

THE ONLINE USE OF MARKEDNESS INFORMATION IN
L1 AND L2 SPANISH GENDER AGREEMENT

By

Copyright 2015

Beatriz López Prego

Submitted to the graduate degree program in
Linguistics and the Graduate Faculty of the University
of Kansas in partial fulfillment of the requirements for
the degree of Doctor of Philosophy.

Co-chair: Alison Gabriele, Ph.D.

Co-chair: Robert Fiorentino, Ph.D.

Annie Tremblay, Ph.D.

Harold Torrence, Ph.D.

Amy Rossomondo, Ph.D.

Date Defended: August 18th 2015

The Thesis Committee for Beatriz López Prego
certifies that this is the approved version of the following dissertation:

THE ONLINE USE OF MARKEDNESS INFORMATION IN
L1 AND L2 SPANISH GENDER AGREEMENT

Co-chair: Alison Gabriele, Ph.D.

Co-chair: Robert Fiorentino, Ph.D.

Date approved: August 18th, 2015

ABSTRACT

The present study focuses on the acquisition and processing of gender agreement by second language (L2) learners of Spanish, whose first language (L1; English) lacks gender. Some L2 theories argue that these learners will not be able to acquire gender, and will have to resort to different strategies to process it in their second language (Hawkins, 2009), particularly in long-distance agreement dependencies (Clahsen & Felser, 2006; Clahsen, Felser, Neubauer, & Silva, 2010). Other theories argue that it is possible for those learners to acquire gender, but they may experience difficulty accessing target gendered forms, due to the computational burden of using a second language (Haznedar & Schwartz, 1997; Prévost & White, 2000). The current study addresses these theories by investigating how native speakers and advanced L2 learners use the gender markedness information (masculine vs. feminine) conveyed by the first element in a long-distance agreement dependency in particular, to process the second agreeing element in the dependency. In addition, it is investigated whether native speakers performing a task under processing burden show similar patterns to L2 learners in their processing of gender agreement (Hopp, 2010; McDonald, 2006; López Prego & Gabriele, 2014). This latter approach attempts to test whether specific error patterns in L2 learners emerge due to processing difficulty, or to a flawed representation of the gender feature. Thus, the study contributes unique data to answer the following questions: whether advanced L2 learners can establish long-distance agreement dependencies; whether they can develop a native-like representation of the gender feature in their L2, when they lack gender in their L1; and whether they can use gender information in a native-like manner in their online processing of agreement. These questions were tested in a self-paced reading task in

which a grammaticality judgment was provided after each sentence. The group of native speakers performing under processing burden was additionally asked to decide whether a string of numbers presented before each sentence was the same or different from a string presented after the grammaticality judgment was supplied.

The main results of the study showed that the advanced L2 learners tested, like the native speaker control group, were sensitive to gender agreement violations in long-distance agreement dependencies. In addition, both groups revealed a significant facilitation effect from the marked (feminine) feature in their processing of long-distance agreement dependencies, crucially, in grammatical sentences. Finally, the native speakers performing under processing burden showed some weak patterns that nevertheless resembled those in the L2 learner group. Thus, the findings from the present study support theories that posit computational difficulty as the source of agreement variability in L2 learners, and run counter to theories proposing a grammatical deficit in the L2 grammar as the cause of agreement errors in learners.

Table of Contents

1.	Introduction	1
2.	Distributed Morphology and Markedness Theory	8
3.	The use of featural information in L1 agreement.....	12
4.	L2 morphological variability	21
5.	The use of featural information in L2 agreement.....	36
6.	Present study	43
6.1.	Predictions	44
6.2.	Method.....	47
6.2.1.	Stimuli	47
6.2.2.	Stimuli controls.....	51
6.3.	Procedure and tasks	51
6.4.	Self-paced reading tasks.....	52
6.5.	Participants.....	55
6.5.1.	Native Speakers	55
6.5.2.	L2 Learners.....	55
6.6.	Analyses and results	56
6.6.1.	RQ1: Are L2 learners sensitive to gender agreement errors in long-distance agreement dependencies?.....	57
6.6.1.1.	Ungrammatical vs. Invariant.....	57
6.6.1.1.1.	Native Controls	59
6.6.1.1.1.1.	Summary.....	61
6.6.1.1.2.	L2 Learners.....	62
6.6.1.1.2.1.	Summary.....	63
6.6.1.2.	Ungrammatical vs. Gendered	64
6.6.1.2.1.	Native Controls	66
6.6.1.2.2.	L2 Learners.....	68
6.6.1.2.2.1.	Summary.....	69
6.6.2.	RQ2: Does the marked status of a feature facilitate processing in L2 learners and native speakers?.....	70
6.6.2.1.	Grammatical Gendered vs. Grammatical Invariant.....	70
6.6.2.1.1.	Native Controls	72
6.6.2.1.1.1.	Summary.....	74
6.6.2.1.2.	L2 Learners.....	75
6.6.2.1.2.1.	Summary.....	77
6.6.2.1.3.	Correlation analysis.....	78
6.6.3.	RQ3: Will native speakers performing under a processing burden reveal similar patterns to L2 learners?.....	79
6.6.3.1.	Ungrammatical vs. Invariant.....	80
6.6.3.1.1.	Summary.....	82
6.6.3.2.	Ungrammatical vs. Gendered	82
6.6.3.2.1.	Summary.....	84
6.6.3.3.	Grammatical Gendered vs. Grammatical Invariant.....	84
6.6.3.3.1.	Summary.....	86
6.6.4.	Acceptance Rates.....	86
6.6.4.1.	Summary	89
6.6.5.	Gender assignment task.....	90
7.	Discussion	90

7.1.	L2 sensitivity to gender agreement violations in long-distance dependencies	
	91	
7.2.	Use of gender information in online processing	95
7.3.	Native Speakers performing under processing burden	99
7.4.	Limitations of the study	101
8.	Conclusion	102

1. Introduction

Numerous studies have investigated the question of whether second language (L2) learners can fully acquire features absent from their native language (L1) (Franceschina, 2005; Hawkins & Casillas, 2008; Hawkins & Chan, 1997; Hopp, 2010; Keating, 2009; McCarthy, 2007, 2008; McDonald, 2006; Montrul, Foote, & Perpiñán, 2008; Prévost & White, 2000; White, Valenzuela, Kozłowska-Macgregor, & Leung, 2004, etc.), many of these studies focusing on the acquisition of gender in Romance languages, by English-speaking learners. Findings have been contradictory, with some studies showing native-like performance in learners (Alemán-Bañón, Fiorentino, & Gabriele, 2014; Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Gabriele, Fiorentino, & Alemán-Bañón, 2013; Hopp, 2010, 2013; López Prego & Gabriele, 2014; Prévost & White, 2000; White et al., 2004), and others showing advanced learners performing highly accurately, but falling short of target-like performance (Franceschina, 2005; Hawkins, 2001; Hawkins & Casillas, 2008; Hawkins & Chan, 1997; Keating, 2009; Tsimpli & Dimitrakopoulou, 2007). The inconsistency in these findings seems to at least in part be related to the different nature of the tasks employed in those studies. When L2 learners are tested using offline or untimed tasks, they tend to be on target, while their performance is more variable in online or timed tasks. This fact suggests that differences between L2 learners and native speakers may be more quantitative, rather than qualitative in nature. That is, the difference in performance may be related to the amount of processing resources available in the L2, compared to the L1, rather than differences in grammatical representations (Hopp, 2010; McDonald, 2006).

Processing in the L2 has generally been considered to be less efficient than in the

L1, and by some, also less dependent on certain kinds of linguistic information (Clahsen & Felser, 2006; Clahsen, Felser, Neubauer, & Silva, 2010). Thus, a key question that has emerged is whether L2 learners can show evidence of native-like online processing. This is particularly interesting for the specific topic of agreement dependencies, the focus of the present study. In agreement dependencies, the morphological form of a word (e.g. an adjective, a determiner, a pronoun, etc.) is dependent on the features (gender, and number in Spanish) of another word (e.g. a noun) within the phrase or sentence. For instance, in example (1) below, finding the pronoun *la* ('her') generates an expectation for a referent of feminine gender and singular number, since the features of the pronoun depend on the referent (*María*) it is replacing.

- (1) Cuando **la**_{FemSg} vi en el congreso, **María** estaba muy contenta.
When I saw **her** at the conference, **María** was very happy.

Thus, the question of whether and how both L1 and L2 speakers use the featural information of an element (*la*, in the example) to more efficiently process another agreeing element (*María*) is an interesting issue to investigate.

A number of L1 studies (Akhutina, Kurgansky, Polinsky, & Bates, 1999; Pearlmutter, 2000; Wicha, Moreno, & Kutas, 2003, 2004; Wicha, Orozco-Figueroa, Reyes, Hernandez, Gavaldón de Barreto, & Bates, 2005; Wagers, Stroud, McElree, & Phillips, 2009) have shown that native speakers experience a slowdown in processing when the expectation for a specific feature is not met. For example, in (2a) (from Wagers et al., 2009), finding a demonstrative (*those*) generates the expectation of a noun. Since the demonstrative is in its plural form, it is expected that the noun (*monkeys*) will also be in the plural form. When that expectation is not met, as in (2b) where the noun is in its

singular form, there is a slowdown in processing, relative to the matching condition in (2a).

(2) The girl told everyone that...

a. **those** mischievous face-making **monkeys** at the zoo...

b. **those** mischievous face-making ***monkey** at the zoo...

...was/were her friend(s).

These findings suggest that comprehenders actively anticipate the featural information encoded by lexical items that necessarily follow, and agree with previous elements in an utterance. Interestingly, some of these studies (Wagers et al., 2009) also find that native speakers may only use the information encoded by specific feature values, for example by plural number (e.g. these monkeys vs. these *monkey), but not by singular (e.g. that monkey vs. that *monkeys). This asymmetrical pattern has been linked to Markedness theory¹ (Battistella, 1990; Greenberg, 1966; Jakobson, 1975; Trubetzkoy, 1939). According to this theory, privative oppositions (e.g. singular vs. plural) have a hierarchical structure, with the more general, or *unmarked* element (singular) being the default, and indicating just the presence of the feature (number), and the most specific, or *marked* element (in this case plural) indicating a specific feature value (plural specification) (Battistella, 1990). Thus, it has been claimed that native speakers make use only of the marked feature value information in agreeing elements, during online processing of agreement (Wagers et al., 2009).

¹ See Battistella (1990) for a comprehensive discussion of Markedness theory.

²In the present study marked/unmarked are matched with specified/underspecified, following other authors

In addition, asymmetrical patterns like the one above have also been related to morphological theories like Distributed Morphology (DM; Halle & Marantz, 1993; Harley & Ritter, 2002). DM also posits a hierarchical, thus asymmetrical organization of features in the grammar, and it additionally proposes specific mechanisms to explain morphological paradigms across languages, and how agreement dependencies between features operate. Nevertheless, despite the fact that researchers have used these specific theories to explain asymmetrical feature patterns in native speakers, the experimental question of what specific role each feature plays in online processing remains rather unexplored (Akhutina et al., 1999; Alemán-Bañón, Rothman, & Miller, 2015; Wagers et al., 2009). The present study attempts to fill this gap by examining whether native speakers make different use of feminine, versus masculine gender in the processing of noun-adjective agreement dependencies.

In L2 research, following L1 research, some studies have also investigated whether L2 learners make use of featural information in the processing of agreement relations. Although some studies find evidence of this type of processing in L2 learners (Dussias et al., 2013; Hopp, 2013), others speak to the contrary (Grüter, Lew-Williams, & Fernald, 2012; Lew-Williams, & Fernald, 2010). Some researchers have proposed that adult L2 learners are in fact incapable of establishing agreement dependencies in a native-like manner, due to a deficit in the adult L2 grammar (Clahsen & Felser, 2006; Clahsen et al., 2010; Hawkins, 2009; Hawkins & Casillas, 2008; Hawkins & Chan, 1997; Tsimpli & Dimitrakopoulou, 2007). This hypothesis was advanced to try to account for the fact that even very proficient learners show variable performance establishing morphosyntactic dependencies. Thus, some L2 theories posit a critical period to acquire

features absent from the L1 (Hawkins & Casillas, 2008; Hawkins & Chan, 1997; Tsimpli & Dimitrakopoulou, 2007), or to acquire hierarchical, grammatical representations, specifically predicting adult L2 learners to fail at establishing long-distance agreement dependencies (Clahsen & Felser, 2006; Clahsen et al., 2010). The current study tests these theories by examining the specific question of whether L2 learners are sensitive to gender agreement violations as evidenced in reading time slowdowns. Crucially, the learners in this study started acquiring Spanish late in life, and were highly proficient at the time of testing.

Other studies on L2 morphosyntax have focused on the nature of the variability observed in agreement morphology in learners. Numerous L2 studies have reported the presence of asymmetries between feature values in L2 learner agreement errors, particularly in production (McCarthy, 2008), but crucially, also in comprehension (McCarthy, 2008; Hopp, 2013; López Prego & Gabriele, 2014). This phenomenon has been reported most frequently in English-speaking learners of Spanish (Bruhn de Garavito, 2003; Bruhn de Garavito & White, 2002; McCarthy, 2007, 2008, 2012; Montrul et al., 2008; White et al., 2004) and French (Prévost & White, 2000; Renaud, 2010, 2011, 2012). For example, a frequently observed error of a learner of Spanish, as illustrated in (3) below, is to provide the (mismatching) masculine form of an adjective in the context of a feminine noun.

(3) La casa es *viejo.
the_{FemSg} house_{FemSg} is old_{MSg}

The opposite pattern, illustrated in (4), providing the (mismatching) feminine form of an adjective in the context of a masculine noun, is not as frequently attested.

(4) El libro es *vieja .
the_{MSg} book_{MSg} is old_{FemSg}

According to a specific proposal by Hawkins (2009), these errors result from a deficit in the adult L2 grammar. In order to make up for this deficit, learners establish agreement dependencies by tracking co-occurrence frequencies in the input, and memorizing exceptions (Hawkins, 2009, p. 72). Specifically, L2 learners are claimed to assign the more frequent masculine gender (in Spanish and Romance languages) to nouns by default, and memorize feminine nouns together with their agreeing elements as they are encountered, as exceptions. A prediction that follows from Hawkins's proposal is that frequency differences among the memorized *feminine* forms should result in differences in the speed of retrieval of these forms in L2 learners (see Hopp, 2013). Since native speakers do not have a deficient grammar, they do not need to resort to the same strategies as L2 learners. Thus, these frequency differences in the input should not impact native speakers. This specific proposal is tested in the current study by examining whether frequency differences in feminine items affect L2 learners and native speakers in their establishment of gender agreement dependencies.

In contrast to Hawkins, other researchers argue that the asymmetrical agreement patterns observed in L2 learners result from the computational difficulty associated with using the L2 (Hopp, 2010, 2013; López Prego & Gabriele, 2014; Prévost & White, 2000; White et al., 2004). These researchers, following an original proposal by Prévost & White (2000), have drawn a connection between the L2 patterns, and the native speaker theories outlined above that posit similar asymmetries in native grammatical representations (Halle & Marantz, 1993; Harley & Ritter, 2002). Specifically, Prévost & White (2000)

proposed that the asymmetrical patterns observed in L2 learners may actually reflect a native-like representation of features, as that hypothesized in Distributed Morphology (Halle & Marantz, 1993; Harley & Ritter, 2002). The feature asymmetries would occasionally surface as errors in L2 learners due to processing difficulty accessing target forms. If L2 learners do, in fact, have a native-like, asymmetrical representation of gender, the question that arises is whether learners will use gender information similarly to native speakers, in their processing of agreement. The current study addresses this question by examining, as for native speakers, whether learners make different use of feminine, versus masculine gender in their online processing of agreement, in grammatical sentences.

Since these latter theories argue that agreement errors in L2 learners emerge due to the increased demands of processing the L2, some studies have tested this hypothesis by exploring whether native speakers performing a task under increased processing burden show similar patterns to L2 learners (Hopp, 2010, 2013; López Prego & Gabriele, 2014; McDonald, 2006). Interestingly, the results from these studies supported the hypothesis tested, suggesting that L2 learners and native speakers may share similar representations of features that surface as errors during burdensome processing. Following these previous studies, the current study seeks to tease apart the computational and representational theories that try to explain L2 agreement variability by testing a group of native speakers performing the same task as L2 learners, and a native speaker control group under an added computational burden.

The current dissertation is structured as follows: first, the main L1 morphological theories proposing a hierarchical representation of features will be explained. Next, a

number of L1 studies on the processing of agreement will be reviewed. Then, the relevant literature and main theories on L2 agreement will be summarized. Subsequently, some L2 studies investigating the online processing of agreement in learners will be presented. Finally, the present study will be explained in detail, followed by the results obtained, and a discussion of the findings in light of the theories of native and non-native processing.

2. Distributed Morphology and Markedness Theory

According to Distributed Morphology (DM; Halle & Marantz, 1993), the Morphology component of the grammar (Morphological Operations), similar to the Syntax component, generates structure; morphological features are organized in hierarchies, rather than in unstructured bundles. The morphosyntactic features produced by the syntax are subsequently subject to morphological operations that generate the featural content of vocabulary items (Harley & Noyer, 1999). This featural content may be ‘blank’, resulting in *underspecified* or *unmarked* morphemes —also called *elsewhere* or *default* forms—, or not, giving place to *specified*, or *marked* morphemes² (Harley & Noyer, 1999; Harley & Ritter, 2002). The notion of *underspecification* implies that all redundant information can be dispensed with from abstract representations (Bobaljik, 2002). In the case of contrastive feature values like masculine versus feminine, or singular versus plural, it is not required that both features are represented; this would constitute redundant information. Instead, the marked morpheme is the one that is specified for a given feature (e.g. feminine, for feminine gender), while the underspecified or unmarked morpheme is

²In the present study marked/unmarked are matched with specified/underspecified, following other authors (Cowper, 2005; Harley, 1994; Harley & Ritter, 2002; McCarthy, 2007, 2008).

void of specific featural content, and only indicates the presence of the feature (e.g. masculine only indicates the presence of gender).

With regards to the categorization of feature values as marked or unmarked, Battistella (1990) proposed a series of distributional and syntactic criteria as a systematic way to determine markedness. Below are some of the relevant criteria for Spanish gender and number that suggest that masculine and singular are the unmarked values for each feature:

(a) Syntactic distribution: unmarked items have a wider-ranging distribution, occurring in more syntactic contexts, compared to marked items. In Spanish, masculine, rather than feminine, is the gender usually assigned to borrowed inanimate nouns, for example *el*_{MSg} *email*.

(b) Indeterminateness of meaning: unmarked items have a more general meaning, while marked items have a more specific meaning. In Spanish the singular form of nouns can be used with a plural meaning, but not vice versa; *la margarita es bonita*, ‘the daisy is beautiful’, can either mean that one daisy is beautiful, or that daisies in general are beautiful. This is also an example of syntactic distribution, since the singular is used in a plural context; the use of the plural in a singular context is not possible.

(c) Neutralization: unmarked items lose some meaning specification in certain contexts. For example, in Spanish the word *padre*, ‘father’, loses the masculine meaning in a plural context. That is, the word *padres* (lit. ‘fathers’) refers to the father and the mother together (i.e. ‘parents’). This indicates again that masculine is the unmarked form for gender.

