go-past time showed a main effect of group (t = 3.40, p < .001) and grammaticality (t = 3.77, p <.001), and a group*grammaticality interaction (t = -3.07, p < .001). According to the model, the native Spanish speakers had reading times that were 17% longer in the ungrammatical condition relative to the grammatical one, and post hoc analyses confirmed that this difference was statistically significant (β = -.17, t = -3.77, p = .001). The L2 Spanish learners had reading times that were 3% shorter on the ungrammatical sentences, but this difference was not statistically significant (β = .03, t = 0.56, p = .943).

Table 4.9
Model Results for Go-Past Time in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of being a native speaker for the grammatical condition	Intercept	5.78	0.05	117.85	< .001
Overall main effect of being an L2 learner	Group	0.21	0.06	3.40	< .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.17	0.05	3.77	< .001
Interaction between group and grammaticality	Group*Grammaticality	-0.20	0.07	-3.07	< .001
	Random Effects				
	Subject	0.023			
	Item	0.010			
	Residual	0.243			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Total time. The results of the linear mixed-effects model for total time in the N-ADJ condition are presented in Table 4.10. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero ($t = 95.77$, $p < .001$). Total time showed a main effect of group ($t = 3.40$, $p = .001$) and grammaticality ($t = 6.62$, $p < .001$), and a group*grammaticality interaction ($t = -4.11$, $p < .001$). According to the model, the native Spanish speakers had reading times that were 29% longer in the ungrammatical condition relative to the grammatical one, and post hoc analyses confirmed that this difference was statistically significant (β = -.29, $t = -6.62$, $p = < .001$). The L2 Spanish learners had reading times that were 3% longer, and post hoc analyses confirmed that this difference was not statistically significant (β = -.03, $t = -0.70$, $p = .896$).

Table 4.10
Model Results for Total Time in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of being a native speaker for the grammatical condition	Intercept	5.91	0.06	95.77	< .001
Overall main effect of being an L2 learner	Group	0.23	0.07	3.40	= .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.29	0.04	6.62	< .001
Interaction between group and grammaticality	Group*Grammaticality	-0.26	0.06	-4.11	< .001
	Random Effects				
	Subject	0.036			
	Item	0.030			
	Residual	0.243			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Summary of results. Three of the four models (first-pass time, go-past time and total time) in the N-ADJ condition showed a main effect of group, with L2 learners reading more slowly than native Spanish speakers. The same three models also showed a main effect of grammaticality, and a group*grammaticality interaction. While the native speakers consistently evidenced longer reading times (between 9% and 29%) in the ungrammatical condition relative to the grammatical one, the L2 learners showed little difference in reading times between the grammatical and ungrammatical conditions.

Null Nominal Agreement

Descriptive statistics. Table 4.11 depicts the mean fixation times for each of the four fixation measures for the N-DROP condition. The native Spanish speakers generally tended to have shorter reading times than L2 Spanish learners. The native Spanish speakers evidenced increased reading times for ungrammatical regions for all four fixation measures with the largest

increase showing up in total time, and the smallest increase for first fixation duration. The L2

learners evidenced reading times that were roughly equivalent for all four reading times.

Table 4.11

Mean (Standard Deviation) Fixation Times for N-DROP Condition

	First Fix.	First-Pass	Go-Past	Total
Native Speakers				
G	242.84	301.56	374.02	487.91
	(76.25)	(145.00)	(238.72)	(308.39)
UG	258.67	357.72	449.08	624.87
	(100.76)	(180.19)	(273.58)	(420.26)
Difference	15.83	56.16	75.06	136.96
L2 Learners				
G	261.69	376.11	449.69	627.37
	(93.21)	(189.30)	(264.70)	(423.08)
UG	261.88	381.30	481.96	614.71
	(91.99)	(220.57)	(458.01)	(473.02)
Difference	0.19	5.19	32.27	-12.66

Note. First Fix, First Fixation Duration; First-Pass, First-Pass Time; Go-Past, Go-past Time; Total, Total Time; G, Grammatical; UG, Ungrammatical. All times are presented in milliseconds.

First fixation duration. The results of the linear mixed-effects model for first fixation

time in the N-DROP condition are presented in Table 4.12. The average native Spanish speaker's

reading times for the grammatical condition were statistically different from zero ($t = 190.62$, $p <$

.001). However, first fixation duration did not show a main effect of group ($t = 1.26$, $p = .215$) or

grammaticality ($t = 1.26$, $p = .208$). The final model did not include an interaction term.

65

Table 4.12

Model Results for First Fixation Duration in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of being a native speaker for the grammatical condition	Intercept	5.46	0.03	190.62	< .001
Overall main effect of being an L2 learner	Group	0.04	0.04	1.26	= .215
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.03	0.02	1.26	= .208
	Random Effects				
	Subject	0.011			
	Item	0.003			
	Residual	0.095			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

First-pass time. The results of the linear mixed-effects model for first-pass time in the

N-DROP condition are presented in Table 4.13. The average native Spanish speaker's reading

times for the grammatical condition were statistically different from zero ($t = 118.01$, $p < .001$).

First-pass time showed a main effect of group ($t = 3.73$, $p < .001$) and grammaticality ($t = 3.86$, p

$<.001$), and a group*grammaticality interaction ($t = -2.80$, $p = .005$). The native Spanish speakers

had reading times that were 15% longer in the ungrammatical condition relative to the

grammatical one, and post hoc analyses revealed that this difference was statistically significant

($β = -.29$, $t = -6.62$, $p = < .001$). The L2 Spanish learners did not evidence any percent change in

reading time, which was confirmed by post hoc analyses ($β = .01$, $t = 0.15$, $p = 1.000$).

Table 4.13
Model Results for First-Pass Time in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of being a native speaker for the grammatical condition	Intercept	5.61	0.05	118.01	< .001
Overall main effect of being an L2 learner	Group	0.21	0.06	3.73	< .001
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.15	0.04	3.86	< .001
Interaction between group and grammaticality	Group*Grammaticality	-0.15	0.05	-2.80	= .005
	Random Effects				
	Subject	0.022			
	Item	0.014			
	Residual	0.182			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Go-past time. The results of the linear mixed-effects model for go-past time in the N-DROP condition are presented in Table 4.14. The average native Spanish speaker's reading times for the grammatical condition were statistically different from zero ($t = 95.80$, $p < .001$). Go-past time showed a main effect of group ($t = 2.78$, $p = .007$) and grammaticality ($t = 3.50$, $p <.001$), and a group*grammaticality interaction ($t = -2.60$, $p = .009$). The native Spanish speakers had reading times that were 15% longer in the ungrammatical condition relative to the grammatical one, and post hoc analyses confirmed that this difference was statistically significant ($\beta = -0.15$, $t = -3.50$, $p = .003$). The L2 Spanish learners had reading times that were 1% slower, but this difference was not statistically significant ($\beta = 0.01$, $t = 0.22$, $p = .996$).

Table 4.14
Model Results for Go-Past Time in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of being a native speaker for the grammatical condition	Intercept	5.79	0.06	95.80	< .001
Overall main effect of being an L2 learner	Group	0.18	0.07	2.78	= .007
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.15	0.04	3.50	< .001
Interaction between group and grammaticality	Group*Grammaticality	-0.16	0.06	-2.60	= .009
	Random Effects				
	Subject	0.029			
	Item	0.032			
	Residual	0.242			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Total time. The results of the linear mixed-effects model for total time in the N-DROP condition are presented in Table 4.15. The average native Spanish speaker's reading times for the grammatical condition was different from zero ($t = 72.51$, $p < .001$). Total time showed a main effect of group ($t = 2.66$, $p = .009$) and grammaticality ($t = 4.30$, $p < .001$), and a group*grammaticality interaction ($t = -3.64$, $p < .001$). The native Spanish speakers had reading times that were 19% longer in the ungrammatical condition relative to the grammatical one, and post hoc analyses confirmed that these differences were statistically significant ($\beta = -0.19$, $t = -4.31$, $p < .001$). The L2 Spanish learners had reading times that were 4% shorter in the ungrammatical condition, but this difference was not statistically significant ($\beta = 0.04$, $t = 0.88$, $p = .818$).

Table 4.15
Model Results for Total Time in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of being a native speaker for the grammatical condition	Intercept	6.03	0.08	72.51	< .001
Overall main effect of being an L2 learner	Group	0.22	0.08	2.66	= .009
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.19	0.05	4.30	< .001
Interaction between group and grammaticality	Group*Grammaticality	-0.23	0.06	-3.64	< .001
	Random Effects				
	Subject	0.061			
	Item	0.073			
	Residual	0.257			

Note. Group and grammaticality are both categorical variables with two levels. Native speaker reading times on grammatical sentences were taken as the reference category (Intercept).

Summary of results. Three of the four models (first-pass time, go-past time and total time) in the N-DROP condition showed a main effect of group, with L2 learners reading more slowly than native Spanish speakers. The same three models also showed a main effect of grammaticality, and a group*grammaticality interaction. While the native speakers consistently evidenced longer reading times (between 15% and 19%) in the ungrammatical condition relative to the grammatical one, the L2 learners showed little difference in reading times between the grammatical and ungrammatical conditions.

General Summary of Statistical Results

This section presented the results on native and L2 learners' reading times for grammatical and ungrammatical sentences with grammatical gender agreement under three conditions: DET-N, N-ADJ, and N-DROP. Table 4.16 summarizes the statistical results related to whether or not both native speakers and L2 learners slowed down in the ungrammatical condition. To give the reader an indication of the magnitude of the sensitivity to violations across

the different sentential contexts, Figure 4.1 summarizes the participants' percent change in

reading time, as estimated by the coefficients in the linear mixed-effects models described above.

Positive numbers indicate a longer reading time in the ungrammatical condition.

Table 4.16
Summary of Statistical Analyses

	First-Fixation	First-Pass	Go-Past	Total Time
	Slowed down in ungrammatical condition?	Slowed down in ungrammatical condition?	Slowed down in ungrammatical condition?	Slowed down in ungrammatical condition?
Native Speakers				
DET-N	No	Yes	Yes	Yes
N-ADJ	No	Yes	Yes	Yes
N-DROP	No	Yes	Yes	Yes
L2 Learners				
DET-N	No	Yes	Yes	Yes
N-ADJ	No	No	No	No
N-DROP	No	No	No	No
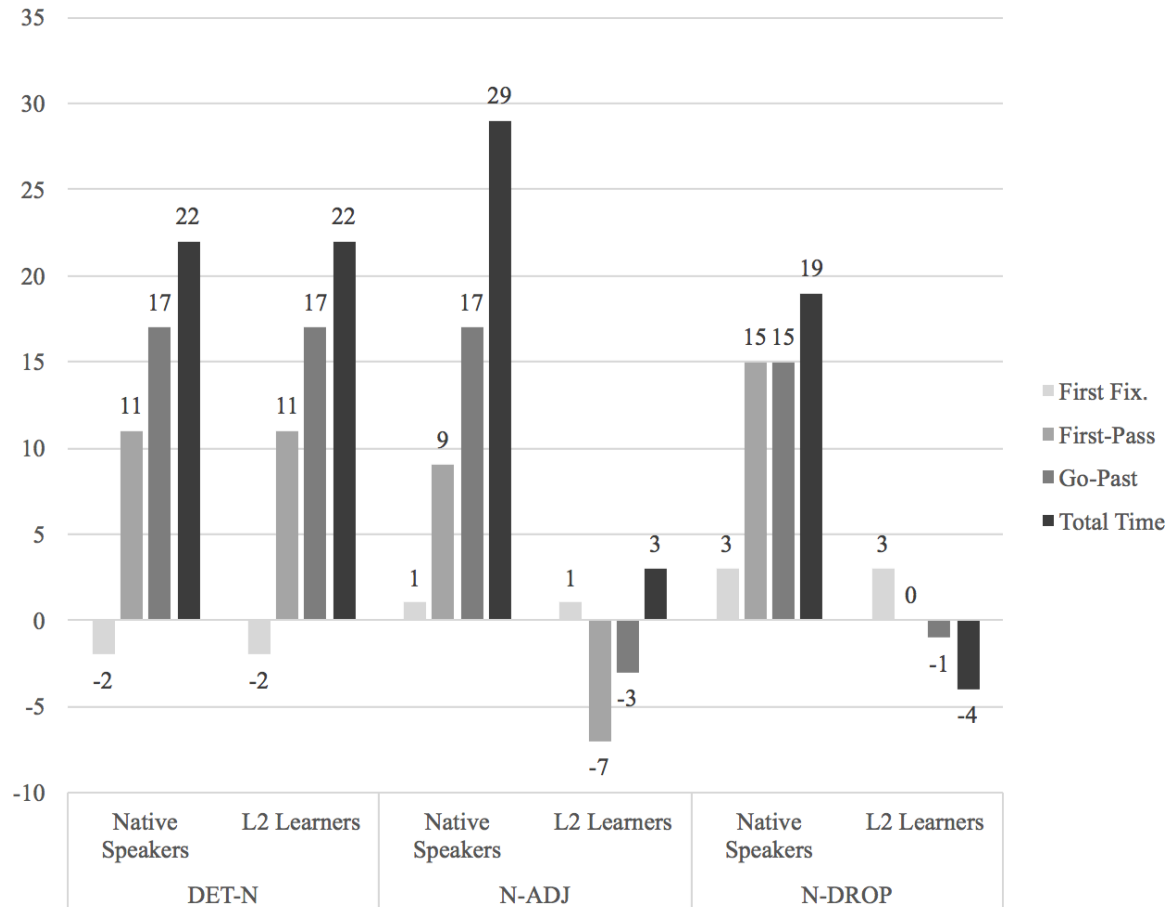
Figure 4.1
Participants' percent change in reading time between grammatical and ungrammatical sentences

Additional Statistical Analyses: L2 Learners' Reading Times on Grammatical and Ungrammatical Sentences by Gender

In the previous section, the L2 learners were sensitive to violations of DET-N agreement, as evidenced by their statistically slower reading times on first-pass, go-past and total time on ungrammatical areas of interest. However, because nouns of both genders were grouped for these analyses, the results cannot determine if the gender of the target nouns plays a role in the L2 learners' sensitivity to ungrammaticalities. That is, it could be that L2 learners' sensitivity to violations is asymmetrical. Previous research has indicated that L2 learners exhibit a masculine

default (e.g., Franceschina, 2001; McCarthy, 2007; Montrul et al., 2008; White et al., 2004), which would mean that the learners may exhibit greater sensitivity to violations such as (11a) compared to (11b):

(11) a. *la*   *proyecto*

    the FEM   project MASC

    'the project'

  b. *el*   *manzana*

    the MASC   apple FEM

    'the apple'

In this section, I aim to determine: (a) if the sensitivity to grammatical violations that L2 learners evidenced in the DET-N condition is asymmetrical and, (b) if the L2 learners' apparent lack of sensitivity found in previous statistical analyses can be attributed to the gender of the target noun acting as an intervening variable.

In the analyses of the previous section, the fixed effects of the mixed-effects models were group (native Spanish speakers, L2 Spanish learners) and grammaticality (grammatical, ungrammatical). Subject and item were included as random effects. One benefit of linear mixed-effects models is that many factors can be included as fixed and random effects in a single model. The number of factors that can be included in a given model, however, is not limitless: complex models with many fixed and random effects require large sample sizes so the model can be adequately fit to the data. The gender of the target noun (masculine, feminine) could have been included as a third fixed effect in the models from the previous section; however, when I tried to do this the models failed to converge, meaning that the model was too complex for the

sample size. Results from models that fail to converge are not reliable and therefore should not be reported.

Instead, I opted to run a separate set of analyses on only the L2 learners' reading times. In this way, group (native Spanish speakers, L2 learners) could be dropped as a fixed effect and replaced by gender (masculine, feminine). Therefore, the fixed effects for these additional analyses were gender (masculine, feminine) and grammaticality (grammatical, ungrammatical), and subject and item were included again as random effects[23]. I conducted a log transformation on all the outcome variables (i.e., the different reading time measures) to correct for problems of normality and heteroscedasticity. The data were analyzed using models with and without an interaction between gender and grammaticality. Interactions are only reported when they were significant.

Descriptive Statistics

Table 4.17 depicts the L2 learners' mean fixation times for each of the four fixation measures for all three conditions.

---

[23] Even though I ran these analyses on a subset of the data, thus lowering the sample size, none these models failed to converge.

Table 4.17

Mean (Standard Deviation) Fixation Times for L2 Learners by Gender of Target Noun

	First Fix.	First-Pass	Go-Past	Total
DET-N				
Masculine				
G	284.92	481.30	642.21	811.39
	(125.72)	(306.58)	(429.17)	(469.77)
UG	248.19	554.37	730.19	934.08
	(93.10)	(303.76)	(493.16)	(509.61)
Difference	-36.73	73.07	87.98	122.69
Feminine				
G	262.89	482.68	610.45	752.18
	(103.87)	(261.55)	(397.87)	(437.05)
UG	261.56	495.41	695.93	942.49
	(106.89)	(262.17)	(432.87)	(508.27)
Difference	-1.33	12.73	85.48	190.31
N-ADJ				
Masculine				
G	294.21	341.76	449.74	520.82
	(115.56)	(136.05)	(316.33)	(296.78)
UG	288.79	342.48	446.33	570.88
	(107.86)	(179.56)	(309.34)	(337.90)
Difference	-5.42	0.72	-3.41	50.06
Feminine				
G	302.91	389.72	473.23	541.52
	(113.74)	(148.39)	(269.84)	(281.43)
UG	292.71	362.00	481.69	569.94
	(105.80)	(189.19)	(346.40)	(379.16)
Difference	-10.20	-27.72	8.46	28.42
N-DROP				
Masculine				
G	269.22	390.47	472.30	625.61
	(105.27)	(198.26)	(271.12)	(436.54)
UG	265.87	376.82	510.27	613.66
	(102.69)	(204.32)	(568.68)	(487.10)
Difference	-3.35	-13.65	37.97	-11.95
Feminine				
G	253.91	361.26	426.51	629.18
	(78.54)	(179.19)	(257.20)	(410.64)
UG	257.78	385.86	452.92	615.80
	(79.79)	(236.76)	(305.51)	(460.10)
Difference	3.87	24.60	26.41	-13.38

Note. First Fix, First Fixation Duration; First-Pass, First-Pass Time; Go-Past, Go-past Time; Total, Total Time; G, Grammatical; UG, Ungrammatical. All times are presented in milliseconds.

In the DET-N condition, the L2 learners evidenced substantially longer reading times on ungrammatical sentences. Numerically, this difference was greater for the masculine nouns during first-pass, but for the feminine nouns in total time. For the N-ADJ and N-DROP conditions, the L2 learners did not evidence substantially longer reading times on the ungrammatical sentences for feminine or masculine noun. These descriptive statistics indicate that the L2 learners may evidence asymmetric sensitivity in the DET-N condition, but likely not in the N-ADJ and N-DROP condition. Below, I will explore in detail the statistical analyses performed on these data in the DET-N condition by reviewing the analyses for each of the four fixation measures: first fixation, first-pass, go-past and total time. The statistical analyses for the N-ADJ and N-DROP condition are located in Appendix F.

First Fixation Duration

The results of the linear mixed-effects model for first fixation duration in the DET-N condition are presented in Table 4.18. The L2 learners' reading times when reading grammatical sentences with a masculine target noun were statistically different from zero ($t = 159.70$, $p <$ .001). First fixation duration showed no main effect of gender ($t = -0.40$, $p = .691$) or grammaticality ($t = -1.78$, $p = .075$), although the latter did approach significance. The final model did not include an interaction term.

Table 4.18

Gender Model Results for L2 Learners' First Fixation Duration in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	5.54	0.04	159.70	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.01	0.04	-0.40	= .691
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.06	0.04	-1.78	= .075
	Random Effects				
	Subject	0.008			
	Item	0.000			
	Residual	0.142			

Note. Gender and grammaticality are both categorical variables with two levels. L2 learner reading times on grammatical sentences with a masculine target noun were taken as the reference category (Intercept).

First-Pass Time

The results of the linear mixed-effects model for first-pass time in the DET-N condition are presented in Table 4.19. The L2 learners' reading times when reading grammatical sentences with a masculine target noun were statistically different from zero ($t = 104.53$, $p < .001$). First-pass time showed no main effect of gender ($t = -0.71$, $p = .484$) or grammaticality ($t = 1.59$, $p = .113$). The final model did not include an interaction term.

Table 4.19
Gender Model Results for L2 Learners' First-Pass Time in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	6.06	0.06	104.53	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.04	0.06	-0.71	= .484
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.08	0.05	1.59	= .113
	Random Effects				
	Subject	0.024			
	Item	0.007			
	Residual	0.275			

Note. Gender and grammaticality are both categorical variables with two levels. L2 learner reading times on grammatical sentences with a masculine target noun were taken as the reference category (Intercept).

Go-Past Time

The results of the linear mixed-effects model for go-past time in the DET-N condition are presented in Table 4.20. The L2 learners' reading times when reading grammatical sentences with a masculine target noun were statistically different from zero ($t = 102.26, p < .001$). Go-past time showed no main effect of gender ($t = -0.58, p = .566$), but did show a main effect of grammaticality ($t = 2.68, p = .008$), with L2 learners reading in the ungrammatical condition 13% more slowly than the grammatical condition. The final model did not include an interaction term.

Table 4.20

Gender Model Results for L2 Learners' Go-Past Time in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	6.29	0.06	102.26	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.04	0.06	-0.58	= .566
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.13	0.05	2.68	= .008
	Random Effects				
	Subject	0.029			
	Item	0.009			
	Residual	0.282			

Note. Gender and grammaticality are both categorical variables with two levels. L2 learner reading times on grammatical sentences with a masculine target noun were taken as the reference category (Intercept).

Total Time

The results of the linear mixed-effects model for total time in the DET-N condition are presented in Table 4.21. The L2 learners' reading times when reading grammatical sentences with a masculine target noun were statistically different from zero ($t = 110.32, p < .001$). Total time showed no main effect of gender ($t = -0.83, p = .417$), but did show a main effect of grammaticality ($t = 3.94, p < .001$), with L2 learners reading in the ungrammatical condition 19% more slowly than the grammatical condition. The final model did not include an interaction term.