Specifically with respect to gender, some data coming from studies investigating gender agreement in Spanish-speaking children seem to support Battistella's (1990) classification of the masculine and feminine feature values (Anderson, 1999; Brisk, 1976; Mariscal, 2009; Pérez-Pereira, 1991). Pérez-Pereira (1991), for example, found that Spanish-speaking children tend to more frequently assign masculine than feminine gender to unknown nouns, a finding that has also been reported for adult native speakers (Natalicio, 1983). Similarly, Brisk (1976) and Mariscal (2009) find that children's gender agreement errors more frequently constitute misuses of masculine agreement with feminine nouns, than feminine agreement with masculine nouns (although these authors do not attribute the results to markedness). Interestingly, this pattern of errors is also found in a study by Anderson (1999), which reports data on two English-Spanish bilingual children undergoing attrition in Spanish.

Under markedness theory, then, since masculine and singular are the unmarked, default values for gender and number (respectively), they signal gender, and number, (respectively) broadly: masculine only indicates the presence of gender, and singular only indicates the presence of number. In contrast, feminine indicates specifically feminine gender, and plural indicates specifically plural number³. The representation of features as specified or underspecified, rather than as binary (\pm), results in an asymmetrical representation of features, key to the DM approach. DM shares this view of features being organized in hierarchies, with markedness theory (Battistella, 1990; Greenberg, 1966; Jakobson, 1975; Trubetzkoy, 1939). Markedness theory imposes a hierarchical structure onto oppositions (e.g. singular vs. plural) that otherwise would be considered

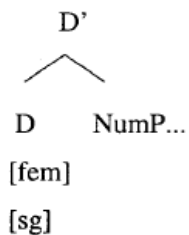
³ In the case of number, the status of singular as unmarked and plural as marked seems to be rather universal in languages with a two-way number distinction (Battistella, 1990).

equivalent (Battistella, 1990). However, DM goes beyond markedness theory in that it constitutes a complete morphological theory, with specific proposals on linguistic mechanisms such as the establishment of agreement dependencies.

The mechanism that DM proposes for agreement is the following: In order to insert the correct morpheme in the syntax, its features need to be checked against those specified in the syntactic context. According to the *Elsewhere Principle* or *Blocking Principle*, the item with the greatest number of matching features should be the one inserted, thus preventing the insertion of an underspecified morpheme when a more specified one is needed, and is available. When there is no perfect match of features with any of the available items, the elsewhere or default form is inserted. This means that the vocabulary item inserted can be either a *perfect* match, or contain a *subset* of the relevant features, but it cannot represent a *mismatch* or clash of features. Under this approach, the insertion of underspecified or default forms is not ruled out. Since these forms are void of featural content, their insertion doesn't result in a clash with the features specified in the syntax, but rather in the insertion of a subset of those.

(5)

a) *syntactic context*



b) *competing morphemes*

la ↔ [fem]

el ↔ *elsewhere*

outcome

wins competition

blocked

(adapted from McCarthy, 2007:19)

To illustrate how the *Elsewhere Principle* works, in example (5a) with the Spanish determiners, the syntactic context is stipulated as feminine, and singular. The competing morpheme *la* in (5b) is underspecified for number (unmarked), while for gender it is specified as feminine (marked). In contrast, the competing morpheme *el* is underspecified for both gender and number (unmarked). Since *la* contains a subset of the features specified in the syntax [fem], and *el* doesn't, *la* is the morpheme with the greatest number of features matching the syntax, and therefore, it should be the vocabulary item inserted in the syntactic context. In this case the *Elsewhere Principle* prevented the insertion of an underspecified item when a more specified one matching the syntax was available. The *Elsewhere Principle* also inhibits the insertion of clashing features. In the same example (5) *las*, specified as feminine and plural, would be another competing morpheme. Nevertheless, if this item were inserted in the syntax, its plural specification would clash with the singular feature denoted in the syntactic context. In this case, the singular default form *la* is the appropriate vocabulary item to be inserted in the syntactic context.

3. The use of featural information in L1 agreement

While many L1 studies have investigated the processing of gender or number features in agreement, few of those studies have focused on how native speakers use the information encoded by specific feature values in their online processing (Akhutina, et al., 1999; Alemán-Bañón, et al., 2015; Wagers et al., 2009). Rather, most of the data informing our knowledge of this issue comes in the form of findings secondary to the purpose of those studies.

For example, several studies investigating number attraction effects in subject-verb agreement report finding differences in the behavior of participants with singular

versus plural number (Antón-Méndez, Nicol, & Garrett, 2002; Bock & Miller, 1991; Bock & Cutting, 1992; Bock & Eberhard, 1993; Eberhard, Cutting, & Bock, 2005; Eberhard, 1997; Nicol, Forster, & Veres, 1997; Pearlmutter, 2000; Vigliocco & Frank, 1999). Number attraction effects refer to number agreement errors on a verb caused by interference from a plural local NP that mismatches the (singular) number of the head of the subject (e.g. “**the key** to the cabinets *were...”). These studies converge in finding that the attraction phenomenon is not so frequently attested when the number mismatch involves a singular local NP mismatching a plural head (e.g. “**the keys** to the cabinet *was...”). Thus, these studies have concluded that singular subjects are more susceptible to interference from modifying plural NPs, in comparison to plural subjects with modifying singular NPs. Eberhard (1997), relying on markedness theory, proposes that it is the fact that the marked plural form of the head noun signals the presence of the feature specification, that makes it less prone to attraction errors. Since a feature specification is present, it is easier to track during agreement. In contrast, unmarked features (e.g. singular), as explained above, are underspecified. Thus, when the head noun is in its singular form, there is no specification to track for agreement, which makes it more susceptible to the interference of a plural local noun. Eberhard’s proposal contemplates the possibility that the asymmetries observed are the result of frequency differences between singular and plural. However, Eberhard argues that asymmetries are better explained by featural differences, given, for example, that in attraction phenomena, more errors result from processing the more frequent, unmarked singular form of the subject head noun (e.g. “**the key** to the cabinets *were...”), than the less frequent, marked plural form (e.g. “**the keys** to the cabinet *was...”). These results leave open the question of

whether this pattern results from the unmarked status of the head noun, or from the marked status of the local attractor noun. Interestingly, a self-paced reading study by Pearlmutter (2000) finds that hierarchical structure interacts with markedness in attraction effects, and also suggests that it is mainly the markedness status of the head noun that results in the particular attraction pattern. Specifically, Pearlmutter finds that with singular subjects (6), interference at the verb phrase can be caused by either higher (6a), or lower (6b) modifying nouns in the tree.

(6)

- a. The **lamp** near the **paintings** of the house was damaged in the flood.
- b. The **lamp** near the painting of the **houses** was damaged in the flood.

However, with plural subjects (7), only the higher modifying noun (7a) interferes. These results suggest that the markedness status of the head noun (as opposed to the local noun) determines whether a local attractor noun will interfere in the dependency or not.

(7)

- a. The **lamps** near the **painting** of the house were damaged in the flood.
- b. The **lamps** near the paintings of the **house** were damaged in the flood.

Pearlmutter explains these results by proposing that the marked status of the subject in (7) as plural makes it more resistant to interference, particularly if the potentially disrupting modifying noun is lower in the tree. Thus, he suggests that the strength of the number marking of the subject head is influenced by the (hierarchical) complexity of the subject NP.

While there is plenty of evidence within attraction studies pointing to a clear difference between the impact of plural versus singular number in processing, not so much evidence has been found within these studies for an asymmetry between gender feature values. Whereas some studies do find such an asymmetry (Vigliocco & Franck, 1999, in French; Vigliocco & Franck, 2001, in Italian; Antón-Méndez et al., 2002, in Spanish), it is in the opposite direction to that predicted by Markedness theory, that is, more attraction errors when the head noun is marked feminine, and the local NP is masculine, than the reverse pattern. In addition, these effects do not seem to be strong, since they are not found in other attraction studies investigating the same languages (e.g. Acuña-Fariña, Meseguer, & Carreiras, 2014, in Spanish; Vigliocco & Franck, 1999, in Italian). It is important to consider, however, that the fact that gender asymmetries don't emerge in the specific context of sentences testing attraction, does not rule out an asymmetric representation of gender. Rather, it would be evidence that the asymmetry doesn't emerge in the specific context of attraction.

Within the realm of ERPs (event-related potentials), not many studies have examined markedness differences between feature values, or some aspect related to feature value asymmetries. For example, Kaan (2002) conducted an ERP study in order to investigate the effects of linear distance, and number interference in subject-verb agreement in Dutch. One of her manipulations involved the grammaticality of the number agreement dependency between the subject and the verb, along with the number of an intervening object, as shown in (8).

(8)

a. Omdat volgens het reglement de trainer de atleten had moeten inschrijven mochten ze niet starten.

because according-to the regulations the coach the athletes had_{sg} need sign-up allowed they not start

“Because the coach ought to have signed up the athletes as was required by the regulations, they were not allowed to start.”

b. Toen na de excursie de toeristen de reisleidster wilden gaan trakteren protesteerde de chauffeur.

when after the tour the tourists the guide(fem) wanted_{pl} go treat protested the driver

“When the tourists wanted to treat the female guide [to, e.g., a drink] after the tour, the driver protested.”

Her findings were interesting in that there were more judgment errors in sentences where the subject was singular and the intervening object was plural (8a), consistent with the results from attraction studies, but the ERP patterns obtained were somewhat unexpected. Specifically, there was a larger P600 (the usual brain response to grammatical violations) in the conditions where the subject and intervening object were both singular, and the verb was plural, as in (9):

(9)

Hoewel volgens het gerucht de keizer de dissident *zullen gaan verbannen is er veel tegestand.

although according to the rumor the emperor the dissident will_{pl} go ban is there a lot of opposition

“Although the emperor will ban the dissident according to the rumor, there is a lot of opposition”

It is possible, as Kaan proposes, that these errors are more salient, or harder to repair, although the reason why is not apparent. One possibility is that having two elements with the same number (the singular subject, and object), reinforced the singular feature, such that finding the marked feature on the plural verb made the error more salient than others.

Nevertheless, the issue of how attraction, markedness, and grammaticality may interact, and impact ERP responses remains a question for future research.

Another ERP study finding markedness-related effects is Deutsch and Bentin (2001), this time in subject-verb gender agreement in Hebrew. Although the purpose of this study was to investigate the impact of syntactic and semantic factors on the processing of gender agreement, the authors report differences in brain responses related to number markedness effects. Specifically, a larger P600 was observed when the gender agreement violations occurred in plural contexts, compared to singular contexts, an effect that the authors attribute to plural marking in Hebrew being salient.

Finally, one ERP study that directly addresses the issue of markedness in native speakers is Alemán-Bañón et al. (2015). In this study, the authors examined gender and number agreement violations in Spanish native speakers, by manipulating the markedness (in gender and number) of the nouns and adjectives involved in the violations. The sentences in (10) sample the gender violations:

(10)

- a. Andrés alquiló un **coche** que parecía ***barata** durante la excursión.
Andres rented a **car_M** that looked **cheap_{Fem}** during the excursion.
- b. Carlos fotografió una **catedral** que parecía ***inmenso** para una revista.
Carlos photographed a **cathedral_{Fem}** that looked **huge_M** for a magazine.

The participants read the sentences while their brain responses were recorded, and at the end of each sentence, they provided a grammaticality judgment. The results revealed that, while markedness did not impact the P600 response in their study, it did modulate another response, the N400. The N400 has been related to cases where a prediction for finding a specific form is not met (Guajardo & Wicha, 2014). The specific pattern found

in the study was that the N400 emerged for gender and number violations where the first element (the noun) was marked (feminine or plural), and the subsequent element (the adjective) was unmarked (masculine or singular; 10b in the case of gender). The N400 did not emerge in the opposite contexts, where the noun was unmarked, and the adjective, marked. To explain these effects, the authors argued that the information provided by the marked feature on the noun allows to make a prediction for an upcoming feature on an agreeing element. When that prediction is not met, an N400 effect emerges. In the context where the noun is in the unmarked form the prediction is not generated, since the unmarked feature does not provide feature information, thus explaining the absence of the N400 response.

This type of asymmetry where specific effects emerge only for the marked feature has also been found in some earlier behavioral studies. For example, in a cued shadowing study on gender agreement in Russian, Akhutina et al. (1999) found that the marked feminine information encoded on adjectives had a facilitative effect on the processing of the gender encoded by a subsequent noun modified by that adjective. Specifically, shadowing (repeating after hearing) the noun was faster after a feminine adjective, compared to nouns following masculine and neuter adjectives.

Finally, a key study by Wagers et al. (2009) investigating differences in the processing of singular and plural features in English, makes a specific proposal for the processing of marked, versus unmarked features, relying on Markedness theory and Distributed Morphology. Specifically, the study investigates how singular and plural features are maintained in focal attention, when an expectation for agreement with those features is generated. The focus of attention is defined as a holding device within

working memory, in which several items can be maintained at the same time (Gilchrist & Cowan, 2011). Those items are ‘privileged’ with respect to other items in working memory; they are highly accessible to cognitive processes, they are in awareness, and they are protected from decay and interference from other items (Gilchrist & Cowan, 2011). Wagers et al. (2009) tested the hypothesis that when comprehenders expect a specific feature in an utterance, they maintain it in the focus of attention. For example, after encountering a determiner with a plural feature, comprehenders will expect to find the same feature marked on an upcoming noun, and thus maintain the plural specification in the focus of attention. In addition, based on the postulations from markedness, Wagers et al. (2009) predicted an interaction between maintenance in the focus of attention, and the markedness status of number features. That is, they hypothesized that the ‘survival’ of number features in the focus of attention is impacted by whether the feature is marked (plural) or unmarked (singular). For example, because the marked plural feature in (11d) below indicates the presence of plural number, it is expected to survive longer in the focus of attention relative to the unmarked singular in (11b), which only indicates the presence of the number feature.

(11) The girl told everyone that...

- a. **that** mischievous face-making ***monkeys** at the zoo
- b. **that** ***monkeys** at the zoo
- c. **those** mischievous face-making ***monkey** at the zoo
- d. **those** ***monkey** at the zoo

...was/were her friend(s).

Thus, Wagers et al. predicted that adding intervening modifiers between a plural determiner and a mismatching singular noun (e.g. 11c) would not affect the rate of detection of the agreement violation between the two. In contrast, the rate of violation detection was expected to be slower if the determiner was singular, and the mismatching noun was plural (e.g. 11a). In other words, in the examples above, the authors predicted no difference in the rate of error detection between (11c) and (11d), and a slower rate in (11a) compared to (11b).

Wagers et al. used a SAT (Speed Accuracy Tradeoff) task, which estimates accuracy as a function of time. In the task, participants read the sentences word by word, and were asked to judge their acceptability (acceptable, or unacceptable) by providing a series of 17 responses on each sentence, each response being cued by a tone. The first tone was presented 200 milliseconds after the onset of the last word in each sentence. After the first response, another tone would be presented, and a second response had to be provided. The tone-response series continued in this manner until the 17th tone, and the participants could change their response along the series if their opinion or degree of confidence with respect to acceptability changed. The results showed that, as predicted, agreement errors in sentences like (11c) and (11d) were detected at the same rate despite the intervening material, while in sentences like (11a) the rate of error detection was slower than in sentences like (11b). Wagers et al. argued that the marked, plural feature is maintained in focal attention throughout the intervening material, so there is no need to retrieve it for feature-checking at the noun. In contrast, the singular feature is more easily displaced from the focus of attention. Its representation is not robustly maintained

through the intervening material, and thus may need to be retrieved at the noun. The result is the observed decrease in the rate of violation detection.

Wagers et al.'s (2009) proposal for the processing of marked, versus unmarked features brings theories like Distributed Morphology and Markedness together with the results from the numerous studies reporting feature asymmetries in processing. The hypothesis that marked features stay longer in the focus of attention than unmarked features is consistent with Distributed Morphology's proposal that marked features reliably indicate the presence of a feature specification, and unmarked features indicate the presence of a feature, such as gender, but not a specific feature value. This hypothesis is going to be of special relevance for the current study, which is one of the few directly addressing the question of how different feature values affect the processing of gender agreement, in both native speakers and L2 learners. Before tackling the topic of the use of specific feature information in L2 agreement processing, the next section provides some general background on L2 learner behavior with agreement dependencies, and the main L2 theories that have been proposed to explain that behavior.

4. L2 morphological variability

Agreement morphology is notoriously problematic for L2 learners, even at advanced levels of proficiency. Many studies show that the agreement errors observed in learners are systematic (Bruhn de Garavito, 2003; Bruhn de Garavito & White, 2002; McCarthy, 2007, 2008, 2012; Montrul et al., 2008; Prévost & White, 2000; Renaud, 2010, 2011, 2012; White et al., 2004). For example, in L2 Spanish and French, where agreement has been studied most extensively, errors emerge in the form of characteristic asymmetries

across learners. In Spanish, the target language in the current study, there are two genders: masculine, and feminine. Canonical masculine and feminine nouns consistently end in *-o*, and *-a*, respectively. Specifically, 99.9% of nouns that end in *-o* are masculine, and 96.3% of nouns that end in *-a* are feminine (Teschner & Russel, 1984). Thus, in the case of canonical nouns, lexical gender is transparent, while in the case of non-canonical nouns such as *el mapa* “the_{Masc} map_{Masc},” it is not. With regards to number, the Spanish system distinguishes between singular and plural. The canonical singular form of nouns is unmarked, while the canonical plural is formed by adding an *-s* to the root of the noun. Both gender and number are expressed within the Determiner Phrase (DP henceforth) through agreement between the noun, determiner, and any adjectives present.

One of the situations where the observed systematic errors usually emerge, is when learners supply agreement morphology on adjectives (which agree with the nouns they modify). Specifically, error types like (12), where masculine (12a) and singular (12b) forms are overextended to feminine and plural contexts (respectively), occur more frequently than (13), where feminine (13a) and plural (13b) forms are incorrectly used in masculine and singular contexts, respectively (McCarthy, 2008). Some researchers have argued that these patterns reflect use of ‘default’ morphology by learners (McCarthy, 2008; Prévost & White, 2000; White et al., 2004)—defaults being defined as underspecified forms that appear in both target-like and non-target-like positions (McCarthy, 2008), where masculine and singular are the underspecified or default forms, and feminine and plural are the specified or marked forms.

- (12) a. **La** casa es *viejo.
 The_{FemSg} house_{FemSg} is old_{MSg}

- b. Las casas son *vieja_.
 The_{FemPl} houses_{FemPl} are old_{FemSg}
- (13) a. El libro es *vieja.
 The_{MSg} book_{MSg} is old_{FemSg}
- b. La casa es *viejas.
 The_{FemSg} house_{FemSg} is old_{FemPl}

Several L2 theories have been proposed trying to account for the fact that these agreement errors are persistent, even through late stages of acquisition (Franceschina, 2005; Lardiere, 1998a, 1998b, 2000; McCarthy, 2008). Some of these theories argue, in general terms, that there is some kind of representational deficit in the L2 grammar that is responsible for learners' errors. These theories reject the possibility of native-like attainment in L2 learners. In contrast, other group of theories claims that native-like attainment is possible for learners, but their performance may not accurately reflect their competence. From this standpoint, L2 agreement errors result from the increased computational burden associated with processing a second language.