Table 4.21

Gender Model Results for L2 Learners' Total Time in DET-N Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	6.52	0.06	110.32	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.04	0.05	-0.83	= .417
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.19	0.05	3.94	< .001
	Random Effects				
	Subject	0.041			
	Item	0.000			
	Residual	0.291			

Note. Gender and grammaticality are both categorical variables with two levels. L2 learner reading times on grammatical sentences with a masculine target noun were taken as the reference category (Intercept).

General Summary of Statistical Results

This section presented the results L2 learners' reading times for grammatical and ungrammatical sentences depending on the gender of the target noun (masculine or feminine). In the DET-N condition, the L2 learners evidenced longer reading times in the ungrammatical condition relative to the grammatical condition for both go-past time and total time, but not first fixation duration or first-pass time. There was no statistically significant interaction for any of the models, meaning the sensitivity to the ungrammaticalities were not mediated by the gender of the target noun. For the N-ADJ and N-DROP condition (see Appendix F for analyses), no statistically significant main effect for gender or grammaticality was evidenced in any of these models, and there were no interactions between gender and grammaticality.

Sensitivity to Violations of Gender Agreement on an Individual Basis

In this section I explore whether the participants evidenced sensitivity to violations of grammatical gender agreement on an individual basis. I will examine this sensitivity in two

different ways. First, I will explore participants' sensitivity as evidenced by their responses on

the reading questionnaire administered immediately after the eye-tracking experiment. Then, I

will explore the L2 learners' sensitivity as evidenced by their individual reading times. Finally, I

will compare the L2 learners' individual reading times to their responses on the self-reports.

Participants' Sensitivity as Evidenced by Self-Reports

Immediately after finishing the eye-tracking experiment, participants were asked five

questions to determine whether they were aware that the sentences were embedded with

grammatical violations. These questions can be found in the section titled **reading**

questionnaire in Chapter 3.

The participants were first asked if they noticed anything strange about the sentences they

read during the eye-tracking experiment and if so, what. Of the 25 L2 learners, 15 reported errors

in the sentences they read. The others either did not report anything strange about the

experimental sentences, or commented solely on the comprehension questions. Of those 15 L2

learners, 10 stated that there were problems with grammatical gender agreement. For example,

one L2 learner noted, "…The grammar was incorrect at times in that the gender of the article

didn't match the gender of the noun" while another stated, "There were some grammatical errors.

For example something like 'el madre' instead of 'la madre.'" Of the 10 L2 learners that reported

seeing violations of grammatical gender agreement, seven reported violations of DET-N

agreement, while the other three did not specify the context of the agreement violations. No L2

learner specifically mentioned violations of N-ADJ or N-DROP agreement.

For question 1, all the native Spanish speakers reported that the sentences were indeed

strange, and 26 of them reported that it was because of ungrammaticalities. Of those 26

participants, 11 specified that there were problems with grammatical gender agreement. Of those

11 participants, five reported violations of DET-N agreement, one reported a violation of N-ADJ agreement, and five did not report a specific context.

Question 2 asked participants a more pointed question: whether there were any grammatical errors in the sentences they read during the eye-tracking experiment. All the native Spanish speakers reported that there were grammatical errors, while only 19 of the 25 L2 learners reported seeing errors. The six participants who stated that they did not notice any grammatical errors were not asked the next three questions and proceeded directly in the online survey to the vocabulary posttest.

The other 19 L2 Spanish learners and all the native speakers then navigated to question 3, where they were asked to name the types of grammatical errors they saw and to provide examples when possible. Fourteen of the L2 learners reported gender agreement errors. Of those 14, 12 reported DET-N agreement violations while 1 reported N-ADJ agreement violations. One participant mentioned agreement violations without specifying a context. No L2 learner reported N-DROP violations[24].

Of the 27 native Spanish speakers, 16 reported gender agreement errors. Of those 16, 11 reported DET-N agreement violations and two reported N-ADJ agreement violations. Three participants mentioned gender agreement violations without specifying one specific context. No native Spanish speaker reported N-DROP violations.

For question 4, I asked participants to check off from a list all the types of grammatical violations they noticed in the sentences. Some of the errors did appear in the experimental

---

[24] This may be because participants were not sensitive to these violations and/or because they did not have the metalinguistic vocabulary to describe these violations, as N-DROP is not usually taught explicitly in Spanish language classrooms.

stimuli while others did not. This was to determine if participants were simply marking all grammatical violations on the list.

Table 4.22 summarizes the participants' responses for question 4. The three types of agreement under investigation in the current study are highlighted in gray. The native Spanish speakers were roughly equally sensitive to all three types of grammatical gender agreement, while the L2 learners were most sensitive to DET-N agreement (16 participants), followed by N-ADJ agreement (13 participants), and then N-DROP agreement (11 participants).

Table 4.22
Participants' Reported Sensitivity to Violations in Sentence Processing Task

Grammatical violation	Violation appeared in stimuli?	# of native speakers reported violation	# of L2 learners reported violation
- ADVERBS appeared in the wrong place in the sentence	yes	21/27	3/19
- incorrect tense (present, past, etc.) was used	yes	21/27	12/19
- incorrect gender agreement between articles (e.g., el, la) and nouns	yes	25/27	16/19
- ADJECTIVES appeared in the wrong place in the sentence	yes	22/27	8/19
- incorrect agreement between subjects and verbs	yes	23/27	10/19
- subjunctive was used incorrectly	no	11/27	2/19
- por and para were used incorrectly	no	9/27	3/19
- ser and estar were used incorrectly	no	11/27	5/19
- incorrect gender agreement between nouns and adjectives	yes	25/27	13/19
- incorrect gender agreement between nouns and null nominals (e.g., el rico, la cómica)	yes	26/27	11/19

For the final question, I asked participants what percentage of the experimental sentences they thought contained grammatical errors. The L2 learners answered that an average of 29% of sentences (range: 3%-75%) contained grammatical errors and the native Spanish speakers answered that an average of 71% did (range: 18% to 99%). In reality, half of the sentences

contained violations of some kind. The difference between the saliency of grammatical violations is striking: native speakers tend to over-estimate the number of errors while L2 learners underestimate them.

L2 Learners' Sensitivity on an Individual Basis

To calculate individual participants' sensitivity to the grammatical gender agreement violations, I subtracted each participant's mean total reading time in the grammatical condition from the ungrammatical condition in each of the three syntactic contexts (DET-N, N-ADJ, N-DROP). I selected the total time measurement because it was the latest processing measure, and therefore the one most likely to evidence sensitivity, if there were any. Table 4.23 shows the change in reading time for the L2 Spanish learners.

Table 4.23

Individual Change in Reading Time for L2 Spanish Learners

Participant	DET-N	N-ADJ	N-DROP
A	-299	-12	-470
B	-265	113	132
C	-158	-213	-24
D	-84	136	98
E	-62	8	-26
F	-60	78	132
G	-10	56	-86
H	57	-9	-17
I	-46	-2	146
J	-26	174	-338
K	-15	143	-101
L	143	62	-25
M	176	-2	-194
N	262	16	13
O	290	20	-420
P	291	-83	-33
Q	502	106	-120
R	564	120	-217
S	174	-22	155
T	178	-155	323
U	495	12	171
V	537	-63	183
W	188	249	-156
X	236	226	564
Y	815	164	224
Mean	155	45	-3

Note. All times in are milliseconds. Any difference over 140 ms is highlighted in gray.

In Table 4.23, all differences over 140 ms are highlighted in gray to facilitate the interpretation of the table[25]. This threshold is, admittedly, arbitrarily defined, just as any threshold would be. I selected it by examining the data for the DET-N condition and finding a natural break in the data: participants in the DET-N condition either evidenced at least 140 ms of

---

[25] A similar procedure was used by Keating (2009) when comparing L2 learners' sensitivity across conditions. Keating, however, selected three different thresholds depending on the condition under investigation (125 ms for agreement in the DP, 117 ms in the VP, and 161 ms in the subordinate clause). In the interest of uniformity, I have instead selected to use a single threshold for all conditions.

sensitivity or none at all (range: -299 – 57). I then applied this threshold to the other two conditions. These data should be interpreted with caution, though, as altering the threshold can yield different interpretations of the same data.

Using 140 ms as a threshold, 14 out of a total of 25 L2 learners were sensitive to violations of DET-N agreement, five were sensitive to violations of N-ADJ agreement and seven were sensitive to N-DROP agreement. The greatest average reading time change was in the DET-N condition (155 ms), followed by N-ADJ agreement (45 ms). Even though more participants in the N-DROP condition were sensitive to the violations of agreement than in the N-ADJ condition when considering the 140 ms threshold, the participants in the N-DROP condition read the grammatical sentences on average at equal speed (-3 ms).

These data show a great deal of variability at the individual level: some participants did not show sensitivity to violations of grammatical gender agreement in any of the three conditions (8 participants), some were sensitive only to DET-N agreement (7), only to N-ADJ agreement (2), only to N-DROP agreement (1), to both DET-N and N-ADJ agreement (1), to both DET-N and N-DROP agreement (4), or to all three types of agreement (2).

Comparison of L2 Learners' Reported Sensitivity and Individual Reading Times

I triangulated the L2 learners' responses to questions 1, 2, 3, 4, and 5 on the reading questionnaire with their individual reading times to determine whether their reported sensitivity was reflected in their reading times. The results are in Table 4.24. Question 1 asked if participants noticed anything strange about the experimental sentences, question 2 asked if they saw any grammatical errors, question 3 asked participants to list any grammatical errors they saw in a free recall fashion, question 4 asked participants to check off the errors they saw from a list,

and question 5 asked participants to report the percentage of sentences they believe contained an

error.

Table 4.24

Comparison of Reported Sensitivity and Individual Change in Reading Time for L2 Spanish Learners

Participant	Q1	Q 3	DET-N Δ Reading Time	Q4	N-ADJ Δ Reading Time	Q4	N-DROP Δ Reading Time	Q4	Q5 (%)
A	/	/	-299	/	-12	/	-470	/	/
B	/	/	-265	/	113	/	132	/	/
C	ambiguous	DET-N	-158	yes	-213	yes	-24	yes	25
D	DET-N	DET-N	-84	yes	136	yes	98	yes	60
E	DET-N	DET-N	-62	yes	8	yes	-26	no	25
F	none	N-ADJ	-60	yes	78	yes	132	yes	/
G	/	/	-10	/	56	/	-86	/	/
H	/	/	57	/	-9	/	-17	/	/
I	none	other	-46	yes	-2	no	146	no	4
J	none	DET-N	-26	yes	174	yes	-338	yes	10
K	/	/	-15	/	143	/	-101	/	/
L	none	other	143	yes	62	no	-25	no	3
M	none	other	176	yes	-2	yes	-194	yes	40
N	ambiguous	DET-N	262	yes	16	no	13	yes	20
O	none	other	290	no	20	no	-420	no	10
P	/	/	291	/	-83	/	-33	/	/
Q	ambiguous	DET-N	502	yes	106	yes	-120	yes	75
R	DET-N	DET-N	564	yes	120	no	-217	no	25
S	none	ambiguous	174	yes	-22	yes	155	no	25
T	DET-N	DET-N	178	yes	-155	yes	323	yes	40
U	DET-N	DET-N	495	no	12	yes	171	yes	35
V	DET-N	DET-N	537	yes	-63	no	183	no	20
W	none	DET-N	188	yes	249	yes	-156	yes	40
X	none	other	236	no	226	yes	564	no	20
Y	DET-N	DET-N	815	yes	164	yes	224	yes	50
Mean			155		45		-3		29

Note. All times in are milliseconds. Any difference over 140 ms is highlighted in gray. A slash (/) indicates that participants indicated on Q 2 that they did not notice any grammatical violations in the experimental stimuli. For Q1 and Q 3, "none" indicates that the learners did not report seeing any agreement errors; "other" indicates that participants did not report seeing gender agreement errors, but rather a different error found in the distractor sentences; and "ambiguous" indicates that the participant reported seeing gender agreement errors, but did not specify the syntactic context.

In general, L2 learners that did not report any grammatical violations (Q2 – those indicated in the table with a slash) were not sensitive to the violations in their reading times; however, reported sensitivity to violations of grammatical gender agreement was not always manifested in reading times.

For question 1, of the 10 participants that reported seeing gender agreement errors, seven reached the 140 ms threshold of sensitivity in their reading times on at least one of the three conditions while three did not. Of the seven participants that specifically reported seeing DET-N errors, five evidenced DET-N sensitivity in their reading times while two did not. For question 3, of the 12 participants that reported seeing DET-N agreement violations, eight evidenced sensitivity in their DET-N reading times while four did not. One participant reported seeing N-ADJ violations for question 3, but did not reach the 140 ms threshold. The participant that mentioned gender errors but did not specify a context (i.e., 'ambiguous') was sensitive to DET-N and N-DROP agreement, but not N-ADJ agreement. For question 4, sensitivity in reading times coincided with the L2 learners' reported sensitivity 57% of the time, and when it did not, participants were more likely to over-report (72%) than under-report (28%) sensitivity. For question 5, the percentages that the participants reported must be interpreted with caution, as participants estimated the percentage of all the violations they saw during the experiment, both in terms of target items and distractors.

Summary of Results

The results of the reading questionnaire indicate the all the native speakers and most of the L2 learners reported seeing errors during the experiment. Of the three gender agreement violations under investigation, DET-N agreement errors were the most salient. In terms of individual reading times, there was a great deal of variability across participants. Reading times

indicated that the L2 learners were most sensitive to DET-N violations, but performed relatively similarly on N-ADJ and N-DROP items. Finally, the L2 learners' reading times coincided with their reported awareness on the reading questionnaire often, but not always. The results from the L2 learners' reading times and reported sensitivity coincided in that the DET-N violations were the most salient both in terms of reading times and reported sensitivity on the reading questionnaire.

CHAPTER 5: GENERAL DISCUSSION

Summary and Discussion of the Findings

The first two research questions that guided this dissertation were whether native Spanish speakers and L2 learners are sensitive to violations of DET-N, N-ADJ and N-DROP gender agreement during an online processing task and whether any evidenced sensitivity to the grammatical violations was contingent on the type of agreement under investigation. In the primary statistical analyses, the native Spanish speakers showed sensitivity to violations of grammatical gender agreement in all three syntactic contexts. The native speakers' sensitivity was evident in their first-pass, go-past and total reading times. L2 learners, on the other hand, were only sensitive to violations of DET-N agreement, and this sensitivity was evident in their first-pass, go-past and total reading times. It is unsurprising that sensitivity to violations of agreement was not evident during first-fixation because, these times generally reflect word identification processes (Clifton et al., 2007), and some previous research has also not found sensitivity to violations of grammatical gender agreement in first fixation times (e.g., Keating, 2009). Even though participants did not evidence sensitivity to violations of grammatical gender agreement during first fixation, they did evidence sensitivity during first-pass time, which, like first fixation, is considered an early-processing measure (Pickering et al., 2004). The results of this study seem to suggest that when sensitivity to violations did occur, it was robust and happened during both early and late stages of processing.

Taken together, these results provide preliminary evidence that the L2 learners do not have a representational deficit for grammatical gender agreement because, although they are not sensitive to the violations in all conditions, they are sensitive in at least one, according to the statistical analyses. Furthermore, an analysis of individual L2 participants' performance

indicated that at least 50% of the participants showed sensitivity to DET-N violations, suggesting some form of representation. The findings of this study therefore generally align with non-representational deficit approaches (e.g., Prévost & White, 2000; Schwartz & Sprouse, 1996) and suggest that learners can indeed acquire functional features that are not instantiated in their L1.

The additional statistical analyses examined whether the L2 learners exhibited evidence of a masculine default (e.g., McCarthy, 2008; White et al., 2004). While the L2 learners exhibited sensitivity to agreement violations in the DET-N condition for go-past time and total time, this sensitivity did not vary depending on the gender of the target noun. For the N-ADJ and N-DROP conditions, the L2 learners showed no sensitivity to agreement violations, regardless of the gender of the noun. Therefore, the L2 learners' reading times did not show any evidence of default morphology. Some researchers argue that L2 learners' underspecification errors during production, although non-nativelike, do not necessarily suggest a representational deficit, because the pattern of errors may be taken as evidence of a functioning system (e.g., White et al., 2004). For comprehension studies that examine native speakers' online processing of agreement, previous studies yield contradictory findings: while some have found that native speakers evidence an asymetrical representation for gender (e.g., Alemán Bañón & Rothman, 2016; Romanova & Gor, 2017), others have not (e.g., Acuña-Fariña, Meseguer, & Carreiras, 2014). Therefore, the L2 learners' lack of gender agreement asymmetry during online processing in this dissertation does not provide clear evidence for or against a non-representational account.

Even though the L2 learners showed some sensitivity to the violations in the DET-N condition, they did not perform like native speakers in the N-ADJ and N-DROP condition. I predicted that L2 learners would evidence the greatest sensitivity to violations of DET-N agreement followed by N-ADJ and then N-DROP. This prediction was only partially borne out

in the data. The statistical results indicated that participants were sensitive to violations of DET-N agreement, but were not sensitive to violations of N-ADJ and N-DROP agreement. When I analyzed L2 learners' responses individually, the participants showed the greatest sensitivity to violations in DET-N agreement both in terms of the number of participants deemed sensitive to the violations (14) and the average amount of sensitivity across participants (155 ms). However, like the statistical analyses, the individual data could not determine whether the N-ADJ or N-DROP condition showed higher rates of sensitivity: while the L2 learners had a greater average sensitivity for violations of N-ADJ agreement (45 ms), a greater number of participants read the ungrammatical areas of interest at least 140 ms more slowly in the N-DROP (seven in the N-DROP condition compared to five in the N-ADJ condition).

The L2 learners' sensitivity to DET-N agreement violations, but not N-ADJ violations is also paralleled in production studies, where learners are more accurate on DET-N agreement (Bruhn de Garavito & White, 2002; Franceschina, 2001; White et al., 2004). The L2 learners' lack of sensitivity to N-ADJ agreement could be due to a processing constraint: unlike the DET-N condition, the N-ADJ contains an AP which makes it longer, more complex, and therefore perhaps more difficult to process (Spinner & Juffs, 2008).

A processing constraint could also explain the lack of sensitivity to violations in the N-DROP condition. For example, in an eye-tracking study, Keating (2009) tested native Spanish speakers' and beginning, intermediate and advanced L2 Spanish learners' (L1 = English) sensitivity to grammatical gender violations on postnominal adjectives located in three syntactic domains: the DP, the VP and a subordinate clause, thereby manipulating both the structural and linear distance between the nouns and adjectives. Keating found that while the native speakers were sensitive to the violations in all three conditions, the beginner and intermediate learners

were not[26], and the advanced L2 learners showed sensitivity only to violations in the DP. Keating interprets the learners' sensitivity in the DP as evidence that they do not have a representational deficit for gender agreement, and ascribes their lack of sensitivity in the VP and subordinate clause as shallow processing (Clahsen & Felser, 2006a, 2006b). The deficit, he argues, is not one of competence, but rather of processing.

Another possible explanation for the L2 learners' lack of sensitivity in the N-DROP condition is that the participants had to compute a syntactic dependency to recover the referent of the null nominal across a linear distance; however, it is unclear whether in the ungrammatical condition the L2 learners (a) were not sensitive to the agreement violation and were therefore recovering the correct antecedent and incorrectly linking it to the pro in the null nominal, or (b) were linking a different antecedent to the pro in the null nominal that does agree in gender.

To illustrate this, consider (12) below. It is possible that the participants either (a) linked the masculine marked null nominal (el_{MASC} $frio_{MASC}$) to the correct antecedent (la_{FEM} $bebida_{FEM}$) because they are not sensitive to the ungrammaticality or, (b) linked the null nominal to an incorrect antecedent (e.g., su $amigo_{MASC}$) because they fail to comprehend the sentence. Example (a) would suggest problems computing agreement, while in example (b) that is not necessarily the case.

---

[26] The L2 learners in this dissertation were likely somewhat similar in proficiency to the intermediate L2 learners in Keating's study. Because Keating's intermediate learners were not sensitive to any violations, he posits that gender is acquired late (p. 525); however, the results of this dissertation suggest that gender is acquired earlier, but is only evidenced in DET-N agreement, not N-ADJ agreement.

(12)   *El hombre toma la bebida caliente y su amigo toma *el frío cuando van al restaurante.*

The man drinks the_{FEM} drink_{FEM} hot and his friend_{MASC} drinks *the_{MASC} small_{MASC} (one) when (they) go to the restaurant.

During the vocabulary posttest, the participants were asked to translate two Spanish sentences that contained null nominals into English to ensure that they could comprehend these types of sentences. While all the L2 learners translated the sentences correctly, the sentences they were asked to translate contained only correct agreement, which does not resolve what the L2 learners would do when the null nominal contained a violation. It is feasible that the L2 learners did not recover any referent at all, which would yield similar reading times on both grammatical and ungrammatical sentences[27]. Therefore, in this study, the difficulty for L2 learners may not be agreement, but rather pro indexing. Recovering referents can be quite difficult for both L1 and L2 language learners. For example, Shin & Cairns (2012) determined that Spanish-speaking children do not develop completely 'adult-like' preferences for overt pronouns in switch-reference contexts until age 14. There was no clear developmental trend for children to develop 'adult-like' preferences for null pronouns in same-reference contexts for any of the ages tested in the study (ages 6 to 15). These examples illustrate that for some structures, extensive input is required to pattern like adult native speakers. This seems to suggest that the L2 learners in this dissertation have simply not had enough input to be as sensitive to the violations as native speakers in all three syntactic categories.

---

[27] This may also explain the advanced learners' lack of sensitivity to N-ADJ agreement violations across clauses in Keating (2009), as those learners also had to compute a long-distance syntactic dependency.