Within the group supporting a grammatical deficit in learners, there are theories like the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006; Clahsen et al., 2010). According to the SSH, grammatical processing in native speakers and L2 learners is qualitatively different. Specifically, grammatical processing in native speakers involves full parsing of complex hierarchical structures. L2 learners, on the other hand, are argued to lack hierarchical representations in their L2 grammar. Consequently, their processing is proposed to predominantly follow a 'shallow' route, which is argued to rely on strong associative form patterns, lexical-semantic information, and world knowledge. Thus, the theory predicts that native-like processing of agreement in L2 learners is restricted to

‘local’ contexts, for example when the agreeing elements are closely adjacent, while processing of ‘non-local’ or ‘long-distance’ agreement, for example across phrase boundaries, is expected to be non-target-like. Keating (2009) is an example of one study showing the patterns predicted by this theory. Keating used eye-tracking to explore the role of distance on the establishment of gender and number agreement in (L1-English) low, intermediate, and advanced, adult learners of Spanish, as well as native speakers. The study included sentences with agreement violations within a DP (local context, 14), across a Verb Phrase (VP; non-local context, 15), or across a Complementizer Phrase (CP; non-local context, 16).

(14) Un **trabajo** ***aburrida** es ideal para alguien que no tolera el estrés.

a_M job_M boring_{Fem} is ideal for somebody who doesn't tolerate stress

(15) Un **trabajo** es bastante ***mala** cuando no ofrece vacaciones o días libres.

a_M job_M is quite bad_{Fem} when it doesn't offer vacation or days off

(16) Un **libro** no se lee rápidamente cuando es ***aburrida** y difícil.

a_M book_M can't be read fast when it is boring_{Fem} and difficult.

The participants were asked to read the Spanish sentences, as well as their English translations while their eye-movements were recorded, and then respond as to whether the English sentences were accurate translations of the Spanish sentences. The results showed that only advanced learners were sensitive to agreement violations, and crucially, only in local contexts (within DP conditions). Keating interpreted his results as consistent with the SSH, and argued for a processing deficit in L2 learners. Nevertheless, this theory has also been challenged by other studies. For example, Alemán-Bañón et al., (2014) conducted an EEG study that tested the detection of gender and number agreement violations within the phrase (17), and across the phrase (18) in advanced, L1-English

learners, and native speakers of Spanish.

(17) El cerebro es un **órgano** muy ***compleja** y el cerebelo también.
the brain is an organ_{MSg} very complex_{FemSg} and the cerebellum too

(18) El **cuadro** es ***auténtica** y el grabado también.
the painting_{MSg} is authentic_{FemSg} and the engraving too

Their results showed a robust P600 (the brain response typically elicited by agreement violations in native speakers) in both L2 learners and native speakers, for both features, in both within, and across the phrase contexts. Interestingly, the two groups showed increased sensitivity in the within-phrase context as compared to the across-phrase context, suggesting that structural distance plays a role in native speaker processing as well. Crucially, in this study, L2 learners were sensitive to agreement violations in a non-local context, contra SSH.

Another theory arguing for a representational deficit in L2 learners is the Representational Deficit Hypothesis, proposed by Hawkins (2009). According to this theory, the specific asymmetries that emerge in L2 learners are the result of faulty representations of features not instantiated in the learners' L1. Specifically, and unlike native speakers, adult L2 learners are argued to build their representations by tracking frequency patterns in the input. Thus, they adopt the most frequent forms (masculine, in Romance languages like Spanish and French) as a default, and they store in memory the less frequent ones (feminine, in Romance languages), together with the nouns with which they co-occur. For example, Hawkins proposes the L2 representation in (19) for French determiners:

(19) *le* » [D, +definite]

la » ____{D, +definite} *cravatte* [*sic*], ____{D, +definite} *tartine*, ____{D, +definite} *forme*

Thus, L2 learners are expected to use the masculine *le* with all definite singular nouns, except with those nouns that the learner has memorized as feminine exceptions. Nevertheless, when the learner has weak memory representations of specific feminine nouns, optionality will emerge, and the learner will produce those nouns both with the feminine, and the masculine determiners.

For the specific language that concerns the current study, Spanish, the fact that masculine is the most frequent form would explain why L2 learners use it as a default. However, as Hawkins points out, masculine is also the underspecified form in feature representations proposed for native speakers (Harley & Ritter, 2002). That is, there is overlap in Spanish between the most frequent form, and the underspecified form. Thus, following this proposal, it is difficult to tease apart whether the asymmetrical patterns that emerge in learners stem from their statistical learning of gender information based on frequency information, or whether the pattern demonstrates a native-like representation of features. Hawkins (2009) reasons that languages like Dutch offer a better test case than Spanish to tease the two possibilities apart. In Dutch, in contrast to Spanish, there is no overlap between the underspecified form and the most frequent form, in the case of adjectives in attributive position. While the most frequent adjective form in Dutch is the bare adjective, the underspecified form in attributive position seems to be the adjective inflected with *-e*. Hawkins discusses a study by Blom, Poliřenská, & Weerman (2008) that takes advantage of this fact to test the production of Dutch adjectives in children learning Dutch as their L1 or their L2, and in adult learners at different proficiency levels.

Blom et al. used an elicited production task to examine the error types produced in the different groups, with the goal of investigating the effects onset of acquisition on the learning of grammatical gender. Blom et al. found that children learning Dutch as their L1, as well as child L2 learners, and proficient adult L2 learners overgeneralized the inflected form of adjectives. In contrast, a group of less proficient adult L2 learners overgeneralized the most frequent bare adjective form. Hawkins argues that age of acquisition determines the way in which features are acquired, with child L2 learners patterning like L1 learners, and late L2 learners patterning according to frequency information. As for the higher proficiency late L2 learners that patterned similar to the younger learners, Hawkins argues that they are becoming sensitive to attributive contexts, where the inflected adjective is more frequent than the bare adjective. Thus, they are still argued to rely on frequency patterns. Nevertheless, this is clearly speculation, since the data provide no evidence that despite showing the same pattern, higher proficiency learners are relying on frequency, while younger learners are relying on grammatical representations. In addition, it is also possible that frequencies in the input have some effect on native speaker grammars; this study did not examine this question. The current study will examine the effects of frequency differences between the two Spanish genders (masculine and feminine) in both L2 learners, and native speakers.

In contrast to the representational theories reviewed above, other researchers have proposed that the attested patterns of variability in L2 learners result from quantitative differences in processing resources, rather than qualitative differences, with respect to native speakers. Specifically, the Missing Surface Inflection Hypothesis (MSIH) (Haznedar & Schwartz, 1997; Prévost & White, 2000) proposes that learners may have

difficulties accessing target morphological forms under a processing burden, for example in spontaneous production or timed tasks, where the immediate nature of the tasks can make computation of agreement more taxing. In other words, real time processing in the L2 may be more effortful than in the L1, leaving less resources available for accurate retrieval of inflected forms (McDonald, 2006; Hopp, 2010). In addition, some researchers have proposed that this processing difficulty may be exacerbated when the features involved in agreement are not shared by the L1 and the L2 (Hopp, 2010; López Prego & Gabriele, 2014). In order to test these hypotheses, some studies have used different manipulations to make processing more effortful for native speakers as well, in order to explore whether they show variability similar to that observed in L2 learners (Hopp, 2010; López Prego & Gabriele, 2014; McDonald, 2006).

In one of the first studies investigating how processing burden affects native speaker performance, McDonald (2006) tested native speakers and L2 learners of English from 15 different language backgrounds. She conducted two experiments investigating various grammatical constructions (word order, regular past tense, S-V agreement, plural agreement, etc). In Experiment 1 she administered both a grammaticality judgment task testing said constructions, and several independent tasks measuring the participants' individual processing abilities, including working memory span, decoding ability, and speed of processing. Both native speakers and L2 learners took the processing measures in English, as the aim was to measure processing capabilities in the native and second language respectively. The results showed that the individual processing abilities measured were significantly poorer for the L2ers (in the L2) than for native speakers, suggesting that the efficiency of the processor is reduced when we use a second language.

The performance of the learners on the grammaticality judgment task revealed better performance on items testing word order and question types, compared to articles and past tense morphology. Moreover, positive correlations were found between the participants' performance in the grammaticality judgment task, and L2 working memory, and decoding ability. In order to further explore the relationship between processing difficulty and performance on grammaticality judgments, a second experiment was conducted.

In this experiment, additional native speaker groups were given different grammaticality judgment tasks containing the same sentences as in Experiment 1. Each task involved the judgment of the sentences under a different 'stress' condition: memory load (low and high), listening through noise, response deadline, and compressed speech. Each condition targeted one of the specific processing capacities previously tested (working memory span, decoding ability, and speed of processing). Results showed that working memory and decoding ability scores correlated with the participants' performance in the conditions burdening these processing abilities, namely the memory load and noise conditions respectively. Crucially, the results revealed a very similar scale of vulnerability in the constructions tested for stressed native speakers and L2 learners in Experiment 1: articles and regular morphology were the most vulnerable structures, and word order was the least affected by the stressors. The noise and high memory load conditions yielded the strongest similarity between natives and learners.

Two important conclusions can be drawn from this study: first, that processing abilities are reduced in a second language; and second, that morphology is considerably susceptible to processing stress, both in the first and second languages. These results

support the proposal that differences in morphology between native speakers and L2 learners may result from the increased computational difficulty of processing sentences in a second language, which may resemble listening to sentences through noise or with a high memory load in the native language.

Following McDonald's steps, Hopp (2010) investigated the processing of several properties of German (case, S-V agreement, word order, auxiliary selection, etc.) in advanced and near-native speakers, as well as native speakers performing under processing burden. The L2 learners had different L1s, either English, Dutch, or Russian, the latter being the most similar to German with respect to the morphological properties tested. The learners were tested in three tasks differing on the computational burden they entailed: an untimed grammaticality judgment task (GJT), a self-paced reading task (SPR task), and a speeded GJT. The results for the advanced group showed that only the L1 Russian learners performed like native speakers, and only in the untimed GJT. Interestingly, while all three near-native groups showed native-like performance in the untimed GJT and in the SPR task, only the L1 Russian near-native group showed native-like performance in the speeded GJT. Hopp concluded that, while learners may be able to retrieve grammatical knowledge in untimed tasks regardless of the L1, they may not be able to access it under increased processing demands. However, similarities between the L1 and L2 may facilitate computation in the L2. In addition to L2 learners, Hopp tested native speakers of German performing under different computational burdens to examine whether native speakers would show "L2-like" patterns under taxing circumstances, despite having intact grammatical competence. Confirmation of this hypothesis would mean that L2 learners may have an intact L2 grammar that emerges as flawed sometimes,

due to processing burden. The native speakers took the speeded GJT at five different rates of presentation. Interestingly, and confirming Hopp's hypothesis, the native speakers in the fastest speed of presentation showed similar declines in accuracy and similar error patterns to near-native speakers in their speeded GJT. This finding brought a new type of evidence in support of computational accounts of morphological variability in L2 learners.

In addition, supporters of the processing burden approach (Prévost and White, 2000; White et al., 2004) have proposed that the observed use of default morphology (e.g. overgeneralizations of masculine) by L2 learners stems from an asymmetrical representation of features, as characterized for native speakers within the Distributed Morphology (DM) framework explained above (Halle & Marantz, 1993; Harley & Ritter).

Prévost and White (2000) were the first (to the author's knowledge) to relate the use of default morphology in L2 learners to the feature hierarchies proposed for native speakers in DM. Recall that in DM the insertion of a subset of the features specified in the syntax (an unmarked, or default form) is allowed, and only results in an agreement violation when a more specified vocabulary item is available (see section 2 above). However, this type of violation does not involve a clash between the features specified in the syntactic context, and those of the vocabulary item, since the vocabulary item lacks specification. In the opposite scenario, where a marked form is inserted in the place of an unmarked form, the specification of the vocabulary item clashes with that of the syntactic context. It is this clash of features that is not allowed under Distributed Morphology. Prévost and White proposed that learners may insert the underspecified, or 'default' form when they have difficulty accessing the target vocabulary item, particularly in the context

of demanding tasks. In these contexts, the parser may halt the search for a ‘perfect match’ of features once it finds a ‘good enough’ match in a subset of those features (the underspecified form), thus freeing processing resources to perform the task. This proposal then, provides a computational explanation for the most frequent type of agreement error in L2 learners, namely the use of default morphology. In addition, it understands the low occurrence of violations involving the insertion of clashing features as compliance with DM’s principle forbidding this type of insertion.

In order to test Prévost and White’s (2000) proposal for the use of default morphology in L2 learners, McCarthy (2008) conducted a study testing learners of Spanish⁴. Following DM, McCarthy matches markedness and the specification of features: Underspecified features are unmarked, and specified features are marked (Harley & Ritter, 2002). McCarthy (2008) hypothesized that L2 errors are instances where underspecified/unmarked forms (e.g. masculine) are overextended to marked (feminine) contexts (20), rather than the opposite pattern, where specified/marked forms are overextended to unmarked contexts (21).

(20) El niño tiene una_{FemSg} manzana_{FemSg}. *LO_{MSg} está comiendo.
 the boy has an_{FemSg} apple_{FemSg} it_{MSg} he is eating
 The boy has an apple. He is eating *it.

(21) El niño tienen un_{MSg} plátano_{MSg}. *La_{FemSg} está comiendo.
 the boy has a_{MSg} banana_{MSg} it_{FemSg} he is eating
 The boy has banana. He is eating *it.

She explored this hypothesis by testing intermediate and advanced English-speaking learners of Spanish in a production and a comprehension task. The production task aimed

⁴ McCarthy investigated both gender and number agreement. For simplicity, and given that the focus of the present study is gender, only gender will be discussed here.

at eliciting adjectives and direct object clitics, also inflected for gender and number in Spanish. Participants were shown pictures with agents acting on objects (e.g. a boy eating an apple) and were asked questions about them by a native speaker of Spanish. The comprehension task was adapted from White et al. (2004). In the task participants were asked to interpret clitics embedded in a story by choosing the referent that corresponded to the clitic, among three pictures. The choice could only be done based on the gender and number of the referents. Her results showed error patterns that were qualitatively similar in both tasks: overextensions of unmarked forms (20) were more frequent than overextensions of marked forms (21), consistent with her prediction.

Following up on McCarthy (2008), and also building on Prévost & White's (2000) proposal, López Prego and Gabriele (2014) conducted a study investigating L2 learner and native Spanish judgments of the agreement error types discussed by McCarthy. In order to test computational and representational accounts of morphological variability, and following Hopp (2010), they also tested other learner and native groups on the same judgments, under different task demands. Examples (22) and (23) are two items showing the ungrammatical conditions tested in the study. In example (22) the head noun was masculine, as was half of the target sentences, and the adjective modifying the noun is in the feminine form, constituting an error of overextension of the marked feature.

(22) Overextension of marked feature (feminine)

*Juan dijo que vio un **colegio** que era **antigua** en Londres.
Juan said that he saw a school_{MSg} that was old_{FemSg} in London.

In example (23) the head noun was feminine, as the other half of target sentences, and the adjective modifying the noun is in the masculine form, constituting an error of overextension of the unmarked feature.

(23) Overextension of unmarked feature (masculine)

*Juan dijo que vio una **tela** que era **fino** en París.

Juan said that he saw a fabric_{FemSg} that was fine_{MSg} in Paris.

The L2 learners took either a Speeded grammaticality judgment task (GJT) in which the sentences were presented word by word at a rapid pace, and a fast grammaticality judgment was required at the end, or an Untimed GJT in which the sentences were presented at once, with no time limit to read or judge them. In addition, three groups of native speakers took the Speeded GJT at three different presentation rates. The same items were used in all tasks. López Prego and Gabriele hypothesized that if, in line with McCarthy's proposal, L2 errors are generally overextensions of underspecified forms, these errors (23) may be harder for learners to detect than overextensions of marked forms in grammaticality judgments. Thus, errors involving overextension of marked forms (22) may be easier to detect. They also hypothesized that if, in line with Prévost and White (2000), the patterns attested in L2 learners of Spanish stem from processing difficulty, native speakers may show similar patterns when performing under processing burden.

The results showed that native speakers in the fastest Speeded GJT performed better with overextensions of marked forms (22), as hypothesized by López Prego and Gabriele, and consistent with DM. Interestingly, L2 learners performed better with the overextensions of unmarked forms (23).

In considering the unexpected L2 pattern, López Prego and Gabriele relied on a proposal put forth in Wagers et al.'s (2009) study on number agreement in English. Recall that from this study, Wagers et al. (2009) concluded that upon encountering a marked feature in an item like a determiner, in which case a noun also carrying a plural feature is expected, the feature information doesn't need to be retrieved again at the noun for feature checking, because it has been maintained in the focus of attention. In contrast, unmarked features are displaced from the focus of attention more easily due to their lack of specification, and therefore, need to be retrieved at the agreeing element to check for agreement. López Prego and Gabriele applied Wagers et al.'s logic to explain their L2 results. Thus, pointing at the word-by-word presentation in their speeded task, they argued that encountering a marked feminine DP first (*una tela* 'a_{FemSg} fabric_{FemSg}') in the overextensions of unmarked features like (23) could have helped the learners detect the agreement error at the adjective (*fino* 'fine_{MSg}'). In contrast, in the overextensions of marked features (22), the unmarked, masculine feature is encountered first (*un colegio* 'a_{MSg} school_{MSg}'), providing no reliable evidence of the presence of a feature. Thus, a prediction about subsequent features following in the sentence may not have been generated. Instead, the authors propose, upon encountering the disagreeing feminine adjective in (22) (*antigua* 'old_{Fem}'), the parser may have, particularly under the processing burden of the Speeded GJT, 'fill in' the gender feature with a feminine specification, leading to an acceptance of the agreement violations in (22), and making these errors harder for learners to detect. Following Hopp (2010), the authors speculated that the processing burden for learners may have been further exacerbated by the fact that

they were dealing with a feature not present in their L1, which would explain the difference in the pattern with native speakers.

This hypothesis, however, could not be directly tested in the study, since the task used didn't provide data on the participants' online processing of the words in the sentences, only on their offline grammaticality judgments at the end of the sentence. It remains unclear what might underlie the difference in the use of markedness information by learners and native speakers in the online processing of agreement. The current study aims at answering that question, in order to further explore the representation of features in L2 learners and native speakers.

5. The use of featural information in L2 agreement

A growing number of L2 studies have started to investigate whether and how L2 learners use featural information in their online processing of agreement. Several of these studies have used a visual world eye-tracking paradigm to investigate this question. However, the results originating from this series of studies are conflicting. While some of these studies suggest that L2 learners are able to use featural information as a predictive cue of agreement, others find the opposite. For example, Lew-Williams and Fernald (2010) conducted a series of visual world experiments testing L2 learners, and native speakers of Spanish on their online processing of gender agreement. In those experiments (most of which used the visual world eye-tracking paradigm), participants were presented with picture displays showing objects with either the same gender (same-gender trials), or different genders (different-gender trials). At the same time, sentences such as *¿Dónde está la_{FemSg} galleta_{FemSg}?* ('Where is the cookie?') or *Encuentra el_{MSg} pájaro_{MSg}* ('Find the

bird') were presented, while the eye movements of the participants were tracked. Thus, for the same-gender trials, the determiner in the sentences didn't provide a gender cue, as the lexical gender of all of the items in the display matched the gender of the determiner, but it did in the different-gender trials, as the lexical gender of only one of the items in the display matched the gender of the determiner. If participants were using the gender information on the determiner to anticipate what object would be mentioned in the sentence, participants' looks to the target object would occur earlier (as soon as the participants heard the determiner) in the different-gender trials than in the same-gender trials. The results showed that this was indeed true for native speakers. However, L2 learners were not able to use the gender of the determiner as a predictive cue, and waited until the noun was revealed in the sentence to initiate looks to the target object. Similar results were obtained in another visual world eye-tracking experiment by Grüter et al. (2012), also investigating learners of Spanish. In this study, it was found that advanced L2 learners were able to use the gender cues in determiners predictively with unfamiliar nouns to which they were exposed at the beginning of the experiment. In contrast, and unlike the native speakers, the learners didn't use the same cues predictively with familiar nouns. Grüter et al. argued that the learning of novel nouns in their experiment (in conjunction with determiners) was more similar to that of L1 acquisition, with the consequent development of strong associations between the nouns and gender nodes. That is, because the learners were exposed for the first time to the novel nouns together with the determiners (auditorily), rather than independently of the determiner (like a new vocabulary item on a textbook), Grüter et al. argue that the learners used co-occurrence relations between the determiners and the nouns, to figure out the gender of the novel

nouns. This would resemble L1 acquisition to a greater extent, and thus result on stronger associations between these nouns and the gender nodes. Consequently, lexical access of these nouns is faster, and gender cues in the online processing of these nouns can be used more efficiently. In contrast, the authors emphasize, the usual conditions for gender learning in the L2 are different from those in the L1 environment; L2 learners do not rely on distributional information, and thus they only develop weak links between gendered items and gender nodes, resulting in slower, and less efficient lexical access.