The third research question that guided this dissertation was whether L2 learners reported seeing violations on a reading questionnaire administered after the eye-tracking experiment. On question 2 of the reading questionnaire, all native speakers and roughly 75% of the L2 learners reported seeing errors in the experimental sentences. The self-reports indicated that the most salient type of errors for the L2 learners was DET-N agreement violations, which also coincided with their reading times. Therefore, even though participants were not told ahead of time that the sentences contained ungrammaticalities, and their attention was directed to meaning through the inclusion of comprehension questions after every experimental sentence, they still developed explicit awareness of the violations.

For the L2 learners, reported sensitivity (or lack thereof) on the reading questionnaire (question 4) coincided with reading times only 57% of the time. The triangulation of participants' reading times with question 4 could results in four possible outcomes, either participants evidenced sensitivity to violations: (a) both on reading times and on the self-report, (b) on reading times but not on the self-report, (c) not on reading times, but on the self-report, or (d) not on reading times or the self-report. In this dissertation, roughly 23% of the L2 learners' responses on question 4 fell in category (a), 12% in (b), 31% in (c) and 35% in (d). Because the goal of studies that examine L2 learners' competence through sensitivity to grammatical violations during processing is to measure a linguistic system that is abstract and implicit, researchers often try to shift participants' attention away from explicit reasoning by having them focus on meaning[28] (e.g., Jiang, 2004, 2007; Jiang et al., 2011; but see Godfroid et al., 2015). This means that any scenario without awareness on the self-reports is ideal: either participants

---

[28] Jiang (2007) and Jiang et al. (2011) discuss integrated and nonintegrated knowledge, which are akin to implicit and explicit knowledge, respectively.

evidence sensitivity implicitly (b) or not at all (d). When participants report awareness, and are sensitive in reading times as in (a), the type of knowledge they are using while processing sentences is difficult to ascertain. Put differently, are the L2 learners sensitive to DET-N violations because they are using metalinguistic knowledge to detect them, or is their sensitivity first detected implicitly, and then rise to the level of metalinguistic awareness? For the native speakers, it is likely that implicit knowledge gives rise to a metalinguistic recognition of the errors in the sentences; however, whether this directionality is also true for L2 learners is an open question. One telling finding is that although the L2 learners evidenced sensitivity in the DET-N condition, they did not in the N-ADJ condition, even though both are explicitly taught in language classrooms. If metalinguistic knowledge were driving sensitivity, then they should show sensitivity on reading times in both conditions. Therefore, it seems likely that the L2 learners' sensitivity on reading times is primarily a reflection of their underlying competence, rather than metalinguistic knowledge.

That said, to ensure that L2 learners' explicit reasoning during processing stems from their underlying system and not the nature of the task they are engaged in, researchers should employ safeguards when creating experiments. Godfroid and Winke (2015) note that "whether eye movements signal implicit or explicit processes will depend to a large extent on the experimental design, target structure, and research questions of the studies involved" (p. 340). Some specific experimental design questions that L2 researchers should pay attention to are: (a) the ratio of distractor/filler sentences to experimental sentences, (b) the percentage of grammatical violations in the distractors, (c) the type of secondary distractor task, and (d) the online methodology employed (for reviews see Jegerski, 2014; Keating, 2014; Marinis, 2010; Roberts, 2012).

In this dissertation, the target sentences I tested represented 40% of all experimental sentences, including distractors/fillers. While similar research has employed roughly the same proportion of experimental sentences and distractors/fillers (e.g., Keating, 2009; VanPatten et al., 2012), more research is needed to determine how and if the percentage of distractors affects sensitivity to violations in experimental sentences. Half of the distractors in this study also contained ungrammaticalities, again coinciding with similar studies (Coughlin & Tremblay, 2013; VanPatten et al., 2012), but this may have also heightened participants' awareness to the violations in the experimental sentences.

Detailed research is also needed as to whether the type of data collection tool researchers implement affects how learners process sentences. Even though eye tracking may allow participants to read more naturally than other online comprehension measures, such as self-paced reading, the issue of reactivity still needs to be further explored (cf. Godfroid & Spino, 2015). One potential pitfall is that because sentences are presented in their entirety, and participants are usually not under a time pressure to read them (but see Godfroid et al., 2015), these eye-tracking studies could invite more explicit reasoning than other types of sentence processing measures, such as non-cumulative self-paced reading, because the sentences are not continuously projected on the screen for reanalysis (Jiang, 2004; Marinis, 2010). Therefore, one could argue that eye tracking is a more natural measure of reading (Dussias, 2010; Roberts, 2012; Witzel et al., 2012), but other online methodologies may be more appropriate for processing research (for discussion see Mitchell, 2004).

Implications of the Findings

The results of this dissertation indicate that L2 learners are indeed sensitive to gender agreement violations during online processing and therefore likely do not have a representational

deficit for gender agreement. The results also indicate that syntactic context can indeed affect sensitivity to grammatical gender agreement violations during online processing. Therefore, researchers should select with care the type of agreement under investigation when assessing L2 learners' competence, as certain types of agreement may be more difficult than others for learners. Learners should be biased for the best when researchers are testing for the ability to compute agreement, that is, the easiest type of agreement should be tested. If not, then it is impossible to determine whether a lack of sensitivity to violations (i.e., during processing tasks) or optionality (i.e., during production) is caused by a representational deficit or another intervening variable, such as syntactic context.

The results also suggest that L2 learners may rely on explicit knowledge when processing sentences that contain grammatical violations, even if their attention is directed to meaning. As a result, online tests that use this experimental design may be measuring explicit knowledge in addition to (or perhaps to the exclusion of) underlying competence. This was likely not a problem in this dissertation, because, even though participants reported high levels of explicit awareness of the violations, it did not make them sensitive in reading times to both syntactic contexts that are taught explicitly (DET-N and N-ADJ agreement). However, researchers using this experimental design to investigate representational deficits should carefully construct their experimental stimuli to limit this potentially confounding variable.

Limitations

All research is limited to the targets at hand, the population at hand, and the conditions at hand, and this dissertation is no exception. This dissertation would have benefitted from a larger sample size as well as a more advanced L2 learner group to tease apart the difficulty of N-ADJ and N-DROP agreement. It also would have benefitted from an independent test of working

memory, to determine if individual differences in working memory affected processing, especially in the N-DROP condition (cf. Keating, 2010; Sagarra & Herschensohn, 2010b).

Another limitation involves the target nouns selected for the eye-tracking study. I selected nouns with canonical –o and –a endings to bias learners with the best (for similar methodological decisions see Keating, 2009; Sagarra & Herschensohn, 2010b); however, this yielded near-ceiling performance on gender assignment in the vocabulary posttest, perhaps over-estimating their ability to assign gender to the target nouns employed in this dissertation. Implementing opaque nouns may have improved the quality of the experimental stimuli by eliminating this confound.

Future Research

This section outlines two future areas of research related to the current study. One future area of research would be to extend this study to investigate different types of grammatical gender agreement (e.g., direct object agreement). This could be done with eye tracking, as in the current study, or with another methodology such as self-paced reading or an oral production task. Studies such as this one would indicate which types of agreement are more and less difficult for L2 learners, which, in turn, can be used to reflect upon previous research investigating whether L2 learners can acquire grammatical gender agreement in their L2 if it is not instantiated in their L1 (e.g., McCarthy, 2008; Sagarra & Herschensohn, 2010a; White et al., 2004).

Another future potential area of research is methodological in nature. The extent to which online processing methodologies measure underlying competence instead of explicit knowledge likely depends on the design of the experiment and the construction of experimental stimuli. Future research should investigate directly the extent to which both the stimuli design and the online methodology chosen affect sensitivity to grammatical violations. These studies should

also triangulate eye-movement data with self-reported sensitivity, to determine what type of

knowledge participants are using to process the sentences.

APPENDICES

Appendix A

Pilot Vocabulary Test

1. Age:
2. Gender:
3. First Language(s)/Mother tongue(s):
4. How many semesters total have you taken Spanish in COLLEGE?
5. What is/are your major(s)? If you have two, please list both.
6. Are you a Spanish minor?
   Yes/No
7. At what age did you begin studying Spanish?
8. How many years did you take Spanish for in HIGH SCHOOL?
9. For how many years total have you actively studied Spanish?
10. Have you visited a Spanish-speaking country? If so, how long were you there for and what was the nature of your visit (e.g., vacation, study abroad, etc.)?
11. Have you studied any language(s) other than Spanish? If so, explain which language(s) you studied and for how long.
12. On the next page you will be asked to translate and rate your knowledge of 32 Spanish nouns. For each word on the next page please do the following:
    A.) Translate the word into English (but do not look up the translation!)
    B.) Rate your knowledge of the word on a scale of 4-0.
     4: I know this word very well; I translated it correctly and rapidly.
    3: I know this word somewhat well; I translated it correctly after some thought.
    2: I'm unsure of this word; I'm unsure if my translation is correct.
    1: I don't think I know this word; I don't think my translation is correct.
    0: I definitely don't know this word; My translation is definitely incorrect.

	English Translation	Word Knowledge				
		4	3	2	1	0
abrigo	___	( )	( )	( )	( )	( )
sombrero	___	( )	( )	( )	( )	( )
cuchillo	___	( )	( )	( )	( )	( )
trabajo	___	( )	( )	( )	( )	( )
bebida	___	( )	( )	( )	( )	( )
pregunta	___	( )	( )	( )	( )	( )
maleta	___	( )	( )	( )	( )	( )
comida	___	( )	( )	( )	( )	( )

Please translate and rate your knowledge.

	English Translation	Word Knowledge				
		4	3	2	1	0

zapato	___	( )	( )	( )	( )	( )
almuerzo	___	( )	( )	( )	( )	( )
ensayo	___	( )	( )	( )	( )	( )
cuaderno	___	( )	( )	( )	( )	( )
revista	___	( )	( )	( )	( )	( )
escuela	___	( )	( )	( )	( )	( )
mochila	___	( )	( )	( )	( )	( )
cerveza	___	( )	( )	( )	( )	( )

Please translate and rate your knowledge.

	English Translation	Word Knowledge				
		4	3	2	1	0
regalo	___	( )	( )	( )	( )	( )
dibujo	___	( )	( )	( )	( )	( )
espejo	___	( )	( )	( )	( )	( )
archivo	___	( )	( )	( )	( )	( )
cocina	___	( )	( )	( )	( )	( )
ventana	___	( )	( )	( )	( )	( )
manzana	___	( )	( )	( )	( )	( )
pintura	___	( )	( )	( )	( )	( )

Please translate and rate your knowledge.

	English Translation	Word Knowledge				
refresco		4	3	2	1	0
partido	___	( )	( )	( )	( )	( )
mercado	___	( )	( )	( )	( )	( )
vestido	___	( )	( )	( )	( )	( )
camisa	___	( )	( )	( )	( )	( )
piscina	___	( )	( )	( )	( )	( )
corbata	___	( )	( )	( )	( )	( )
iglesia	___	( )	( )	( )	( )	( )
refresco	___	( )	( )	( )	( )	( )

13. Please decide whether each noun below is masculine or feminine.

	Masculine	Feminine
abrigo	( )	( )
sombrero	( )	( )
cuchillo	( )	( )
trabajo	( )	( )
zapato	( )	( )
almuerzo	( )	( )
ensayo	( )	( )

cuaderno	( )	( )
regalo	( )	( )
dibujo	( )	( )
espejo	( )	( )
archivo	( )	( )
refresco	( )	( )
partido	( )	( )
mercado	( )	( )
vestido	( )	( )
bebida	( )	( )
pregunta	( )	( )
maleta	( )	( )
comida	( )	( )
revista	( )	( )
escuela	( )	( )
mochila	( )	( )
cerveza	( )	( )
cocina	( )	( )
ventana	( )	( )
manzana	( )	( )
pintura	( )	( )
camisa	( )	( )
piscina	( )	( )
corbata	( )	( )
iglesia	( )	( )

Appendix B

Experimental Stimuli

Determiner-Noun Condition

1. El niño se pone el/*la zapato durante la mañana antes de montarse en el autobús.

2. El muchacho compra el/*la sombrero cuando viaja a México con su mamá y su papá.

3. La madre prepara el/*la almuerzo durante la mañana para su hija y su esposo.

4. El traductor termina el/*la trabajo cuando llega a la oficina a las nueve de la mañana.

5. La mujer recuerda el/*la momento durante la reunión con su jefe y unos clientes.

6. El chico bebe el/*la refresco cuando ve la película en el cine con su familia.

7. La muchacha describe el/*la mercado cuando habla con su madre y su padre por teléfono.

8. La profesora revisa el/*la ensayo durante sus horas de oficina a las tres de la tarde.

9. La artista crea el/*la dibujo durante su clase de arte por la tarde con sus estudiantes.

10. El estudiante termina el/*la proyecto cuando regresa a su casa muy tarde por la noche.

11. El espectador bebe la/*el bebida durante el partido con sus amigos en el estadio.

12. El padre visita la/*el escuela durante la noche para la reunión con el maestro.

13. La muchacha compra la/*el camisa durante el fin de semana en el centro comercial.

14. La mujer visita la/*el iglesia cuando tiene dinero en el banco para donar.

15. El hombre usa la/*el guitarra cuando toca música con su banda en el bar.

16. La cocinera prepara la/*el comida durante la mañana en la cocina con sus asistentes.

17. La experta contesta la/*el pregunta durante la conferencia al final de su presentación.

18. El niño rompe la/*el ventana cuando juega con su amigo por la tarde en el jardín.

19. El abogado come la/*el manzana cuando habla con su cliente a las ocho de la mañana.

20. La madre necesita la/*el cámara cuando va de vacaciones con su esposo y su bebé.

Noun-Adjective Condition

21. El hombre busca el zapato rojo/*roja cuando se viste en la mañana para ir a trabajar.

22. La muchacha lleva el sombrero rosado/*rosada durante el fin de semana en la discoteca.

23. El estudiante come el almuerzo bueno/*buena cuando llega a la cafetería con sus amigos.

24. La mujer termina el trabajo corto/*corta durante la noche en la biblioteca de la universidad.

25. La niña recuerda el momento bello/*bella cuando habla con su madre en la playa.

26. El atleta toma el refresco frío/* fría cuando termina de correr tres millas por la mañana.

27. La cocinera busca el mercado caro/*cara durante el fin de semana en la ciudad.

28. El profesor lee el ensayo largo/*larga cuando tiene tiempo libre por la noche.

29. El artista describe el dibujo/*bonita bonito durante la exhibición del Museo Reina Sofía.

30. La muchacha empieza el proyecto nuevo/*nueva durante la mañana los jueves o viernes.

31. El hombre compra la bebida barata/*barato cuando sale con sus amigos en la noche.

32. La madre visita la escuela linda/*lindo durante la tarde para ver a sus hijos.

33. La chica plancha la camisa limpia/*limpio cuando se despierta por la mañana a las siete.

34. El niño dibuja la iglesia blanca/*blanco durante la clase en la hoja de papel.

35. La muchacha vende la guitarra negra/*negro durante el fin de semana a su compañero.

36. La niña come la comida mala/*malo durante el viaje con su madre y sus amigas.

37. La esposa hace la pregunta tonta/*tonto cuando habla con su esposo en la noche.

38. El padre limpia la ventana sucia/*sucio cuando trabaja afuera con su hijo en el jardín.

39. El hombre compra la manzana pequeña/*pequeño cuando va a la tienda por la tarde.

40. El muchacho vende la cámara vieja/*viejo durante la noche a alguien en Ebay.

Null Nominal Condition

41. El chico lleva el zapato marrón y su hermanito lleva el blanco/*la blanca cuando van al parque.

42. La madre se pone el sombrero horrible pero su hija se pone el bonito/*la bonita durante el partido.

43. La mujer pide el almuerzo simple pero su amiga pide el caro/*la cara cuando salen.

44. El estudiante escribe el trabajo breve y su compañero escribe el largo/*la larga durante la clase.

45. El hombre describe el momento ideal y su tío describe el malo/*la mala durante la conversación.

46. El jefe pide el refresco grande y su empleado pide el pequeño/* la pequeña cuando van a McDonald's.

47. La muchacha visita el mercado normal y su abuela visita el lindo/*la linda durante el día.

48. La estudiante lee el ensayo horrible y su amiga lee el bueno/*la buena cuando hacen la tarea.

49. El pintor ve el dibujo reciente y su estudiante ve el viejo/*la vieja cuando estudian.

50. El profesor describe el proyecto interesante y el chico describe el tonto/*la tonta durante la reunión.

51. El hombre toma la bebida caliente y su amigo toma la fría/*el frío cuando van al restaurante.

52. La niña prefiere la escuela horrible pero su mamá prefiere la limpia/*el limpio durante las visitas.

53. La madre lleva la camisa elegante y su hija lleva la sucia/*el sucio cuando salen.

54. La mujer visita la iglesia normal y su hermana visita la bella/*el bello durante la semana.

55. El músico toca la guitarra azul pero su amigo toca la rosada/*el rosado durante el concierto.

56. La mujer come la comida especial y su hermana come la barata/*el barato cuando cenan.

57. El estudiante hace la pregunta interminable y su amigo hace la corta/*el corto durante la clase.

58. El arquitecto quiere la ventana tradicional y el cliente quiere la nueva/*el nuevo durante la cita.

59. La niña come la manzana verde pero su hermana come la roja/*el rojo cuando desayunan.

60. El muchacho usa la cámara gris pero el fotógrafo usa la negra/*el negro cuando sacan fotos.

Appendix C

L2 Learner Background Questionnaire

1. What is your participant number?
1. Age:
2. Gender:
3. What is your native language?
4. Year in college:
   ( ) Freshman
   ( ) Sophomore
   ( ) Junior
   ( ) Senior (4th year)
   ( ) Senior (5th+ year)
   ( ) Other: _____
5. What is/are your major(s)? If you have two, please list both.
6. If you are a Spanish major, how many Spanish classes (NOT including the ones you're enrolled in now) do you still need to take to finish your Spanish Major?
7. Are you a Spanish minor?
8. When do you expect to graduate from MSU?
9. Please select all the courses you have previously taken or are currently taking at MSU:
   [ ] SPN 101 Elementary Spanish I
   [ ] SPN 102 Elementary Spanish II
   [ ] SPN 150 Review of Elementary Spanish
   [ ] SPN 201 Second Year Spanish I
   [ ] SPN 202 Second Year Spanish II
   [ ] SPN 310 Basic Spanish Grammar
   [ ] SPN 320 Cultural Readings and Composition
   [ ] SPN 330 Phonetics and Pronunciation
   [ ] SPN 342 Media and Conversation
   [ ] SPN 350 Introduction to Reading Hispanic Literature
   [ ] SPN 412 Topics in Hispanic Culture
   [ ] SPN 420 Spain and its Literature
   [ ] SPN 432 Latin America and its Literature
   [ ] SPN 440 The Structure of Spanish
   [ ] SPN 452 Topics in Spanish Language I
   [ ] SPN 462 Topics in Spanish Literature
   [ ] SPN 472 Topics in the Literatures of the Americas
   [ ] SPN 482 Topics in Spanish Linguistics
   [ ] SPN 490 Independent Study
   [ ] SPN 491 Special Topics in Spanish
   [ ] SPN 492 Senior Writing Project
   [ ] Other(s): _____

10. How many SEMESTERS of Spanish have you taken in COLLEGE (including this semester)?
11. For how many YEARS did you take Spanish in HIGH SCHOOL?
12. For many YEARS did you take Spanish in MIDDLE SCHOOL?
13. At what age did you take your first Spanish class? (Please only consider classes that met for more than an hour or two a week).
14. For how many years total have you actively studied Spanish?
15. How many hours per week do you currently spend...
    Speaking Spanish: _____
    Reading Spanish: _____
    Listening to Spanish: _____
    Writing in Spanish: _____
16. Have you ever studied abroad in a Spanish-speaking country?
17. Have you ever vacationed in a Spanish speaking country?

			If yes, WHERE did you travel?	WHEN did you travel?	FOR HOW LONG did you travel?
	Yes	No			
Response	( )	( )	_____	_____	_____

18. Have you ever completed an internship or service learning experience in a Spanish-speaking environment?
    [ ] No, neither.
    [ ] Yes, an internship.
    [ ] Yes, a service learning experience.
    [ ] If you selected yes, please briefly describe your experience.:
    _____
19. What languages besides Spanish have you studied? (Leave blank if you have not studied any other language).

	Language	For how many years did you study this language?
Language 1	_____	_____
Language 2	_____	_____
Language 3	_____	_____
Language 4	_____	_____

20. Where did you complete your study abroad?
21. What month and year did you LEAVE to travel abroad?
22. What month and year did you RETURN FROM traveling abroad?
23. How long did you study abroad for?
24. What best describes your living situation?
    [ ] I lived with a host family.
    [ ] I lived in a dorm with other Americans.
    [ ] I lived in a dorm with native-Spanish speakers.
    [ ] Other: _____
25. What percentage of the time did you use Spanish (as opposed to English or any other language) while living abroad (this includes speaking Spanish, watching Spanish TV, reading in Spanish, conversing in Spanish etc.)?
26. How did the study abroad experience affect your Spanish abilities?
    ( ) 7 My Spanish was MUCH BETTER when I returned to the US.