In contrast to the studies above, Dussias et al. (2013), and Hopp (2013) find evidence of use of gender information in the online processing of agreement in L2 learners. Dussias et al. also conducted a visual world-eye-tracking paradigm experiment to test whether gender cues on an article would facilitate the processing of a subsequent noun, in L2 learners of Spanish. They tested high and low proficiency English-speaking learners, and a group of low proficiency Italian-speaking learners, as well as a control group of native speakers of Spanish. As in the other visual-world paradigm studies, participants were presented with picture displays while they listened to sentences containing a determiner and a noun naming one of the pictures (e.g. *El estudiante estaba dibujando **el reloj** que vio ayer*, ‘The student was drawing **the clock** that he saw yesterday’). In this study two pictures were presented in each display, depicting a target and a distracter. Again, in some of the trials the target and the distracter had the same gender, while in other trials, they had different genders, such that the gender cue in the determiner was disambiguating. The participants were asked to click on the picture that was mentioned in the sentence, while their eye movements were tracked. In addition, after listening to each sentence participants were asked to perform a plausibility judgment

task. The results showed that both the native speaker group, and the high proficiency English group used the gender cues on the determiner, as they showed anticipatory looks to the target noun prior to hearing it, in gender-different trials. The Italian group only showed anticipatory effects on feminine trials, which the authors hypothesized, could be due to the higher percentage of non-canonical nouns in masculine than in feminine trials⁵, or to the differences between masculine determiners in Spanish and Italian⁶, in addition to the lower proficiency of the Italian group. Finally, the low proficiency English group showed some surprising results, with faster looks only for masculine trials, and only for same-gender trials. Dussias et al. propose the possibility that these learners may have tried to use the masculine gender information in gender-different trials. However, given the low proficiency of the group, trying to use the gender cue may have been substantially taxing, rather than beneficial, resulting in delayed looks. As for the same gender trials, Dussias et al. propose that the learners may not have been making an effort to integrate the gender information when the two pictures had the same gender, thus resulting in faster looks. Given that there was no evidence of predictive processing in feminine trials for this last group, the authors suggest that there may be different time courses involved in learning to process masculine versus feminine gender. Leaving aside the results for the low proficiency groups, importantly, the study finds evidence of online use of gender information in the high proficiency English-speaking group.

Hopp (2013) also used the visual-world paradigm methodology to test L1 English

⁵ Non-canonical nouns present extra difficulty for L2 learners due to the fact that they are phonologically non-transparent, that is, their phonological form doesn't contain a gender cue. Thus, knowing the gender of these nouns requires memorization. If the learner assigns the wrong gender to the noun, this error could be incorrectly interpreted as an agreement error.

⁶ While Spanish only has the masculine singular determiner *el*, Italian has two: *il* and *lo*, which are phonologically conditioned.

learners of German on their processing of gender. The ultimate goal of the study was to investigate the causes of morphological variability in adult L2 learners, and thus, Hopp tested some of the L2 theories explained above that are also relevant to the current study. He specifically investigated Grüter et al.'s (2012) lexical proposal (summarized above) that L2 learners are variable providing agreement morphology because of their weaker links between nouns and gender nodes, compared to native speakers. In order to test this hypothesis, Hopp examined the relationship between gender assignment in a production task, and the predictive use of gender cues in a visual-world paradigm task. Hopp also tested the Missing Surface Inflection Hypothesis (MSIH; Prévost & White, 2000; White et al. 2004). Recall that this theory posits that morphological variability in learners stems from difficulty accessing target morphological forms, specifically emphasizing increased processing demands as the source of difficulty. Contrasting with these theories, Hopp tested Hawkins's (2009) Representational Deficit Hypothesis (RDH), which, as already discussed, postulates a representational impairment for adult L2 learners learning L2 properties not present in their L1. Hopp specifically tested Hawkins's (2009) claim that learners whose L1 lacks gender resort to frequency related strategies in order to manage gender in their L2. Recall that the strategy Hawkins posits is initially pairing all nouns with the most frequent, default form of the article, and incrementally memorizing exceptions when nouns are encountered with other article forms. German has three genders: masculine, feminine, and neuter, this last one being the marked form. Crucially, the gender system of German is phonologically and semantically opaque, which minimizes the possibility of L2 learners relying on form associations between determiners and nouns to process agreement dependencies. In addition, Hopp used the

visual-world paradigm to test RDH's specific prediction that anticipatory looks will only occur for the non-default (neuter), memorized determiner-noun pairs, since a default gender cue should not activate a specific set of memorized nouns. On the other hand, according to Grüter et al.'s (2012) lexical learning hypothesis and the MSIH, predictive use of gender cues should be evident across the three genders in German.

Hopp (2013) tested a group of advanced, and near-native L2 learners of German and a group of German native speakers in a production task, and a comprehension task. The participants were first tested on their knowledge of lexical gender in the production task. In this task they were asked to name objects on four-picture displays, along with a determiner and an adjective describing the color of the object. Next, the same objects were presented in the comprehension visual-world task, in which the learners listened to sentences like "Where is the yellow X?", or "Where do you see two X?" while they looked at the four-picture displays. The participants' eye movements were tracked through the presentation of the sentences.

For statistical analyses the L2 learners were divided into those who were consistent in their production of gender for a particular noun (gender-consistent), and those who showed inconsistencies (gender-inconsistent). The results revealed that only the gender-consistent group and the native speaker group used gender information in a predictive manner. The results also revealed a relationship between the performance of the gender-consistent group in the gender assignment production task, and the use of gender cues predictively in this group, thus supporting Grüter et al.'s (2012) hypothesis that only learners with strong gender nodes can take advantage of gender cues in predictive processing. In addition, both of these groups used the three genders in a

predictive manner, thus countering the RDH, and providing evidence for lexical accounts of morphological variability (Grüter et al., 2012; Prévost & White, 2000; White et al., 2004).

The L2 studies outlined above offer a general picture of the research conducted on the online processing of gender agreement in the L2. The relative novelty of this type of research, and the disparity of results obtained demand further investigation into the question of whether L2 learners can use featural information online, for agreement purposes. If so, the understudied question of whether they use only marked feature information, and whether that is also the case for native speakers becomes essential to test for similarities and differences between native speaker and L2 feature representations. In addition, exploring whether frequency differences play a role in any feature asymmetries that may emerge in both groups would speak to whether L2 learners and native speakers rely on similar mechanisms to build gender representations. Specifically, while some researchers have argued that agreement error patterns emerging in advanced L2 learners may actually reflect a native-like representation of features (Prévost & White, 2000; White et al., 2004), others claim that such error patterns reflect frequency-based memorization strategies in L2 learners, but markedness-related feature representations in native speakers. The current study addresses these issues in an attempt to tease apart L2 theories positing a grammatical deficit in L2 learners (Clahsen & Felser, 2010; Hawkins, 2009), versus theories arguing for the potential of native-like L2 grammars that emerge in a non-native manner due to the processing burden associated with using an L2 (Prévost & White, 2000; White et al., 2004).

6. Present study

The current study investigates how advanced L1-English learners and native speakers of Spanish process long-distance gender agreement dependencies online. In order to do this, a self-paced reading experiment was conducted testing gender agreement between a preposed adjective (*blanco*, in 24) and a noun (*vestido*, in 24) in a long-distance dependency (across a Complementizer Phrase; CP), as shown in (24).

(24) a. Como es **blanco**, he decidido que modificaré **el vestido** que dejé en la entrada.
“As it is white_{MSG}, I have decided that I’ll modify the dress_{MSG} I left in the hall.”

The study directly addresses three main questions:

RQ1: Are L2 learners sensitive to long-distance gender agreement violations?

This question is investigated by manipulating the grammaticality of the gender agreement dependency between the preposed adjective and the noun. By means of this manipulation the study directly tests the Shallow Structure Hypothesis, which proposes that L2 learners cannot process agreement in a long-distance dependency.

RQ2: Do L2 learners and native speakers use markedness information in their processing of agreement? This question is examined by manipulating the markedness status of the features involved, as well as the gender information provided by the preposed adjective, in grammatical sentences. This manipulation tests whether L2 learners and native speakers represent and use features similarly, as proposed by the Missing Surface Inflection Hypothesis. Additionally, the relationship between frequency differences in the input and the use of markedness information in both groups is examined. This exploratory analysis tests the predictions of the Representational Deficit Hypothesis, which argues that the feature asymmetries observed in L2 learners are related to a frequency-based compensatory strategy, rather than native-like feature

representations.

RQ3: Do native speakers performing under processing burden show similar patterns to L2 learners? In order to explore this question an additional group of native speakers is tested on the same experiment, with an added computational burden. This experimental approach attempts to tease apart computational theories that attribute agreement variability in learners to processing issues, from representational theories that posit a grammatical deficit in learners as an explanation for agreement errors (Hopp, 2010; López Prego & Gabriele, 2014; McDonald, 2006).

By examining these questions and testing these theories the study informs our knowledge of the similarities/differences existing in agreement processing in L2 learners and native speakers, specifically for features not present in the learners' L1. Thus, the study sheds light on the nature of feature representations in native speakers and L2 learners, and in turn, on the source of the systematic agreement errors observed in L2 learners. The study provides new and unique data to help determine whether L2 learners and native speakers have similar asymmetrical representations of gender features that surface under processing burden (López Prego and Gabriele, 2014; Prévost & White, 2000; White et al., 2004), or whether L2 feature representations are different from native speakers', and result on agreement errors (Clahsen & Felser, 2006; Clahsen et al., 2010; Hawkins, 2009; Keating, 2009). Lastly, the study furthers our knowledge of what is ultimately possible for learners regarding the processing of agreement, as it focuses on advanced learners.

6.1. Predictions

The predictions that each of the theories tested make for the research questions are

explained below:

RQ1: Are L2 learners sensitive to gender agreement errors in long-distance agreement dependencies?

The Missing Surface Inflection Hypothesis predicts that advanced L2 learners can potentially show sensitivity to agreement violations in these dependencies, just like native speakers. In contrast, the Representational Deficit Hypothesis predicts that, unlike native speakers, L2 learners will not be sensitive to the violations, because they have a flawed representation of the gender feature. Similarly, the Shallow Structure Hypothesis predicts a lack of sensitivity in L2 learners, because the violations occur across phrase boundaries.

RQ2: Does the marked status of a feature facilitate processing in L2 learners and native speakers?

Under the Missing Surface Inflection Hypothesis, L2 learners and native speakers share an asymmetrical representation of features like that proposed in Distributed morphology. If that is the case, both learners and native speakers are predicted to show a facilitation effect after encountering a marked (feminine) feature on the adjective (e.g. Wagers et al., 2009). Facilitation may not emerge after the unmarked (masculine) feature, since under Distributed Morphology unmarked features don't provide reliable evidence of the presence of a feature specification.

The Representational Deficit Hypothesis makes the same prediction as the Missing Surface Inflection Hypothesis for native speakers. However for L2 learners, the theory predicts that a facilitation effect after a marked feature is related to frequency differences in the input, and not to a native-like representation of gender. Specifically, L2 learners are argued to memorize the non-default, marked (feminine) form of agreeing

elements such as adjectives, together with the nouns with which they co-occur in the input. Thus, when learners encounter the marked form of an adjective (feminine), a set of memorized feminine nouns will be activated, facilitating the processing of a subsequent feminine noun (see Hopp, 2013 for discussion). Crucially, then, the higher the frequency with which a specific feminine adjective occurs with a specific noun, the larger the effect of facilitation that would be observed. Therefore, a relationship between frequency of co-occurrence of noun-adjective pairs and facilitation effects should be found only in L2 learners, and only for feminine items. This relationship should not emerge in native speakers, since they don't need to memorize exceptions to the default to compensate for a deficient gender representation, and not for masculine items, as these are not stored.

Finally, under the Shallow Structure Hypothesis, facilitation effects across phrases in L2 learners are ruled out, since the theory proposes that L2 learners cannot establish agreement in long-distance dependencies.

RQ3: Will native speakers performing under processing burden reveal similar patterns to L2 learners?

Computational accounts of morphological variability in L2 learners, like McDonald (2006) and the Missing Surface Inflection Hypothesis, predict that if L2 learner patterns result from processing issues rather than a grammatical deficit, similar patterns may emerge in native speakers performing under processing burden.

In contrast, accounts that argue for a grammatical deficit in L2 learners as their source of variability in agreement predict that, since native speakers have an intact grammar, they should not resemble L2 learners. Theories like the Shallow Structure Hypothesis and the Representational Deficit Hypothesis fall under this perspective.

6.2. Method

6.2.1. Stimuli

As mentioned above, the target sentences in the experiment involved gender agreement between a preposed adjective and a noun across a CP, introduced by the complementizer *que*, “that”. An example of one set of items is given below in (25):

(25)

Marked-Unmarked

a. Como es ***blanca**, he decidido que modificaré **el abrigo** que dejé en la entrada.

Invariant-Masculine

b. Como es **suave**, he decidido que modificaré **el abrigo** que dejé en la entrada.

Masculine-Masculine

c. Como es **blanco**, he decidido que modificaré **el abrigo** que dejé en la entrada.

“As it is soft/white*_{FemSg/MSg}, I have decided that I’ll modify the_{MSg} coat_{MSg} that I left in the hall.”

Unmarked-Marked

d. Como es ***blanco**, he decidido que modificaré **la chaqueta** que dejé en la entrada.

Invariant-Feminine

e. Como es **suave**, he decidido que modificaré **la chaqueta** que dejé en la entrada.

Feminine-Feminine

f. Como es **blanca**, he decidido que modificaré **la chaqueta** que dejé en la entrada.

“As it is soft/white*_{MSg/FemSg}, I have decided that I’ll modify the_{FemSg} jacket_{FemSg} that I left in the hall.”

In these sentences the adjective is preposed with respect to the noun it describes, so that an expectation of encountering the noun later in the sentence is generated. This allows for the possibility of finding effects of online processing as early as the determiner in the sentences. In addition, this structure allows having intervening material between the adjective and the noun, thus allowing to test long-distance agreement.

In half of the items in each set the head noun was masculine (25a, b, and c), and in the other half, feminine (25d, e, and f), all in their singular form. For each masculine/feminine noun there were three conditions, one of them ungrammatical (25a,

and d), and two of them grammatical (25b, c, e, & f). The conditions in each set are named according to the markedness of the gender feature on the adjective and the noun in the sentence. For example, the ungrammatical condition (25a) is labeled **Marked-Unmarked** because the adjective, which is encountered first, is in its marked, feminine form, while the noun (or rather the whole determiner phrase) is unmarked, masculine. The counterpart ungrammatical condition with a feminine noun (25d) is labeled **Unmarked-Marked** because the adjective is in the unmarked, masculine form, while the noun is marked, feminine. In one of the grammatical conditions (25b, and 25e), the adjective was gender-invariant, that is, it had the same form for masculine, and feminine nouns. Thus, (25b) and (25e) are labeled **Invariant-Masculine** and **Invariant-Feminine**, respectively. This condition was included as a baseline against which to compare both the ungrammatical conditions, and the second type of grammatical conditions, explained in what follows. What makes this condition a good baseline is the fact that the invariant adjectives don't carry gender information, which helps avoid any effects on critical regions resulting from the potential markedness difference between masculine and feminine adjectives. In the other grammatical conditions (25c, and 25f), the adjective was gendered, that is, it carried gender information, and it agrees with the noun it modifies. Thus, (25c) and (25f) are labeled **Masculine-Masculine** and **Feminine-Feminine**, respectively. Comparing these conditions to the invariant baseline, where the adjectives don't provide gender information, allows us to investigate whether and how markedness information is used in the processing of grammatical sentences. That is, if the gender information on the adjective is used in agreement processing, we should see a difference between the gendered conditions and the invariant conditions. In addition, if marked

features are processed differently than unmarked features, thus facilitating the processing of an upcoming feature, the difference may emerge between the gendered conditions (25c, and 25f), but not between the invariant conditions (25b, and 25e). Crucially, exploring these hypotheses in grammatical sentences as opposed to the usual paradigm of agreement violations allows the examination of agreement processing in naturally occurring sentences.

Thirty-six sets of target sentences like those in (25) above were created, with 6 items per each of the 6 conditions⁷. These target items were distributed across six lists, in a Latin-square design. Thus, all participants saw all items, but only one version of each item. Each subject read 12 ungrammatical target sentences, 6 with a masculine noun, and 6 with a feminine noun, and 24 grammatical target sentences, 12 with an invariant adjective (half with masculine nouns, and half with feminine nouns), and 12 with a gendered adjective (half with masculine nouns, and half with feminine nouns)⁸. All the target nouns had canonical gender marking, and were controlled for frequency. In addition, all the adjective-noun pairs were also controlled for frequency, as detailed in the following section. These frequencies were used in correlation analyses explained in the analyses section (see Appendix II for the frequencies of all target adjective-noun pairs). The invariant adjectives were used three times each, due to the difficulty⁹ of finding a different adjective for each stimuli set. In order to control for any effects emerging from these repetitions, the gendered adjectives were also used three times each.

⁷ See Appendix I for a complete list of experimental items.

⁸ The total number of grammatical and ungrammatical items in the experiment was balanced in the fillers, as explained in the description of the filler items.

⁹ The number of invariant adjectives in Spanish is limited, and the list was further reduced by the requirement that the adjectives represent concrete, rather than abstract properties, so that they were associated with the intended nouns, rather than with some entity outside the sentences.

Thirty-six sets of distracters targeted the distinction between Preterite, and Imperfect. Thus, the participants read an equal number of sentences manipulating gender agreement, and the Preterite/Imperfect distinction. This type of distracter was chosen due to the fact that L2 learners typically consider this distinction to be difficult (Coppieters, 1987; Montrul & Slabakova, 2003), which would help disguise the focus of the study. These distracters included 18 grammatical items, and 18 ungrammatical items. The ungrammatical items were evident violations of the canonical use of the Imperfect, as in example (26), where the Preterite is incorrectly used to refer to a habitual activity in the past.

- (26) a. Cuando mi madre era niña, siempre *estudió en el jardín de su casa.
b. Cuando mi madre era niña, siempre estudiaba en el jardín de su casa.
“When my mother was a child she always studied in the garden of her house.”

In addition, a total of 48 filler items, 36 ungrammatical, and 12 grammatical, were included in order to balance the total number of grammatical, and ungrammatical sentences in the experiment. The ungrammatical items were person violations in clitics, as in (27).

- (27) a. El viernes Enrique *te_{2p} compró a su madre un barco de madera para su habitación.
b. El viernes Enrique le_{3p} compró a su madre un barco de madera para su habitación.
on Friday Enrique to *you/her bought his mom a boat of wood for her bedroom.
“On Friday Enrique bought his mom a wooden boat for her bedroom.”

This specific type of filler was chosen because it was considered that it wouldn't be too taxing for the learners to detect the violations, yet the violations wouldn't be so obvious that they would decrease sensitivity to the target violations.

Each participant read a total of 120 experimental items: 36 target sentences, 36

distracter sentences, and 48 filler items.

6.2.2. Stimuli controls

In the target items only nouns with canonical gender were included. In addition, a series of paired-samples *t*-tests were conducted to test for differences in the average length of nouns, and in the frequency of co-occurrence of the different noun-adjective combinations, as follows: In order to control for average length differences between masculine and feminine nouns, paired *t*-tests were conducted which revealed no significant differences ($t(35) = .47$; $p = .64$). Additionally, the co-occurrence frequencies¹⁰ of the masculine noun-adjective pairs, and the feminine noun-adjective pairs were compared in paired *t*-tests that revealed no significant differences, both for combinations with gendered adjectives (masculine noun-masculine adjective vs. feminine noun-feminine adjective; $t(35) = .67$; $p = .60$), and with invariant adjectives (masculine noun-invariant adjective vs. feminine noun-invariant adjective; $t(35) = .49$; $p = .63$). Finally, paired *t*-tests revealed no significant differences between the co-occurrence frequencies of masculine nouns combined with gendered adjectives, versus masculine nouns combined with invariant adjectives ($t(35) = .93$; $p = .36$), or between feminine nouns combined with gendered adjectives, versus feminine nouns combined with invariant adjectives ($t(35) = 1.50$; $p = .15$).

6.3. Procedure and tasks

Participants first signed an informed consent statement and filled out a linguistic background questionnaire. Then, they performed the self-paced reading task. The L2

¹⁰ These frequencies were tallied from a Google search, due to the difficulty of finding all the pairs in a corpus. The reliability of this method has been tested in several studies (Keller et al., 2002; Keller & Lapata, 2003), which found that bigram frequencies obtained from Google are highly correlated with the bigram frequencies obtained from corpora, and that the frequency of bigrams not found in corpora can be reliably predicted from Google searches.

learners performed two additional tasks measuring cognitive abilities¹¹ following the self-paced reading, and then a proficiency test, and a gender assignment task¹². The testing session for the learners was conducted individually in a computer lab at the University of Kansas, and it lasted approximately 90 minutes. One of the native speaker groups completed only the self-paced reading task, in a testing session that lasted about 30 minutes. These native speakers were tested in small groups in a computer lab at the Universidade de Santiago de Compostela, in Spain. Another group of native speakers performed the same self-paced-reading task, testing the same stimuli, with an additional processing burden as described below. This group also completed the two cognitive measure tasks completed by the L2 learners. Testing in this group was conducted individually in a quiet room, and the session lasted approximately 90 minutes. All the computer tasks were administered using the stimulus presentation program Paradigm (Tagliaferri, 2005). All the participants received payment for their participation in the study.

6.4. Self-paced reading tasks

The stimulus sentences described above were presented in a non-cumulative, self-paced moving window format (Just et al. 1982). In this format, sentences are presented first as groups of dashes on the screen, each dash representing a character in a word. When the participant presses the mouse for the first time, the first group of dashes on the screen corresponding to the first segment (word, or words) on the sentence is replaced by the corresponding segment. When the participant presses the mouse a second time, the

¹¹ The results from these tasks will not be reported in this dissertation.

¹² This was a timed task on the computer in which L2 learners were presented with the target nouns in the self-paced reading task, and were asked to select with the mouse either the masculine, or the feminine form of the article that corresponded to each noun.

segment is replaced by the dashes again, and the second group of dashes in the sentence is replaced by the next segment in the sentence. Thus, with each press of the mouse, the previous segment is replaced by dashes, and the next one is revealed. The sentence in (28) demonstrates the division of target sentences into segments as they were shown in the self-paced reading task. Of these segments¹³, the critical regions were regions 5 (determiner) and 6 (noun), and the spillover region was region 7 (complementizer). Reading times at these regions were used in statistical analyses as a measure of online processing.

(28) Como es blanco,/ he decidido/ que/ llevaré/ la/ gorra/ que/ compré/ en la boutique.
 R1 R2 R3 R4 R5 R6 R7 R8 R9

After the last segment in the sentence, the participants were presented with the words *Bien* “good” and *Mal* “bad” at the left and right of the screen respectively, and were asked to decide whether or not the sentence was a good sentence in Spanish, by choosing one of the options using the mouse¹⁴. The task started with a practice block that included 10 sentences (5 grammatical, and 5 ungrammatical) targeting number agreement. This practice block had the role of biasing the participants to interpret the adjectives in the target sentences as referring to a subsequent noun in the sentence, rather than to a noun outside of the sentence. Thus, in the practice sentences the reference of items potentially

¹³ It was decided that the complex verb form *he decidido* “I have decided” would be presented as one segment in order to avoid priming of the English personal pronoun “he” due to its orthographic overlap with the Spanish form *he* “I have”. In addition, the determiner in the target DP was presented separate from the target noun in order to detect potential effects emerging at the determiner. Previous studies (Wicha et al. 2003, 2004) have in fact found effects of prediction of gender at the determiner, in native speakers of Spanish.

¹⁴ Although the focus of the current study is on online processing, including this offline secondary task after the self-paced reading ensures that the participants are on task, rather than just clicking through the sentences. A grammaticality judgment was chosen as opposed to other tasks because of its success in a previous study (López Prego & Gabriele, 2014) yielding asymmetries in L2 learners and native speakers.

referring to entities outside of the sentence was clearly linked to entities occurring later in the sentence. In addition, feedback was provided on the judgments of all practice items, and participants were instructed to pay attention to the practice sentences, since they would be very similar to the ones in the real experiment. Examples (29) through (31) illustrate some examples of sentences that were included in the practice block. The full set of 10 practice sentences can be found in Appendix III.

- (29) Cuando quise *retomarlas, la novela ya no estaba donde la había dejado.
When I went to resume *them, the novel wasn't where I had left it any more.
- (30) Como quería terminarlo, me llevé el libro para leer en el avión.
Since I wanted to finish it, I took the book with me to read on the plane.
- (31) Dos días después de comprarlas, las entradas bajaron de precio.
Two days after buying them, the tickets went down in price.

The self-paced reading task just described was administered to the L2 learners, and one group of native speakers.

A second group of native speakers took the same self-paced reading task, but a memory burden was added to it. Specifically, before each sentence in the experiment, a 6-digit string was presented for 1500 ms. After the 1500 ms, the participants read the sentence in self-paced reading, and provided the grammaticality judgment. Right after the grammaticality judgment, the same, or a different 6-digit string was presented, and the participants were asked to decide whether it was the same, or different from the string preceding the sentence. Once the participant gave their response, the experiment moved to the following trial. Strings shown after the sentences showed transpositions of two numbers in 50% of the trials, and were identical to the initial string in the other 50%.

6.5. Participants

6.5.1. Native Speakers

A total of 30 native speakers took the same self-paced reading task as L2 learners. This group is referred to as the Native Speaker Control group, and was mostly constituted by undergraduate students from various degree programs at the Universidade de Santiago de Compostela in Spain. Two native speakers were excluded from the analyses due to their low performance in the task. Thus, a total of 28 native speakers were retained as the control group. The mean age in this group was 20.07 years of age.

The native speakers who took the self-paced reading task with the memory burden (N=24) were mostly adults with college degrees, and different occupations at the time of testing. Two native speakers were excluded from analyses due to very low performance on the self-paced reading task. The remaining 22 native speakers were retained for analyses. This group is referred to as the Native Speakers with Burden. The mean age in this group was 27.59.

6.5.2. L2 Learners

A total of 25 L2 learners of Spanish with English as their native language were tested for the study. These learners took the MLA/DELE Spanish proficiency test (Montrul, 2005) after all the computer tasks. This is a pencil and paper, 50-item fill-in-the-blank and multiple-choice test targeting vocabulary and grammar. The test classifies learners as Low proficiency (0-29 points), Intermediate (30-39 points), or Advanced (40-50 points). Only those learners who scored at the Advanced level on the proficiency test, and performed above chance on the grammaticality judgment task were retained for the study.

Thus, a total of 16¹⁵ L2 learners were retained, and included in statistical analyses. The mean score for these learners on the proficiency test was 44.25. They all had a late onset of acquisition for Spanish (after 12 years old), or any other Romance language (mean age of acquisition was 14.69; mean age at the time of testing was 29.69). Most of these L2 learners were recruited from the body of graduate teaching assistants and lecturers in the Department of Spanish and Portuguese at the University of Kansas. Some L2 learners were recruited among contacts of the researcher, and had college degrees in Spanish. All the L2 learners were living in the US at the time of testing.

6.6. Analyses and results

Different sets of analyses were conducted to address each of the three research questions. For each question, analyses were conducted on the reading times (henceforth RTs) of the three critical regions in the target sentences: REGION 5, which corresponds to the Determiner, REGION 6, which corresponds to the Noun, and REGION 7, which corresponds to the complementizer. The determiner and noun in the sentences should agree with the preposed adjective presented in REGION 1. Therefore, any predicted effects may arise in these regions (5 and 6). REGION 7 was also examined in order to capture potential spillover effects at the complementizer. These analyses included only the RTs of the target trials that were correctly judged as grammatical or ungrammatical by the participants. In addition, RTs above or below 2 standard deviations of each

¹⁵ The nine L2 learners excluded from the statistical analyses were eliminated based on their low performance in the conditions with invariant adjectives. Interestingly, these learners showed very high accuracy in the other conditions, likely indicating a specific difficulty with invariant adjectives. This dissertation does not further explore that difficulty, but points to an interesting gap in the agreement literature worthy of study.

participant's individual mean for a specific condition were not included in the analyses¹⁶. All analyses were conducted separately for L2 learners and native speakers, both by participants and by items. The results are reported below as responses to each research question. Statistical significance was set at a *p* value of .05, and values larger than .05 and smaller than, or equal to .10 were considered marginally significant.

6.6.1. RQ1: Are L2 learners sensitive to gender agreement errors in long-distance agreement dependencies?

6.6.1.1. Ungrammatical vs. Invariant

The set of analyses performed to answer this question examined reading times in the ungrammatical violation conditions, using the grammatical conditions with the invariant adjective as a baseline. The conditions examined are repeated below as (32).

(32)

Marked-Unmarked

a. Como es *nueva, he decidido que llevaré el vestido que compré en París.

Invariant-Masculine

b. Como es verde, he decidido que llevaré el vestido que compré en París.

“As it is green/new*_{FemSg}, I have decided that I'll wear the dress_{MSg} I bought in Paris.”

Unmarked-Marked

c. Como es *nuevo, he decidido que llevaré la blusa que compré en París.

Invariant-Feminine

d. Como es verde, he decidido que llevaré la blusa que compré en París.

“As it is green/new*_{MSg}, I have decided that I'll wear the blouse_{FemSg} I bought in Paris.”

¹⁶ As a result of this data trimming, the percentage of data from the critical regions retained in L2 learners was 93.06%. In the case of the control group of native speakers, the percentage was 91.22%. In the burdened group of native speakers the percentage was 86.83%.

A series of 2 x 2 repeated-measures ANOVAs with Grammaticality (grammatical vs. ungrammatical), and Gender of the noun in Region 6 (masculine vs. feminine) as within-subjects factors were conducted on the RTs of the three critical regions (by participants and by items), and on acceptance rates.

The predictions each of the L2 theories tested make for these analyses are the following:

Missing Surface Inflection Hypothesis: This hypothesis predicts a main effect of Grammaticality for both native speakers and L2 learners, with ungrammatical conditions being slower than invariant conditions, reflecting sensitivity to the violations. If markedness plays a role, an interaction between Grammaticality and Gender of the Noun may emerge, reflecting a greater slowdown in the Marked-Unmarked condition (32a) compared to the Invariant-Masculine condition (32b), than in the Unmarked-Marked condition (32c) compared to the Invariant-Feminine condition (32d). That is, if markedness plays a role in the processing of agreement, having a marked feature in the preposed adjective would help detect the ungrammaticality at the critical regions, resulting in a greater slowdown compared to the baseline, than if the preposed adjective is unmarked masculine.

Representational Deficit Hypothesis: if a main effect of Grammaticality were to emerge in L2 learners, it should be driven by a slowdown in the Marked-Unmarked condition (32a) compared to the Invariant-Masculine condition (32b). Thus, the same interaction predicted by the MSIH is predicted by the RDH. In addition, for the RDH, the size of the interaction effect should be related to frequency differences within forms stored in memory, that is, differences in the co-occurrence frequency of feminine noun-

adjective pairs in the target items. In contrast, this relationship with frequency should not emerge for native speakers, nor for masculine noun-adjective pairs in either of the groups. Therefore, this hypothesis will be tested by following-up the potential interaction with correlation analyses between the co-occurrence frequencies of the target feminine noun-adjective pairs, and the effect size of the interaction.

Shallow Structure Hypothesis: No Grammaticality effect, nor an interaction between Grammaticality and Gender of the Noun should emerge in L2 learners. Since the gender violations tested occur in a long-distance dependency, L2 learners should not be sensitive to those errors.

The results are reported below separately for the native speaker controls and the L2 learners.

6.6.1.1.1. Native Controls

Figure 1 below represents the mean RTs in all the regions of the target sentences for the native speaker control group in this analysis.

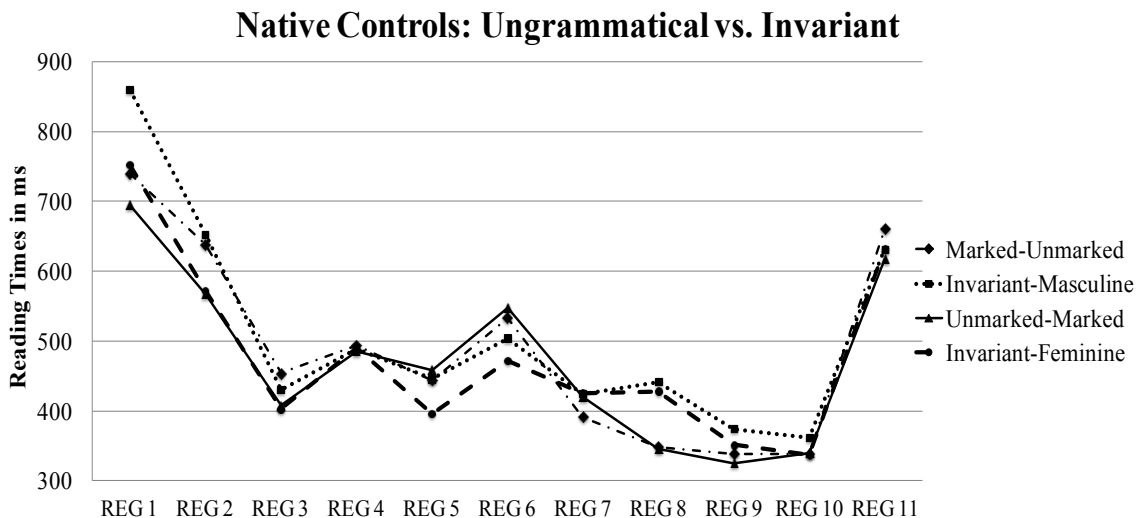


Figure 1. Mean RTs in ungrammatical, and grammatical invariant conditions in the native speaker control group.

Region 5 (determiner): the only effect in this region was a marginal interaction of Grammaticality*Gender of noun ($F_1(1,27) = 2.39, p = .13; F_2(1,35) = 2.87, p = .099$). This interaction was followed up with one-tailed t -tests in order to examine the specific hypothesis that there would be a larger difference between Marked-Unmarked vs. Invariant-Masculine, than between Unmarked-Marked vs. Invariant-Feminine. However, the results showed a somewhat different pattern: Invariant-Feminine was significantly faster than Unmarked-Marked ($t(35) = 1.95, p = .03$), and Marked-Unmarked was marginally faster than Invariant-Masculine ($t(35) = -.11, p = .084$). Thus, the predicted effects did not emerge at the determiner.

Region 6 (noun): the analysis for this region revealed a robust effect of Grammaticality ($F_1(1,27) = 4.58, p < .05; F_2(1,35) = 7.42, p = .01$) at the critical noun. This effect was driven by the ungrammatical conditions being slower than the invariant conditions. Figure 2 illustrates the effect of Grammaticality.

Native Controls Region 6 (Noun)

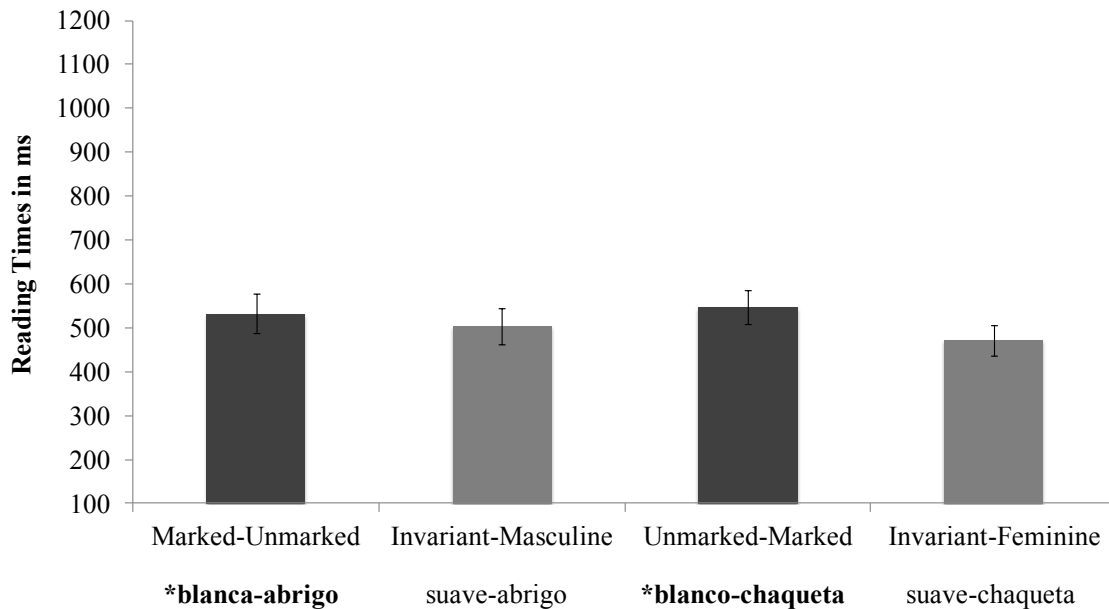


Figure 2. Mean RTs in Region 6 of ungrammatical and invariant conditions in the native speaker control group.

Region 7 (complementizer): No significant effects emerged in this region by participants or by items.

6.6.1.1.1. Summary

The only robust effect in the comparison between ungrammatical conditions and the invariant baseline was an effect of grammaticality in Region 6 (Noun region), which resulted from slower RTs in ungrammatical conditions compared to the invariant conditions. Therefore, as expected, native speakers were sensitive to gender agreement violations in long-distance agreement dependencies. The interaction predicted by the Missing Surface Inflection Hypothesis and the Representational Deficit Hypothesis, however, did not emerge, and natives were equally sensitive to both types of violations in Region 6. Only a marginal interaction (only by items) emerged in Region 5, and it was in an unexpected direction.