( ) 6
( ) 5
( ) 4 My Spanish was THE SAME when I returned to the US.
( ) 3
( ) 2
( ) 1 My Spanish was MUCH WORSE when I returned to the US.
27. Compare your Spanish abilities now to when you first arrived home from your study abroad experience. How would you compare your Spanish abilities now and then?*
( ) 7 My Spanish is MUCH BETTER NOW than when I returned to the US.
( ) 6
( ) 5
( ) 4 My Spanish is THE SAME NOW as when I returned to the US.
( ) 3
( ) 2
( ) 1 My Spanish is MUCH WORSE NOW than when I returned to the US.
28. Please list all the languages you know in order of dominance. Be sure to include your native language in the list:
29. Please list all the languages you know in order of acquisition (your native language first):
30. Please list the percentage of the time you are *currently* and *on average* exposed to each language.
(Your percentages should add up to 100%)
31. Do you wear glasses or contacts?
( ) Yes
( ) No
32. Are you currently having problems reading from books or computer screens that could negatively impact your performance on the tests you are taking today?
( ) Yes
( ) No
33. On a scale from zero to ten, please select your *level of proficiency* in speaking, understanding, and reading SPANISH from the pull-down menus:

Speaking	Understanding spoken language	Reading
( ) 0 - none	( ) 0 - none	( ) 0 - none
( ) 1 - very low	( ) 1 - very low	( ) 1 - very low
( ) 2 - low	( ) 2 - low	( ) 2 - low
( ) 3 - fair	( ) 3 - fair	( ) 3 - fair
( ) 4 - slightly less than adequate	( ) 4 - slightly less than adequate	( ) 4 - slightly less than adequate
( ) 5 - adequate	( ) 5 - adequate	( ) 5 - adequate
( ) 6 - slightly more than adequate	( ) 6 - slightly more than adequate	( ) 6 - slightly more than adequate
( ) 7 - good	( ) 7 - good	( ) 7 - good
( ) 8 - very good	( ) 8 - very good	( ) 8 - very good
( ) 9 - excellent	( ) 9 - excellent	( ) 9 - excellent
( ) 10 - perfect	( ) 10 - perfect	( ) 10 - perfect

34. On a scale from zero to ten, please select how much the following factors contributed to you learning SPANISH:

Interacting with friends	Interacting with family	Reading
( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor
Self instruction	Watching TV	Listening to the radio/music
( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor

35. Please rate to what extent you are currently exposed to SPANISH in the following contexts:*

Interacting with friends	Interacting with family	Reading
( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always	( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always	( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always
Self instruction	Watching TV	Listening to the radio/music
( ) 0 – never	( ) 0 – never	( ) 0 – never

( ) 1 – almost never	( ) 1 – almost never	( ) 1 – almost never
( ) 2	( ) 2	( ) 2
( ) 3	( ) 3	( ) 3
( ) 4	( ) 4	( ) 4
( ) 5 – half of the time	( ) 5 – half of the time	( ) 5 – half of the time
( ) 6	( ) 6	( ) 6
( ) 7	( ) 7	( ) 7
( ) 8	( ) 8	( ) 8
( ) 9	( ) 9	( ) 9
( ) 10 - always	( ) 10 - always	( ) 10 - always

36. In your perception, how much of a non-native accent do you have in SPANISH?
    ( ) 0 - none
    ( ) 1 - almost none
    ( ) 2 - very light
    ( ) 3 - light
    ( ) 4 - some
    ( ) 5 - moderate
    ( ) 6 - considerable
    ( ) 7 - heavy
    ( ) 8 - very heavy
    ( ) 9 - extremely heavy
    ( ) 10 - pervasive

37. Please rate how frequently others identify you as a non-native speaker based on your
    accent in SPANISH:
    ( ) 0 - never
    ( ) 1 - almost never
    ( ) 2
    ( ) 3
    ( ) 4
    ( ) 5 - half of the time
    ( ) 6
    ( ) 7
    ( ) 8
    ( ) 9
    ( ) 10 - always

Appendix D

Native Spanish Speaker Background Questionnaire

1. What is your participant number?
2. Age:
3. Gender:
4. What is your occupation? If you are a student, please type in your area of study (e.g., Spanish literature, mathematics, etc.).
5. Are you a Spanish TA, instructor or professor?
   Yes/No
6. If you answered yes to the previous question, for how many years have you taught Spanish?
7. What languages have you studied and for how long?

	Language	For how many years did you study this language?
Language 1	_____	_____
Language 2	_____	_____
Language 3	_____	_____
Language 4	_____	_____

8. Please list all the languages you know (including your native langauge) in order of dominance:
9. Please list all the languages you know in order of acquisition (your native language first):
10. Please list the percentage of the time you are *currently* and *on average* exposed to each language. (Your percentages should add up to 100%)
11. When choosing to read a text available in all of your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.
12. When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of the time would you choose to speak each language? Please report percent of total time.
13. Please check your highest education level (or the approximate US equivalent to a degree obtained in another country):
    ( ) Less than High School
    ( ) High School
    ( ) Professional Training
    ( ) Some College
    ( ) College
    ( ) Some Graduate School
    ( ) Master's
    ( ) Ph.D./M.D./J.D.
    ( ) Other: _____
14. In what country were you born?
15. At what age did you immigrate to the US?
16. Do you wear glasses or contacts?

Yes/No
17. Are you currently having problems reading from books or computer screens that could negatively impact your performance on the tests you are taking today?
Yes/No
18. Age when you...

	Age
Began acquiring ENGLISH:	___
Become fluent in ENGLISH:	
Began reading in ENGLISH:	___
Became fluent reading in ENGLISH:	___

19. Please list the number of years and months you have spent in each language learning environment:

	Years	Months
A country where ENGLISH is spoken:	___	___
A family where ENGLISH is spoken:	___	___
A school and/or working environment where ENGLISH is spoken:		___

20. On a scale from zero to ten, please select your *level of proficiency* in speaking, understanding, and reading ENGLISH from the pull-down menus:

Speaking	Understanding spoken language	Reading
( ) 0 - none ( ) 1 - very low ( ) 2 - low ( ) 3 - fair ( ) 4 - slightly less than adequate ( ) 5 - adequate ( ) 6 - slightly more than adequate ( ) 7 - good ( ) 8 - very good ( ) 9 - excellent ( ) 10 - perfect	( ) 0 - none ( ) 1 - very low ( ) 2 - low ( ) 3 - fair ( ) 4 - slightly less than adequate ( ) 5 - adequate ( ) 6 - slightly more than adequate ( ) 7 - good ( ) 8 - very good ( ) 9 - excellent ( ) 10 - perfect	( ) 0 - none ( ) 1 - very low ( ) 2 - low ( ) 3 - fair ( ) 4 - slightly less than adequate ( ) 5 - adequate ( ) 6 - slightly more than adequate ( ) 7 - good ( ) 8 - very good ( ) 9 - excellent ( ) 10 - perfect

21. On a scale from zero to ten, please select how much the following factors contributed to you learning ENGLISH:

Interacting with friends	Interacting with family	Reading
( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9

Self instruction	Watching TV	Listening to the radio/music
( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor	( ) 0 - not a contributor ( ) 1 - minimal contributor ( ) 2 ( ) 3 ( ) 4 ( ) 5 - moderate contributor ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - most important contributor

22. Please rate to what extent you are currently exposed to ENGLISH in the following contexts:

Interacting with friends	Interacting with family	Reading
( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always	( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always	( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always
Self instruction	Watching TV	Listening to the radio/music
( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always	( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always	( ) 0 – never ( ) 1 – almost never ( ) 2 ( ) 3 ( ) 4 ( ) 5 – half of the time ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 - always

23. In your perception, how much of a foreign accent do you have in ENGLISH?*
( ) 0 - none
( ) 1 - almost none

( ) 2 - very light
( ) 3 - light
( ) 4 - some
( ) 5 - moderate
( ) 6 - considerable
( ) 7 - heavy
( ) 8 - very heavy
( ) 9 - extremely heavy
( ) 10 - pervasive

24. Please rate how frequently others identify you as a non-native speaker based on your <u>accent</u> in ENGLISH:
( ) 0 - never
( ) 1 - almost never
( ) 2
( ) 3
( ) 4
( ) 5 - half of the time
( ) 6
( ) 7
( ) 8
( ) 9
( ) 10 - always

Appendix E

Vocabulary Posttest

1. What is your participant number?
2. Please mark whether each word is masculine or feminine, and then translate each word into English.
   If you are unsure about a translation, feel free to guess. If you have no idea what a particular translation is, write "I don't know".

	Gender		English Translation
	Masculine	Feminine	
zapato	( )	( )	___
sombrero	( )	( )	___
almuerzo	( )	( )	___
trabajo	( )	( )	___
momento	( )	( )	___
bebida	( )	( )	___
escuela	( )	( )	___
camisa	( )	( )	___
iglesia	( )	( )	___
guitarra	( )	( )	___
refresco	( )	( )	___
mercado	( )	( )	___
ensayo	( )	( )	___
dibujo	( )	( )	___
proyecto	( )	( )	___
comida	( )	( )	___
pregunta	( )	( )	___
ventana	( )	( )	___
manzana	( )	( )	___
cámara	( )	( )	___

3. Please translate the following two sentences into English to the best of your ability:

   El jefe pide el refresco grande y su empleado pide el pequeño cuando van a McDonald's.

   _____

   La niña come la manzana verde pero su hermana come la roja cuando desayunan.

   _____

Appendix F

Additional Gender Analyses

The L2 learners' model results for the gender analyses in the N-ADJ (Tables 6.1 to 6.4)

and N-DROP (Tables 6.5 to 6.8) conditions are presented below. For all analyses, gender and

grammaticality are both categorical variables with two levels. L2 learner reading times on

grammatical sentences with a masculine target noun were taken as the reference category

(Intercept).

Table 6.1
Gender Model Results for L2 Learners' First Fixation Duration in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	5.62	0.03	179.26	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	0.03	0.04	0.69	= .505
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.02	0.04	-0.62	= .536
	Random Effects				
	Subject	0.000			
	Item	0.001			
	Residual	0.142			

Table 6.2

Gender Model Results for L2 Learners' First-Pass Time in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	5.77	0.05	115.15	$< .001$
Overall main effect of reading a sentence with a feminine target noun	Gender	0.10	0.06	1.58	$= .130$
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.06	0.04	-1.72	$= .087$
	Random Effects				
	Subject	0.008			
	Item	0.012			
	Residual	0.155			

Table 6.3

Gender Model Results for L2 Learners' Go-Past Time in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	5.95	0.06	92.56	$< .001$
Overall main effect of reading a sentence with a feminine target noun	Gender	0.08	0.07	1.19	$= .250$
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.03	0.05	-0.66	$= .509$
	Random Effects				
	Subject	0.031			
	Item	0.014			
	Residual	0.236			

Table 6.4

Gender Model Results for L2 Learners' Total Time in N-ADJ Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	6.13	0.08	80.60	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	0.04	0.08	0.44	= .662
Overall main effect of reading an ungrammatical sentence	Grammaticality	0.03	0.04	0.66	= .512
	Random Effects				
	Subject	0.052			
	Item	0.022			
	Residual	0.228			

Table 6.5

Gender Model Results for L2 Learners' First Fixation Duration in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	5.53	0.03	176.48	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.03	0.03	-0.87	= .383
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.00	0.03	-0.05	= .961
	Random Effects				
	Subject	0.009			
	Item	0.000			
	Residual	0.100			

Table 6.6
Gender Model Results for L2 Learners' First-Pass Time in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	5.84	0.06	94.23	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.03	0.07	-0.48	= .635
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.01	0.04	-0.23	= .820
	Random Effects				
	Subject	0.028			
	Item	0.016			
	Residual	0.181			

Table 6.7
Gender Model Results for L2 Learners' Go-Past Time in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	t	p
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	6.01	0.08	73.70	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	-0.07	0.10	-0.76	= .458
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.01	0.05	-0.23	= .815
	Random Effects				
	Subject	0.037			
	Item	0.036			
	Residual	0.248			

Table 6.8

Gender Model Results for L2 Learners' Total Time in N-DROP Condition

Description	Predictor	Coefficient (β)	Standard Error	*t*	*p*
Overall effect of reading a grammatical sentence with a masculine target noun	Intercept	6.25	0.11	57.354	< .001
Overall main effect of reading a sentence with a feminine target noun	Gender	0.01	0.13	0.07	= .944
Overall main effect of reading an ungrammatical sentence	Grammaticality	-0.04	0.05	-0.92	= .358
	Random Effects				
	Subject	0.087			
	Item	0.068			
	Residual	0.256			

REFERENCES

REFERENCES

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua, 143*, 108-128. doi:10.1016/j.lingua.2014.01.013

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research, 30*(3), 275-306. doi:10.1177/0267658313515671

Alemán Bañón, J., & Rothman, J. (2016). The role of morphological markedness in the processing of number and gender agreement in Spanish: an event-related potential investigation. *Language, Cognition and Neuroscience, 31*(10), 1273-1298. doi:10.1080/23273798.2016.1218032

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137-153. doi:10.1162/0898929052880101

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

Bernstein, J. B. (1993). *Topics in the syntax of nominal structure across Romance.* (Doctoral Dissertation), City University of New York.

Bruhn de Garavito, J. (2003). The (dis)association between morphology and syntax: The case of L2 Spanish. In S. Montrul & F. Ordoñez (Eds.), *Linguistic theory and language development in Hispanic languages* (pp. 398-417). Somerville, MA: Cascadilla Press.

Bruhn de Garavito, J., & White, L. (2002). The second language acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. Muñoz Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153-178). Netherlands: Kluwer Academic Publishers.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53-77). Oxford: University Press.

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125-147). Oxford, England: Elsevier.

Carroll, S. (1989). Second language acquisition and the computational paradigm. *Language Learning, 39*(4), 535-594. doi:10.1111/j.1467-1770.1989.tb00902.x

Carstens, V. M. (2000). Concord in minimalist theory. *Linguistic Inquiry, 31*(2), 319-355. doi:10.1162/002438900554370

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Christensen, K. R., Kizach, J., & Nyvad, A. M. (2013). The processing of syntactic islands – An fMRI study. *Journal of Neurolinguistics, 26*(2), 239-251. doi:10.1016/j.jneuroling.2012.08.002

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*(1), 107-126. doi:10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3-42. doi:10.1017/S0142716406060024

Clahsen, H., & Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research, 5*(1), 1-29. doi:10.1177/026765838900500101

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on the mind and brain* (pp. 341-371). Oxford, UK: Elsevier.

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics, 34*(03), 615-646. doi:10.1017/s0142716411000890

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*(3), 369-382. doi:10.1177/0267658312443651

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159-181). New York, NY: Routledge.

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

Dewaele, J., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition, 4*(3), 275-297. doi:10.1017/S136672890100044X

Dowens, M. G., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870-1887. doi:10.1162/jocn.2009.21304

Dufour, R., & Kroll, J. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition, 23*(2), 166-180. doi:10.3758/BF03197219

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics, 30*, 149-166. doi:10.1017/s026719051000005x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172. doi:10.1017/S0272263105050096

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing.* (pp. 97-135). Amsterdam: John Benjamins.

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research, 17*(3), 213-247. doi:10.1177/026765830101700301

126

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: John Benjamins.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing of second language: A review of methodologies and experimental findings. *Second Language Research, 21*(2), 175-198. doi:10.1191/0267658305sr257oa

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition, 37*(2), 269-297. doi:10.1017/S0272263114000850

Godfroid, A., & Spino, L. (2015). Reconceptualizing reactivity of think-alouds and eye-tracking: Absence of evidence is not evidence of absence. *Language Learning, 65*(4), 896-928. doi:10.1111/lang.12136

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 325-348). Amsterdam: John Benjamins.

Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora, 10*(1), 95-125. doi:10.3366/cor.2015.0068

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191-215. doi:10.1177/0267658312437990

Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research, 13*(3), 187-226.

Hernández Pina, F. (1984). *Teorías psicolingüísticas y su aplicación a la adquisición del español como lengua materna*. Madrid: Siglo XXI de España Editores.

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

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*(4), 901-931. doi:10.1016/j.lingua.2009.06.004

Hopp, H. (2012). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*(1), 33-56. doi:10.1177/0267658312461803

Instituto Cervantes. (2008). Diploma de español. Nivel intermedio.   Retrieved from http://www.dele.org/uploads/test/intermedio/textos_orales.pdf

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. doi:10.1016/j.jml.2007.11.007

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research Methods in Second Language Psycholinguistics* (pp. 20-49). New York: Routledge.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603-634. doi:10.1017/S0142716404001298

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*(1), 1-33. doi:10.1111/j.1467-9922.2007.00397.x

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning, 61*(3), 940-967. doi:10.1111/j.1467-9922.2010.00627.x

Keating, G. D. (2005). *Processing gender agreement across phrases in Spanish: Eye movements during sentence comprehension.* (Doctoral Dissertation), The University of Illinois, Chicago, IL.

Keating, G. D. (2009). Sensitivity to violations of gender in native and nonnative Spanish: An eye-movement investigation. *Language Learning, 59*(3), 503-535. doi:10.1111/j.1467-9922.2009.00516.x

Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (Eds.), *Research in Second Language Processing and Parsing* (pp. 113-134). Amsterdam: John Benjamins.

Keating, G. D. (2014). Eye-tracking with text. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 69-92). London: Taylor & Francis.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition, 37*(1), 1-32. doi:10.1017/s0272263114000187

Kreiner, H., Garrod, S., & Sturt, P. (2013). Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes, 28*(6), 829-874. doi:10.1080/01690965.2012.667567

Leeser, M. J., Brandl, A., & Weissglass, C. (2011). Task effects in second language sentence processing research. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research*. Philadelphia: John Benjamins.

Liceras, J. M., Díaz, L., & Mongeon, C. (2000). N-drop and determiners in native and non-native Spanish: More on the role of morphology in the acquisition of syntactic knowledge. In R. P. Leow & C. Sanz (Eds.), *Spanish applied linguistics at the turn of the millenium* (pp. 67-96). Somerville, MA: Cascadilla Press.

Lim, J. H., & Christianson, K. (2014). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics, 36*(6), 1-33. doi:10.1017/s0142716414000290

López Ornat, S. (1988). On data sources on the acquisition of Spanish as a first language. *Journal of Child Language, 15*, 679-686. doi:10.1017/S0305000900012642

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967. doi:10.1044/1092-4388(2007/067)

Marinis, T. (2007a). On-line processing of passives in L1 and L2 Children. In A. Belikova, L. Meroni, & M. Umeda (Eds.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)* (pp. 265-276). Somerville, MA: Cascadilla Proceedings Project.

Marinis, T. (2007b). On-line processing of sentences involving reflexive and non-reflexive pronouns in L1 and L2 children. In A. Gavarro & M. J. Freitas (Eds.), *Proceedings of GALA* (pp. 348-358). Newcastle, UK: Cambridge Scholars Publishing.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139-162). Amsterdam: John Benjamins.

Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language, 36*, 143-171. doi:10.1017/S0305000908008908

McCarthy, C. (2007). *Morphological variability in second language Spanish.* (Doctoral Dissertation), McGill University, Montreal.

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research, 24*(4), 459-486. doi:10.1177/0267658308095737

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: eyetracking, ERPs and beyond* (pp. 15-32). New York: Psychology Press.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3), 503-553. doi:10.1111/j.1467-9922.2008.00449.x

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60*(1), 154-193. doi:10.1111/j.1467-9922.2009.00554.x

Picallo, M. C. (1991). Nominals and nominalizations in Catalan. *Probus, 3*(3), 279-316. doi:10.1515/prbs.1991.3.3.279

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eye-tracking, ERPs and beyond* (pp. 33-50). New York: Psychology Press.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition, 35*(4), 655-687. doi:10.1017/S0272263113000399

Prévost, P., & White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16*(2), 103-133. doi:10.1191/026765800677556046

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*(1-2), 103-121. doi:10.1016/j.specom.2004.02.004

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi:10.1037/0033-2909.124.3.372

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157. doi:10.1037/0033-295X.105.1.125

Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research, 28*(1), 113-127. doi:10.1177/0267658311418416

Roberts, L., & Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research, 29*(4), 413-439. doi:10.1177/0267658313503171

Romanova, N., & Gor, K. (2017). Processing of gender and number agreement in Russian as a second language. *Studies in Second Language Acquisition, 39*(1), 97-128. doi:10.1017/s0272263116000012

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*(1), 1-29. doi:10.1191/0267658306sr259oa

Sagarra, N., & Ellis, N. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition, 35*(2), 261-290. doi:10.1017/s0272263112000885

Sagarra, N., & Herschensohn, J. (2010a). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning, 61*(1), 80-116. doi:10.1111/j.1467-9922.2010.00588.x

Sagarra, N., & Herschensohn, J. (2010b). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8), 2022-2039. doi:10.1016/j.lingua.2010.02.004

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*(1), 40-72. doi:10.1177/026765839601200103

Shin, N. L., & Cairns, H. S. (2012). The development of NP selection in school-age children: Reference and Spanish subject pronouns. *Language Acquisition, 19*(1), 3-38. doi:10.1080/10489223.2012.633846

Snyder, W. (1995). *Language acquisition and language variation: The role of morphology.* (Doctoral Dissertation), MIT.

Snyder, W., Senghas, A., & Inman, K. (2001). Agreement morphology and the acquisition of noun-drop in Spanish. *Language Acquisition, 9*(2), 157-173. doi:10.1207/S15327817LA0902_02

Spinner, P., Gass, S. M., & Behney, J. (2013). Ecological validity in eye-tracking: An empirical study. *Studies in Second Language Acquisition, 35*(2), 389-415. doi:10.1017/s0272263112000927

Spinner, P., & Juffs, A. (2008). L2 grammatical gender in a complex morphological system: The case of German. *International Review of Applied Linguistics in Language Teaching, 46*(4). doi:10.1515/iral.2008.014

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research, 23*(2), 215-242. doi:10.1177/0267658307076546

van Hell, J. G., & de Groot, A. M. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193-211. doi:10.1017/S1366728998000352

VanPatten, B., & Benati, A. G. (2010). *Key Terms in Second Language Acquisition*. London: Continuum.

VanPatten, B., Keating, G. D., & Leeser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism, 2*(2), 109-140. doi:10.1075/lab.2.2.01pat

Vitu, F., O'Regan, K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics, 57*(3), 352-364. doi:10.3758/BF03213060

Wen, S., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. (2010). Proficiency effects and distance effects in nonnative processing of English number agreement. In K. Franich, K. Iserman, & L. Keil (Eds.), *Proceedings of the 34th Boston University Conference on Language Development* (pp. 445-456). Somerville, MA: Cascadilla Press.

White, L., Valenzuela, E., Kozlowska-Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25*, 105-133. doi:10.1017/S0142716404001067

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications.   Retrieved from http://arxiv.org/pdf/1308.5499.pdf

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research, 41*(2), 105-128. doi:10.1007/s10936-011-9179-x