6.6.1.1.2. L2 Learners

Figure 3 below represents the mean RTs in all the regions of the target sentences for the L2 learner group in this analysis.

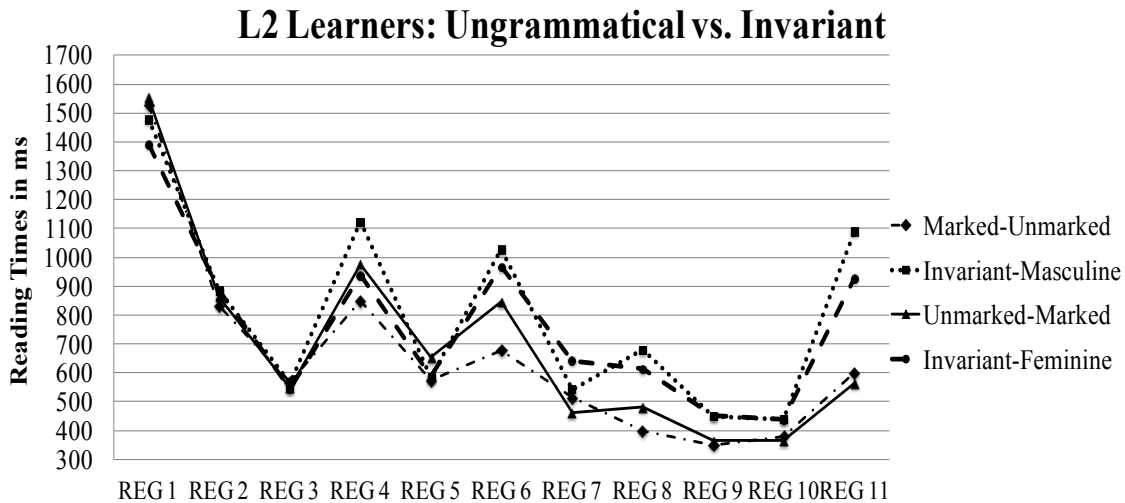


Figure 3. Mean RTs in ungrammatical, and grammatical invariant conditions in the L2 learner group.

Region 5 (determiner): No significant effects emerged in this region.

Region 6 (noun): there was a robust effect of Grammaticality ($F_1(1,15) = 8.41, p = .01$; $F_2(1,35) = 5.89, p < .05$) reflecting slower RTs for invariant conditions compared to ungrammatical conditions, a finding which was not predicted. There was also an interaction of Grammaticality*Gender of Noun, which emerged only in the participant analyses ($F_1(1,15) = 4.76, p < .05$). This interaction was followed-up with a one-tailed t -test that revealed significantly faster RTs in Marked-Unmarked conditions than in Invariant-Masculine conditions ($t(15) = -3.32, p < .01$, but no difference between Unmarked-Marked vs. Inv.-Fem. ($t(15) = -1.41, p = .82$). Figure 4 illustrates these results.

L2 Learners Region 6 (Noun)

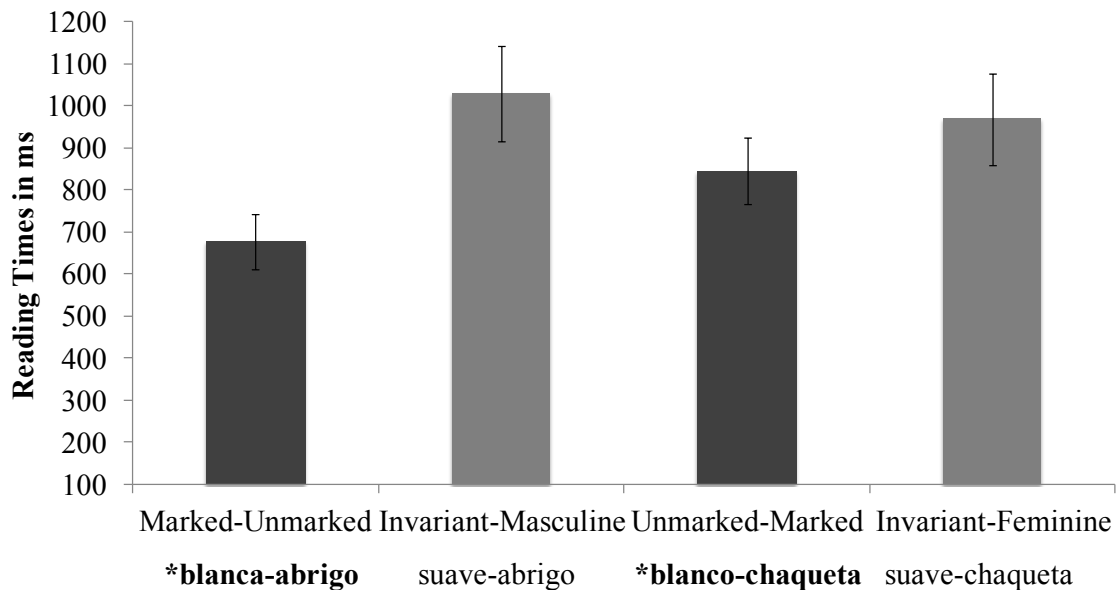


Figure 4. Mean RTs in Region 6 of ungrammatical and invariant conditions in the L2 learner group.

Region 7 (complementizer): the Grammaticality effect which emerged in Region 6 was significant in Region 7 as well ($F_1(1,15) = 6.13, p < .05$; $F_2(1,35) = 3.98, p = .054$).

6.6.1.1.2.1. Summary

The most robust finding for L2 learners when comparing the ungrammatical conditions to the invariant baseline is an effect of Grammaticality. However, this effect was driven by the invariant conditions being significantly slower than the ungrammatical conditions. This result seems to reflect an unexpected difficulty with invariant adjectives in L2 learners that could be obscuring slowdowns in the ungrammatical conditions. Thus, the potential of a Grammaticality effect in L2 learners was followed up with a different analysis, as will be explained below.

6.6.1.2. Ungrammatical vs. Gendered

This set of analyses examines whether learners are sensitive to agreement violations when processing only gendered adjectives, as the inclusion of the invariant adjectives may have obscured sensitivity in the previous analysis. Thus, this analysis compared the ungrammatical conditions to a grammatical gendered baseline instead of the invariant baseline. The conditions examined are repeated below as example (33):

(33)

Marked-Unmarked

a. Como es *nueva, he decidido que llevaré el vestido que compré en París.

Masculine-Masculine

b. Como es nuevo, he decidido que llevaré el vestido que compré en París.

“As it is new_{MSg}/new*_{FemSg}, I have decided that I’ll wear the dress_{MSg} I bought in Paris.”

Unmarked-Marked

c. Como es *nuevo, he decidido que llevaré la blusa que compré en París.

Feminine-Feminine

d. Como es nueva, he decidido que llevaré la blusa que compré en París.

“As it is new_{FemSg}/new*_{MSg}, I have decided that I’ll wear the blouse_{FemSg} I bought in Paris.”

This analysis, however, does not provide an ideal approach to exploring the effects of markedness in a potential interaction between Grammaticality and Gender of the Noun.

This is due to the fact that the adjectives in the grammatical condition with a critical masculine noun are masculine (see 33b), and the adjectives a in the grammatical condition with a critical feminine noun are feminine (see 33d). This means that if the marked feminine adjective results in a facilitation effect at critical regions, the effect would emerge in the feminine baseline, and not in the masculine baseline. Thus, the results reported below mainly focus on whether a main effect of Grammaticality emerges or not for L2 learners.

A series of 2 x 2 repeated-measures ANOVAs with Grammaticality (grammatical vs. ungrammatical), and Gender of the noun in Region 6 (masculine vs. feminine) as within-subjects factors were conducted on the RTs of the three critical regions (by participants and by items), for both L2 learners and the native speaker control group.

The predictions for a Grammaticality effect in L2 learners made by each of the L2 theories tested are the following:

Missing Surface Inflection Hypothesis: a main effect of Grammaticality is predicted for both native speakers and L2 learners: ungrammatical conditions should be slower than grammatical gendered conditions.

Representational Deficit Hypothesis: if a main effect of Grammaticality were to emerge in learners, it should be driven by a difference between the Marked-Unmarked condition, and the Masculine-Masculine condition. This difference should emerge as an interaction, and be tied to frequency differences related to the feminine adjective-masculine noun pairs¹⁷.

Shallow Structure Hypothesis: No Grammaticality effect is predicted for L2 learners, since the gender violations tested occur in a long-distance dependency.

The results are reported below separately for the native speaker control group and the L2 learners.

¹⁷ Although the results in this analysis focus on grammaticality effects, the predictions by the RDH can only be tested by a follow-up frequency analysis on a potential interaction emerging in the direction specified above. Since this interaction did not emerge, the follow-up analysis was not performed.

6.6.1.2.1. Native Controls

Figure 5 represents the mean RTs in all the regions of the target sentences for the native speaker control group in the analysis comparing ungrammatical conditions to grammatical gendered conditions. The results for each region are reported below.

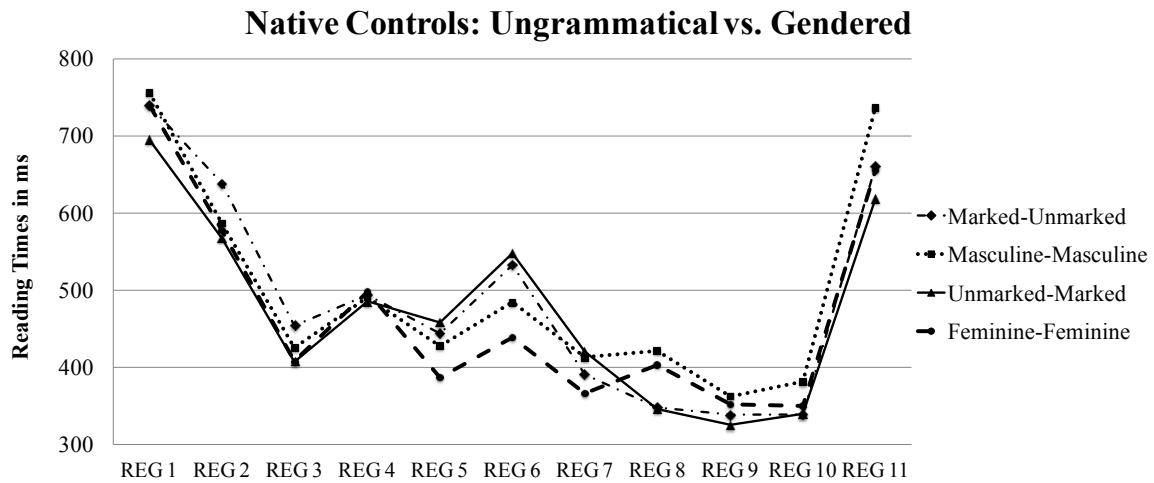


Figure 3. Mean RTs in ungrammatical, and grammatical gendered conditions in the native speaker control group.

Region 5 (determiner): an effect of Grammaticality emerged in this region ($F_1(1,27) = 5.24, p < .05$; $F_2(1,35) = 3.44, p = .07$) indicating slower RTs for ungrammatical conditions than for grammatical conditions. Thus, natives were sensitive to gender violations already at the determiner. Figure 6 illustrates this Grammaticality effect.

Native Controls Region 5 (Determiner)

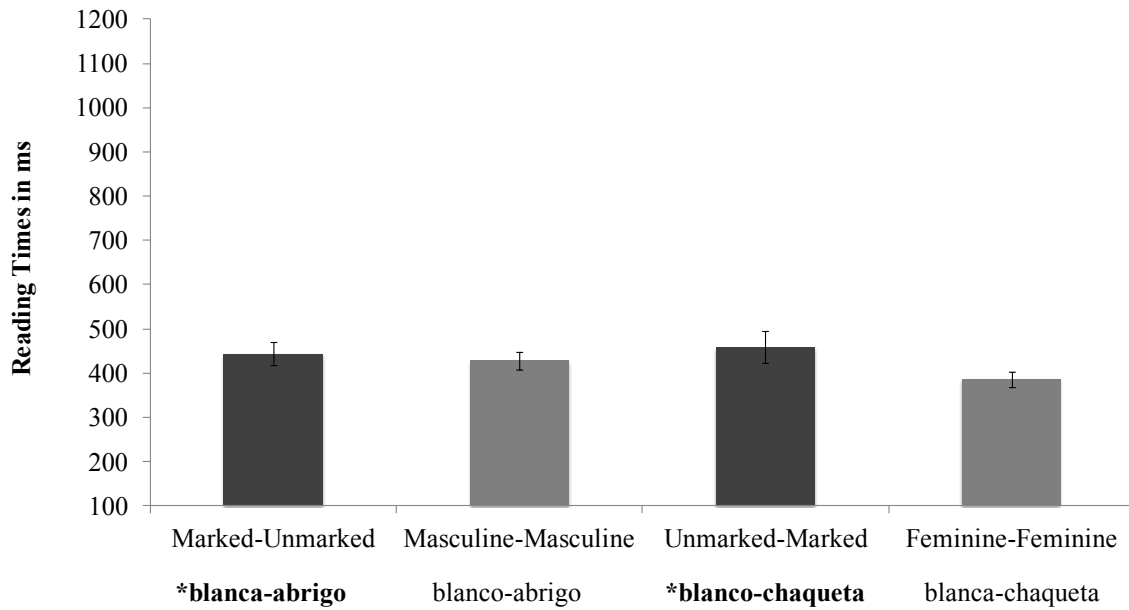


Figure 6. Mean RTs in Region 5 of ungrammatical and grammatical gendered conditions in the native speaker control group.

Region 6 (noun): the same effect of Grammaticality that emerged in Region 5 was significant in Region 6 ($F_1(1,27) = 8.37, p = .01$; $F_2(1,35) = 14.19, p < .01$). Figure 7 illustrates this Grammaticality effect.

Native Controls Region 6 (Noun)

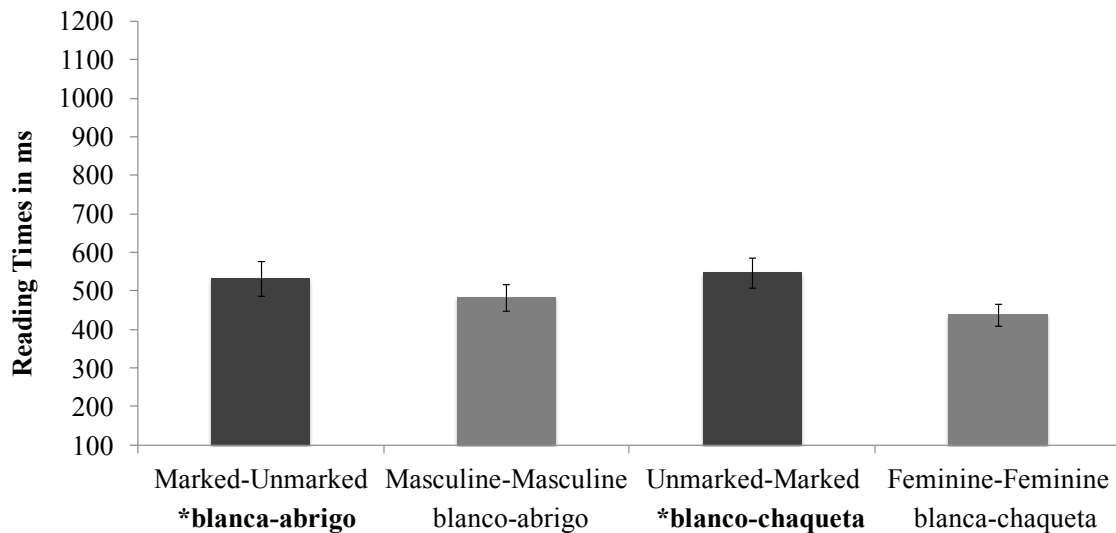


Figure 7. Mean RTs in Region 6 of ungrammatical and grammatical gendered conditions in the native speaker control group.

Region 7 (complementizer): no significant effect of Grammaticality emerged.

6.6.1.2.2. L2 Learners

Figure 8 below represents the mean RTs in all the regions of the target sentences for the L2 learner group in the analysis comparing ungrammatical conditions to grammatical gendered conditions. The results for each region are reported below.

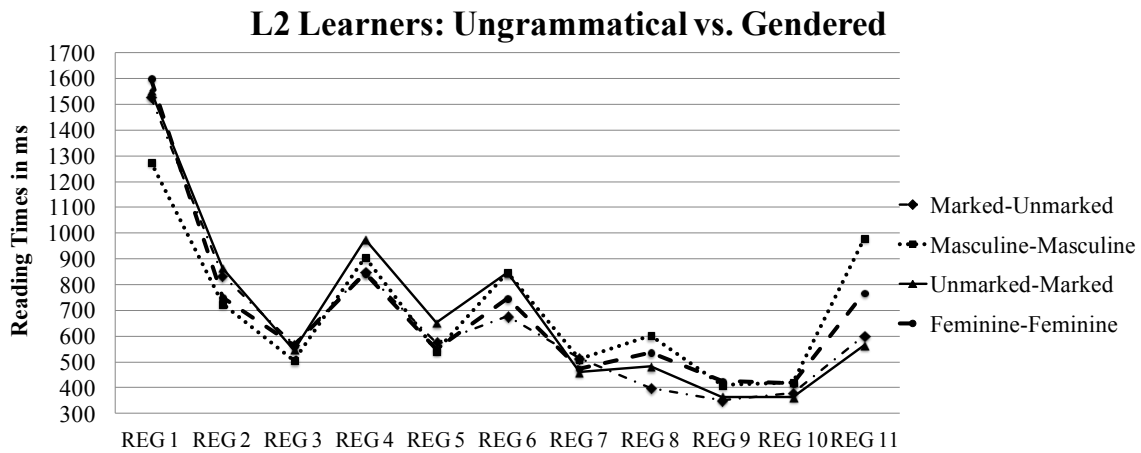


Figure 8. Mean RTs in ungrammatical, and grammatical gendered conditions in the L2 learner group.

Region 5 (determiner): in this region there was a significant effect of Grammaticality ($F_1(1,15) = 3.31, p = .09; F_2(1,35) = 4.74, p < .05$) reflecting slower RTs for ungrammatical conditions compared to grammatical conditions. Thus, this result provides evidence of L2 sensitivity to gender violations already at the determiner. The interaction predicted by the Representational Deficit Hypothesis did not emerge in this region. Figure 9 illustrates the Grammaticality effect.

L2 Learners Region 5 (Determiner)

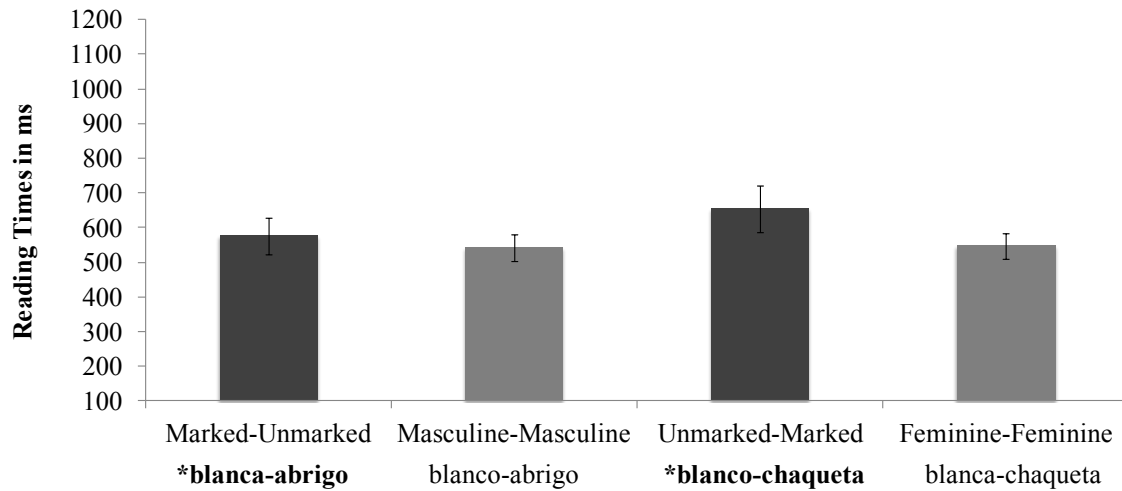


Figure 9. Mean RTs in Region 5 of ungrammatical and grammatical gendered conditions in the L2 learner group.

Region 6 (noun) and 7 (complementizer): no effect of Grammaticality emerged. The interaction predicted by the RDH didn't emerge either.

6.6.1.2.2.1. Summary

Native speakers showed a Grammaticality effect in the expected direction already in Region 5 (Determiner). The effect in this region was significant by participants and marginal by items, and became robust in Region 6 (Noun), where it reached significance in both analyses. L2 learners, who in the analysis with the invariant baseline had shown more difficulty with the invariant conditions than with the ungrammatical conditions, showed a native-like Grammaticality effect in Region 5 (Determiner). The effect was significant by items, and marginal by participants, and similar to native speakers, it reflected slower RTs in ungrammatical conditions, compared to grammatical conditions. This result, contrary to the predictions of the SSH, suggests that L2 learners are sensitive to gender agreement violations in the online processing of long-distance agreement dependencies, and may use gender cues in a predictive manner, as they showed the effect

of Grammaticality at the first critical region (Determiner). In addition, the Grammaticality effect was not driven by a difference between the Marked-Unmarked and the Masculine-Masculine conditions, as predicted by the RDH. The fact that this pattern did not emerge provides some indicative evidence against the RDH.

6.6.2. RQ2: Does the marked status of a feature facilitate processing in L2 learners and native speakers?

6.6.2.1. Grammatical Gendered vs. Grammatical Invariant

The set of analyses performed to answer this question examined reading times in the grammatical conditions only, comparing the gendered conditions with the invariant baseline. The conditions examined are repeated below as (34).

(34)

Masculine-Masculine

a. Como es nuevo, he decidido que llevaré el vestido que compré en París.

Invariant-Masculine

b. Como es verde, he decidido que llevaré el vestido que compré en París.

“As it is new_{M_{Sg}}/green, I have decided that I’ll wear the dress_{M_{Sg}} I bought in Paris.”

Feminine-Feminine

c. Como es nueva, he decidido que llevaré la blusa que compré en París.

Invariant-Feminine

d. Como es verde, he decidido que llevaré la blusa que compré en París.

“As it is new_{F_{emSg}}/green, I have decided that I’ll wear the blouse_{F_{emSg}} I bought in Paris.”

For this analysis a series of 2 x 2 repeated-measures ANOVAs with Adjective Type (gendered vs. invariant), and Gender of the noun in Region 6 (masculine vs. feminine) as within-subjects factors were conducted (by participants, and by items) for each critical region in native speakers and L2 learners.

The predictions for this analysis made by the theories tested are the following:

Missing Surface Inflection Hypothesis: based on the Distributed Morphology framework (Halle & Marantz, 1993; Harley & Ritter, 2002) and Wagers et al.'s (2009) proposal for marked features, the MSIH predicts a an interaction of Adjective Type*Gender of the Noun resulting from a facilitation effect after a marked feature in the Feminine-Feminine condition (34c). This facilitation would yield shorter reading times at critical regions in the Feminine-Feminine condition (34c) compared to the Invariant-Feminine condition (34d), than in the Masculine-Masculine condition (34a) compared to the Invariant-Masculine condition (34b).

Representational Deficit Hypothesis: this theory predicts the same interaction as the MSIH for the ANOVA analysis. However, for the RDH, while those results would be driven by markedness in native speakers, they should be driven by frequency effects in L2 learners. Recall that the RDH argues that for L2 learners, exceptions to the default (thus feminine nouns) are stored in memory with the items with which they co-occur as they are encountered. Thus, depending on the frequency of the stored feminine noun-adjective pairs, there would be more or less of a facilitation effect after the marked adjective. In order to test this hypothesis, an emerging interaction in the predicted direction would be followed up with a correlation analysis between the co-occurrence frequencies of feminine adjective-noun pairs, and a measure of the size of the interaction. This follow-up analysis is explained in detail in a subsequent section.

Shallow Structure Hypothesis: this theory predicts no facilitation effects in L2 learners, since the effects would have to survive throughout the long-distance dependency, something this theory specifically claims not to be possible in L2 learners.

6.6.2.1.1. Native Controls

Figure 10 below represents the mean RTs in all the regions of the target sentences for the native speaker control group in this analysis. The results for each of the critical regions are reported below.

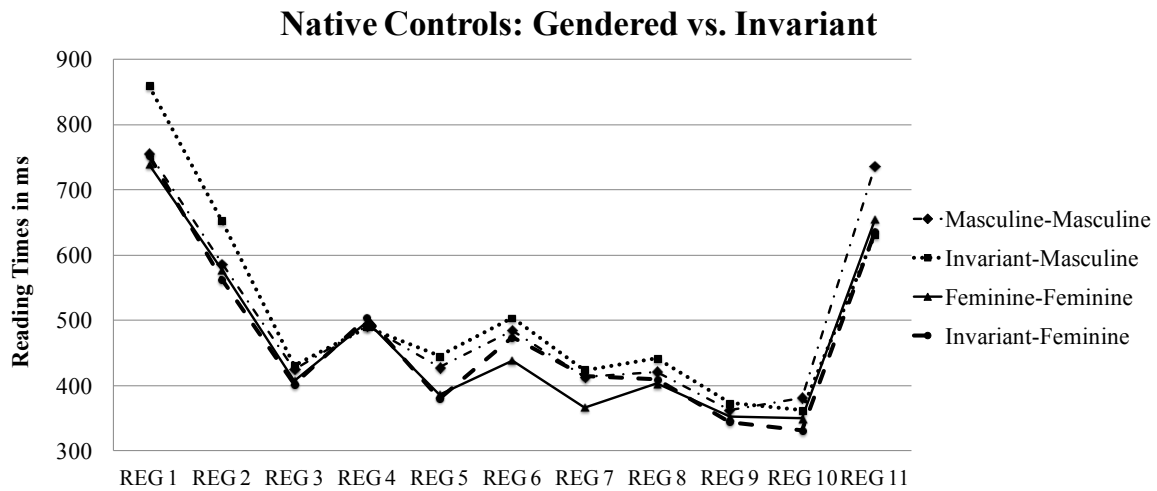


Figure 10. Mean RTs in grammatical gendered and invariant conditions in the native speaker control group.

Region 5 (determiner): in this region there was a main effect of Gender of the Noun ($F_1(1,27) = 10.16, p < .01$; $F_2(1,35) = 5.44, p < .05$) reflecting faster RTs for feminine nouns compared to masculine nouns. Figure 11 illustrates the Gender of the Noun effect.

Native Controls Region 5 (Determiner)

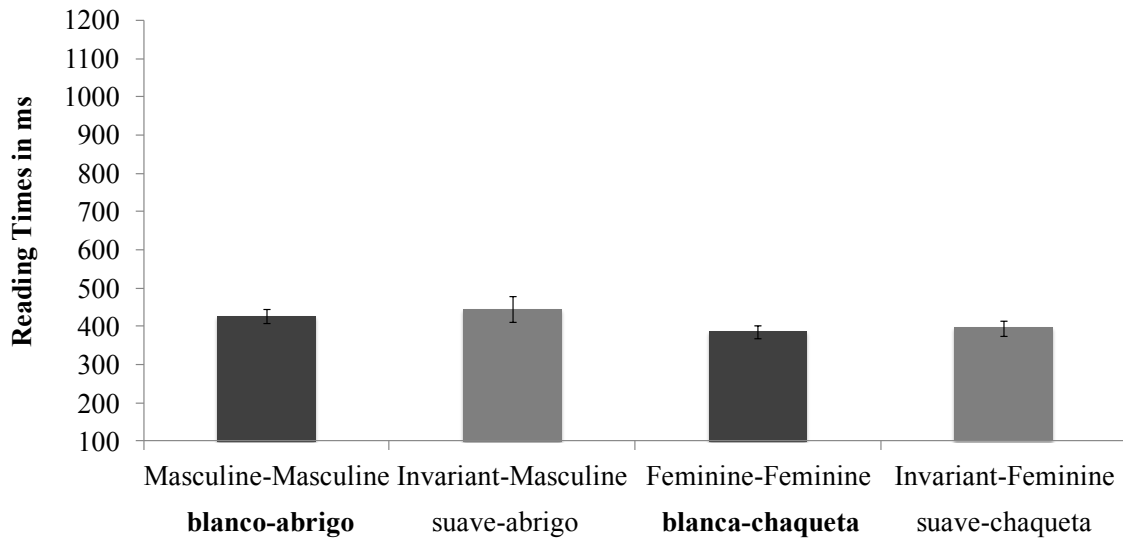


Figure 11. Mean RTs in Region 5 of grammatical gendered and invariant conditions in the native speaker control group.

Region 6 (noun): the same effect of Gender of the Noun as in Region 5 emerged here, as marginal ($F_1(1,27) = 3.31, p = .08$; $F_2(1,35) = 3.87, p = .06$).

Region 7 (complementizer): there was a marginal effect of Gender of the Noun ($F_1(1,27) = 3.16, p = .09$; $F_2(1,35) = 3.11, p = .09$), again reflecting feminine conditions being faster than masculine conditions. There was also a main effect of Adjective Type ($F_1(1,27) = 8.34, p = .01$; $F_2(1,35) = 10.12, p < .01$) reflecting faster RTs for gendered conditions than for invariant conditions. Finally, there was an interaction of Adjective Type*Gender of the Noun ($F_1(1,27) = 7.61, p = .01$; $F_2(1,35) = 6.96, p = .01$). This interaction was followed up with a one-tailed t -test in order to test the hypothesis that the Feminine-Feminine condition would be faster than the Invariant-Feminine condition. The results revealed the predicted pattern ($t_1(27) = -3.69, p < .001$; $t_2(35) = -4.31, p < .001$), and no difference between the Masculine-Masculine condition and the Invariant-

Masculine condition ($t_1(27) = -.66, p = .48; t_2(35) = -.24, p = .18$). Figure 12 illustrates this interaction.

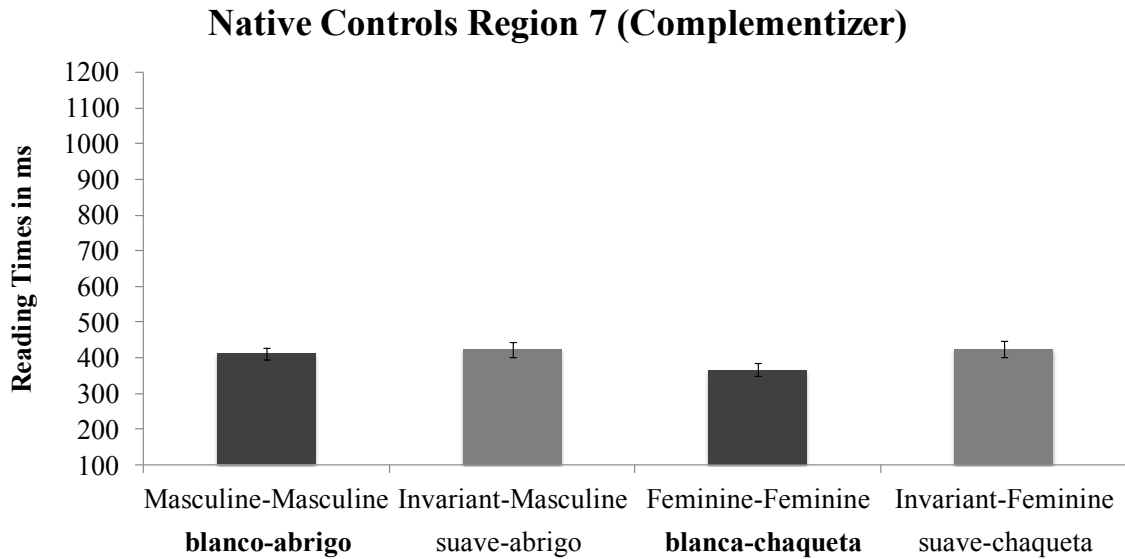


Figure 12. Mean RTs in Region 7 of grammatical gendered and invariant conditions in the native speaker control group.

6.6.2.1.1.1. Summary

In this comparison, in Region 7 (Complementizer) the interaction predicted by the MISH and the RDH for L2 learners emerged, showing a facilitation effect in the Feminine-Feminine condition relative to the Invariant-Feminine condition, and no difference between the Masculine-Masculine condition and the Invariant-Feminine condition. This result supports the asymmetrical representation of features proposed in Distributed Morphology (Halle & Marantz, 1993; Harley & Ritter, 2002). In addition, native speakers showed faster RTs in the conditions with feminine nouns than masculine nouns, in the three critical regions (although the robustness of the effect varied across regions), an effect which was not expected.

6.6.2.1.2. L2 Learners

Figure 13 below represents the mean RTs in all the regions of the target sentences for the L2 learner group in this analysis. The results for each of the critical regions are reported below.

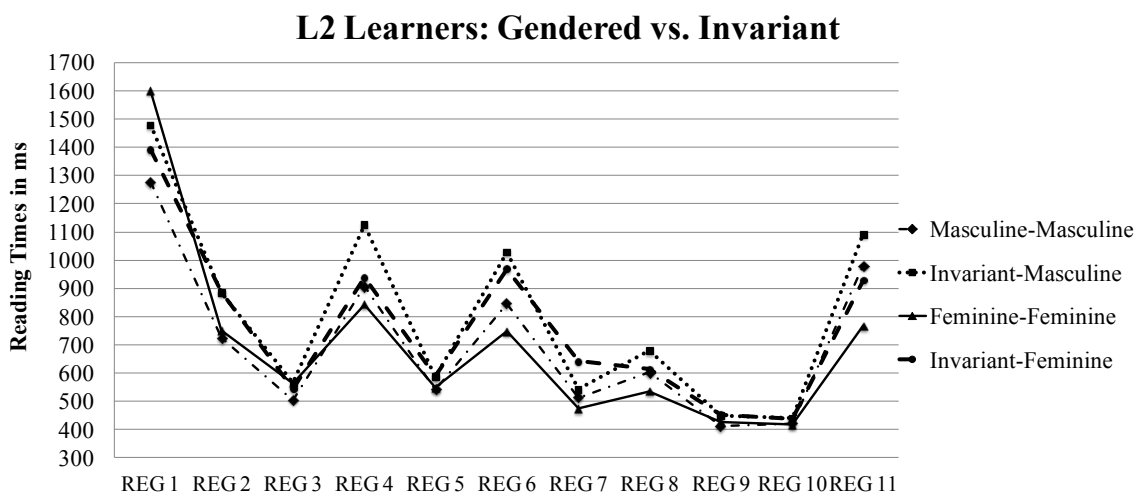


Figure 13. Mean RTs in grammatical gendered and invariant conditions in the L2 learner group.

Region 5 (determiner): in this region there was a marginal effect of Adjective Type ($F_1(1,15) = 3.14, p = .10$; $F_2(1,35) = 3.13, p = .09$), reflecting faster RTs for conditions with gendered adjectives than invariant adjectives.

Region 6 (noun): the Adjective Type effect in Region 5 becomes significant in this region ($F_1(1,15) = 8.60, p = .01$; $F_2(1,35) = 6.95, p = .01$). Figure 14 illustrates this effect.

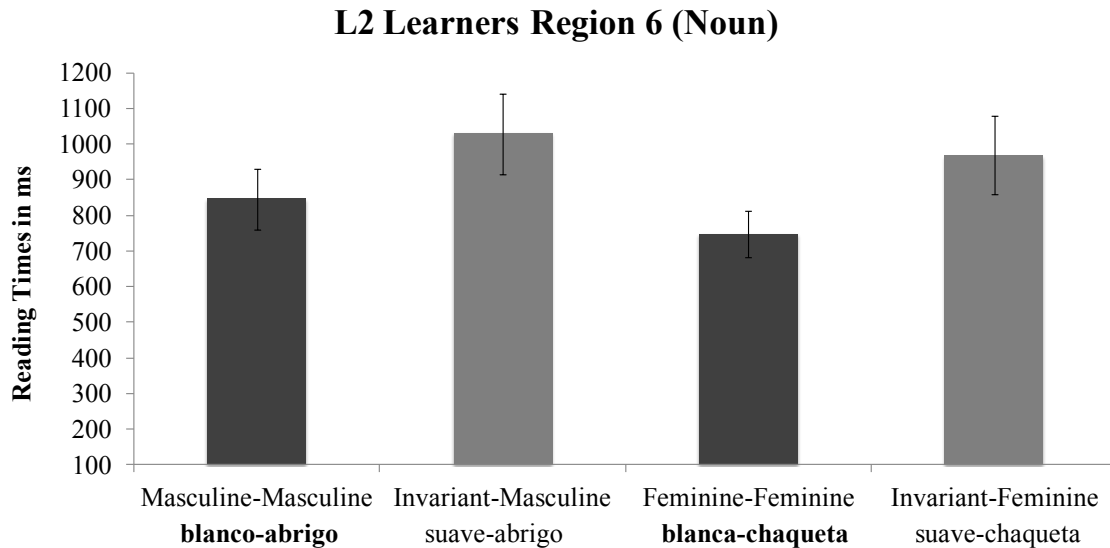


Figure 14. Mean RTs in Region 6 of grammatical gendered and invariant conditions in the L2 learner group.

Region 7 (complementizer): again, there was a significant effect of Adjective Type ($F_1(1,15) = 10.92, p < .01; F_2(1,35) = 12.35, p = .001$) reflecting faster RTs for conditions with gendered adjectives than invariant adjectives. In addition, there was an interaction of Adjective Type*Gender of the Noun that was marginally significant by participants ($F_1(1,15) = 4.40, p = .053$), although not significant by items ($F_2(1,35) = 2.56, p = .12$). This interaction was followed up with a one-tailed t -test in order to test the hypothesis that the Feminine-Feminine condition would be faster than the Invariant-Feminine condition. The results revealed the predicted pattern, both by participants ($t_1(15) = -3.04, p < .01$, and by items ($t_2(35) = -3.20, p < .01$), and no difference between the Masculine-Masculine condition and the Invariant-Masculine condition ($t_1(15) = -1.03, p = .48; t_2(35) = -1.19, p = .76$). Figure 15 illustrates the effects in this region.

L2 Learners Region 7 (Complementizer)

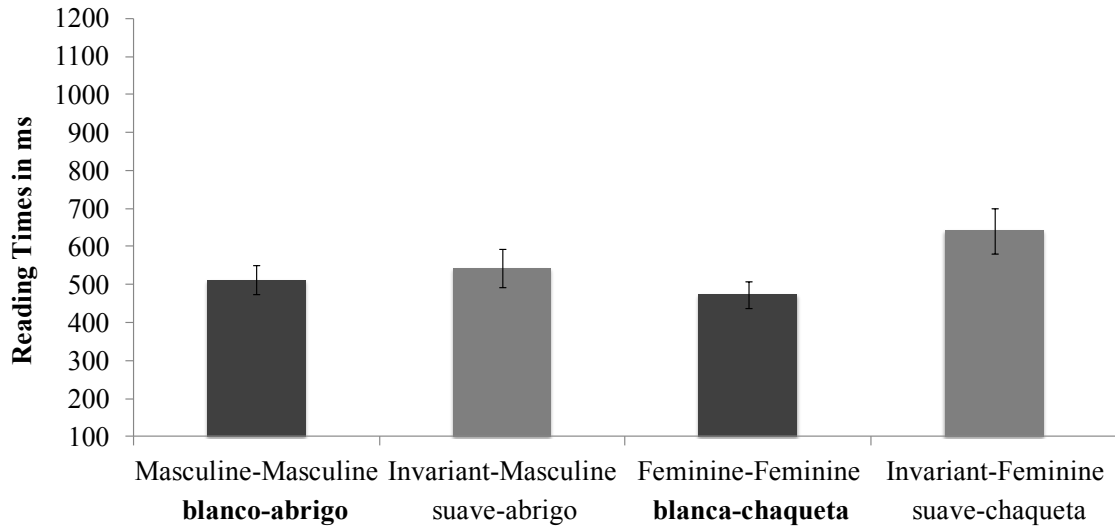


Figure 15. Mean RTs in Region 7 of grammatical gendered and invariant conditions in the L2 learner group.

6.6.2.1.2.1. Summary

In this analysis, L2 learners showed faster RTs in conditions with gendered adjectives compared to invariant adjectives in the three critical regions (with different degrees of robustness). Importantly, the same interaction that emerged in native speakers in Region 7 (Complementizer) also emerged in L2 learners, in Region 7 as well. Although the interaction was weaker in the L2 learner group, it was in the same direction as in the native speaker group, with faster RTs in the Feminine-Feminine condition compared to the Invariant-Feminine condition, and no difference between the Masculine-Masculine condition and the Invariant-Masculine condition. This result is against the Shallow Structure hypothesis, which does not predict any kind of facilitation effect in L2 learners across a long-distance agreement dependency. On the other hand, both the Missing Surface Inflection Hypothesis and the Representational Deficit Hypothesis predict this interaction in L2 learners, although for different reasons. In order to tease apart these two

theories a planned, follow-up correlation analysis was performed as explained below.

6.6.2.1.3. Correlation analysis

The Representational Deficit Hypothesis postulates that in L2 learners, a facilitation effect from the marked feminine feature would be due to feminine forms being stored in memory together with agreeing elements, as exceptions to the masculine default. Thus, the theory would predict that differences in the frequency of co-occurrence of the stored feminine items would result in different facilitation effects; the higher the frequency of co-occurrence of the feminine noun-adjective pairs, the greater the facilitation effect. In order to test this hypothesis a correlation was conducted between the size of the interaction that emerged in Region 7, respectively for L2 learners and native speakers, and the co-occurrence frequencies of the feminine noun-adjective pairs in the target items. The size of the interaction was measured by subtracting the RTs of the Feminine-Feminine conditions from the Invariant-Feminine conditions at Region 7 (separately for L2 learners and native speakers), in order to obtain the size of the facilitation effect for that region. The RDH predicts a positive correlation between the co-occurrence frequencies of the feminine noun-adjective pairs, and the size of the facilitation effect in Region 7, and only in L2 learners, since native speakers don't rely on this compensatory memorization of feminine items. In contrast, the MSIH predicts that if a significant correlation emerges, it should be present in both L2 learners and native speakers.

The results of this analysis revealed no significant correlation, either in the native speaker control group ($r(34) = .18, p = .29$), or in L2 learners ($r(34) = .12, p = .50$)¹⁸.

¹⁸ In order to fully test this hypothesis another correlation was performed for masculine items, by subtracting the RTs of the Masculine-Masculine conditions from the Invariant-Masculine conditions, and

Therefore, against the Representational Deficit Hypothesis, the facilitation effect observed in L2 learners is not indicative of feminine forms being stored in memory as exceptions to the default gender. Instead, the interaction found in both L2 learners and native speakers seems more consistent with Wagers et al.'s (2009) proposal that marked features survive longer than unmarked features in the focus of attention, which would result in the observed facilitation only after a marked feature. In turn, since this effect emerges both in L2 learners and native speakers, the occurrence of the interaction together with the results of the correlation analysis support the Missing Surface Inflection Hypothesis, and Prévost and White's (2000) proposal that L2 learners can acquire gender to native-like levels, and show a mental representation consistent with that proposed in Distributed Morphology for native speakers.

6.6.3. RQ3: Will native speakers performing under a processing burden reveal similar patterns to L2 learners?

In order to answer this question, the same ANOVAs performed for the L2 learners and the native speaker control group were performed on the native speaker group under processing burden. Recall that the task for these native speakers was also self-paced reading with grammaticality judgment on the same items as the other groups, but it had the added burden of remembering a string of numbers presented before each sentence, and deciding whether it was the same or not, as a string of numbers presented after each grammaticality judgment.

correlating this difference with the co-occurrence frequencies of the masculine noun-adjective pairs. This correlation was also conducted separately for native speakers and L2 learners. The results also indicate an absence of correlation both for native speakers ($r(34) = -.06, p = .72$) and L2 learners ($r(34) = .27, p = .11$).

The prediction that computational accounts like the Missing Surface Inflection Hypothesis make for this question is the following: if L2 learner patterns result from processing burden rather than a grammatical deficit, similar patterns may emerge in native speakers performing under processing burden.

In contrast, accounts hypothesizing a grammatical deficit in L2 learners, like the Shallow Structure Hypothesis and the Representational Deficit Hypothesis, predict the following: since L2 learner patterns result from a deficit in their grammar, and native speakers have an intact grammar, native speaker patterns should not resemble L2 learners.

In the following pages the results for each of the analyses conducted in this group are reported.

6.6.3.1. Ungrammatical vs. Invariant

Recall that in this analysis the ungrammatical conditions were compared to the invariant conditions, in order to investigate whether L2 learners are sensitive to gender agreement violations in long-distance dependencies. Figure 16 below illustrates the mean RTs in all the regions of the target sentences for the native speakers under burden in this analysis.

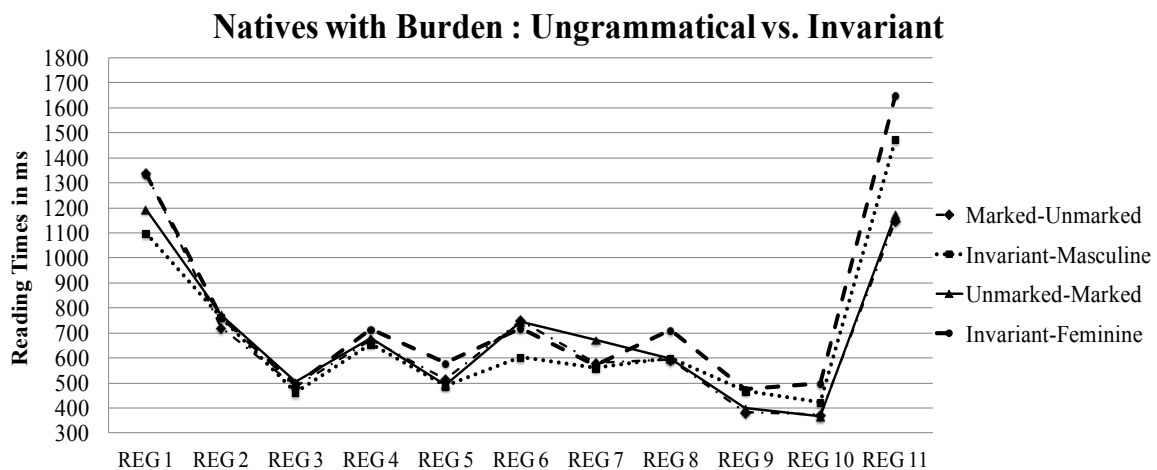


Figure 16. Mean RTs in ungrammatical and invariant conditions in the native speakers performing under processing burden.

The results of this analysis for the native speakers under burden didn't reveal any effects by participants, in any of the critical regions. By items, the only significant effect emerged in Region 5 (Determiner), where there was an interaction of Grammaticality*Gender of the Noun ($F_2(1,35) = 5.27, p < .05$). A two-tailed t -test¹⁹ was conducted to explore the interaction that revealed an interesting result: RTs were slower in the Invariant-Feminine condition than in the Unmarked-Marked condition ($t_2(35) = -2.73, p = .01$), while there was no significant difference between the Marked-Unmarked condition and the Invariant-Masculine condition ($t_2(35) = -.79, p = .43$). Figure 17 below illustrates this interaction.

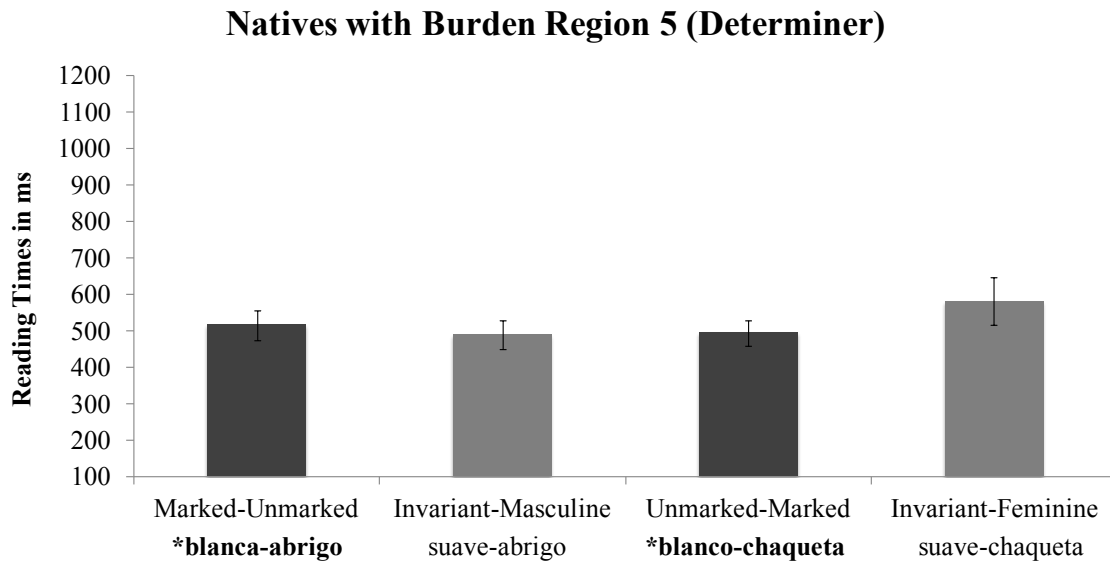


Figure 17. Mean RTs in Region 5 of ungrammatical and invariant conditions in native speakers performing under processing burden.

¹⁹ A two-tailed, rather than a one-tailed t -test, was conducted because the mean RTs of the conditions tested indicated that the interaction that emerged was in an unexpected direction (driven by a difference between the Unmarked-Marked and Invariant-Feminine conditions). Therefore, there was no specific hypothesis to motivate using a one-tailed test.

6.6.3.1.1. Summary

In this analysis, native speakers performing under processing burden, like L2 learners, did not show an effect of grammaticality in any of the regions. The only effect they showed was slower RTs in the Invariant-Feminine condition than in the Unmarked-Marked condition. This result resembles the results for L2 learners in this analysis, in which they showed slower RTs in invariant conditions than in ungrammatical conditions. In this group of native speakers, the effect emerges only by items in Region 5 (Determiner), and only for the Invariant-Feminine condition.

6.6.3.2. Ungrammatical vs. Gendered

Recall that this analysis was performed as a follow-up to the previous analysis, in order to explore the possibility that the slowness of the invariant conditions in L2 learners was obscuring a real grammaticality effect (i.e., ungrammatical conditions slower than grammatical conditions). Thus, in this follow-up, the invariant conditions that served as a baseline in the previous analysis were replaced by grammatical gendered conditions. Figure 18 below the mean RTs in all the regions of the target sentences for the native speakers under burden in this analysis.

Natives with Burden : Ungrammatical vs. Gendered

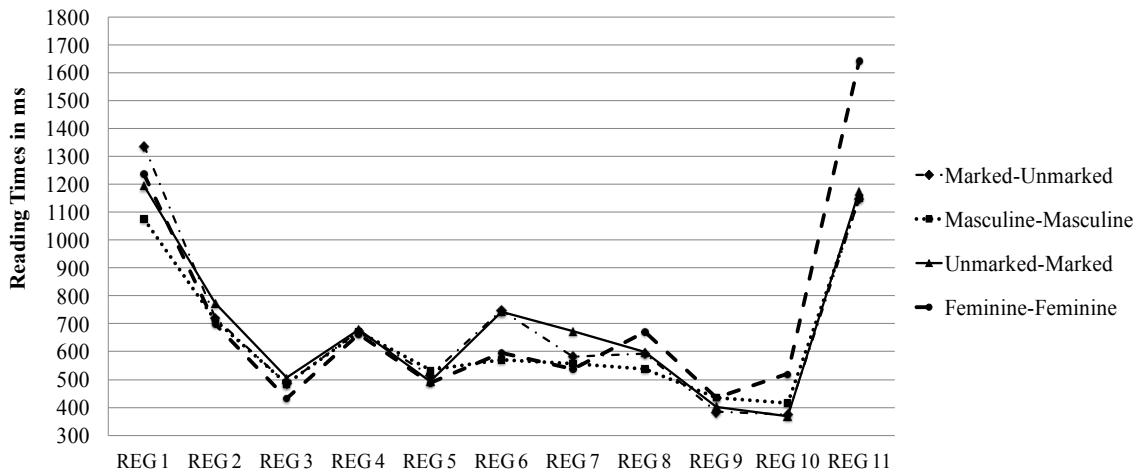


Figure 18. Mean RTs in ungrammatical and grammatical gendered conditions in the native speakers performing under processing burden.

The results for the native speakers performing under processing burden are the following:

Region 5: no Grammaticality effect emerged in this region.

Region 6: in this region there was a robust effect of Grammaticality ($F_1(1,21) = 8.34, p = .01; F_2(1,35) = 9.92, p < .01$), reflecting slower RTs for ungrammatical conditions than for grammatical conditions. Figure 19 illustrates this effect.

Natives with Burden Region 6 (Noun)

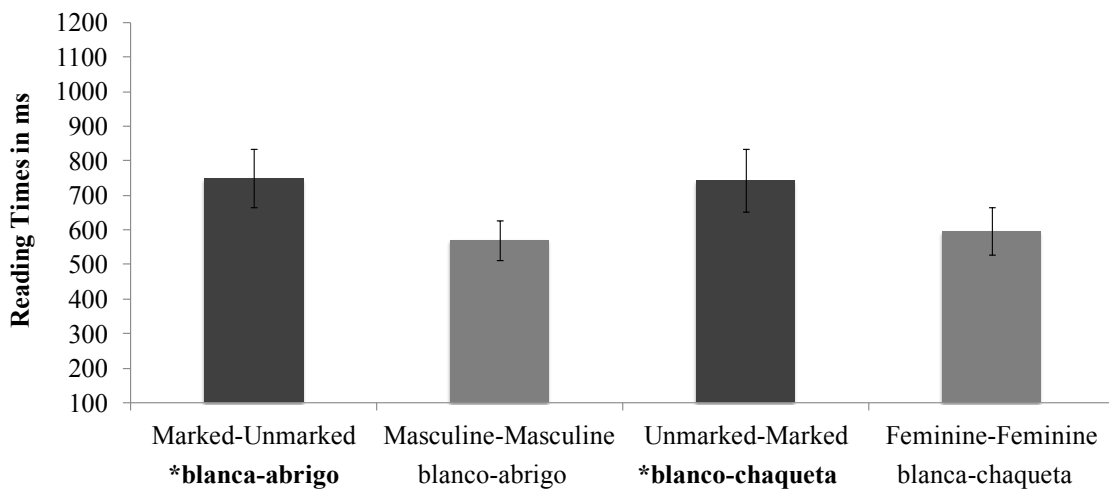


Figure 19. Mean RTs in Region 6 of ungrammatical and grammatical gendered conditions in native speakers performing under processing burden.

Region 7: no Grammaticality effect emerged in this region either.

6.6.3.2.1. Summary

Native speakers under burden did now show an effect of grammaticality. The effect was restricted to one region (the noun region), as it was for L2 learners in this same analysis, while the native speaker control group showed the effect in Regions 5 and 6, although with different strengths. Thus, the processing burden in this native speaker group may have resulted in a more restricted effect of grammaticality.

6.6.3.3. Grammatical Gendered vs. Grammatical Invariant

This analysis focused on the grammatical conditions only, in order to examine whether there is a facilitation of agreement processing effect caused by the presence of a marked feminine feature. Figure 20 illustrates the mean RTs in all the regions of the target sentences in the native speakers performing under burden.

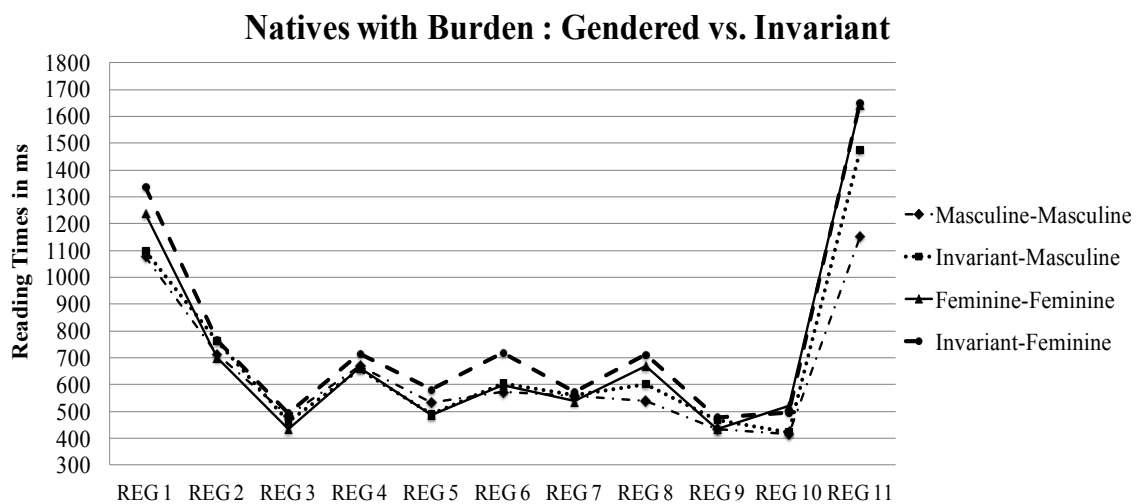


Figure 20. Mean RTs in grammatical gendered and invariant conditions in the native speakers performing under processing burden.

The results for this group in this analysis are the following:

Region 5: there were no effects by participants in this region. By items there was a marginal interaction of Adjective Type*Gender of the Noun ($F_2(1,35) = 3.80, p = .059$). This interaction was followed up with a one-tailed t -test in order to test the hypothesis that the Feminine-Feminine condition would show faster RTs than the Invariant feminine condition, which was confirmed ($t_2(35) = -2.28, p < .05$), and there would be no difference between the Masculine-Masculine condition and the Invariant-Feminine condition, which was also confirmed ($t_2(35) = .55, p = .42$). Figure 21 illustrates this interaction.

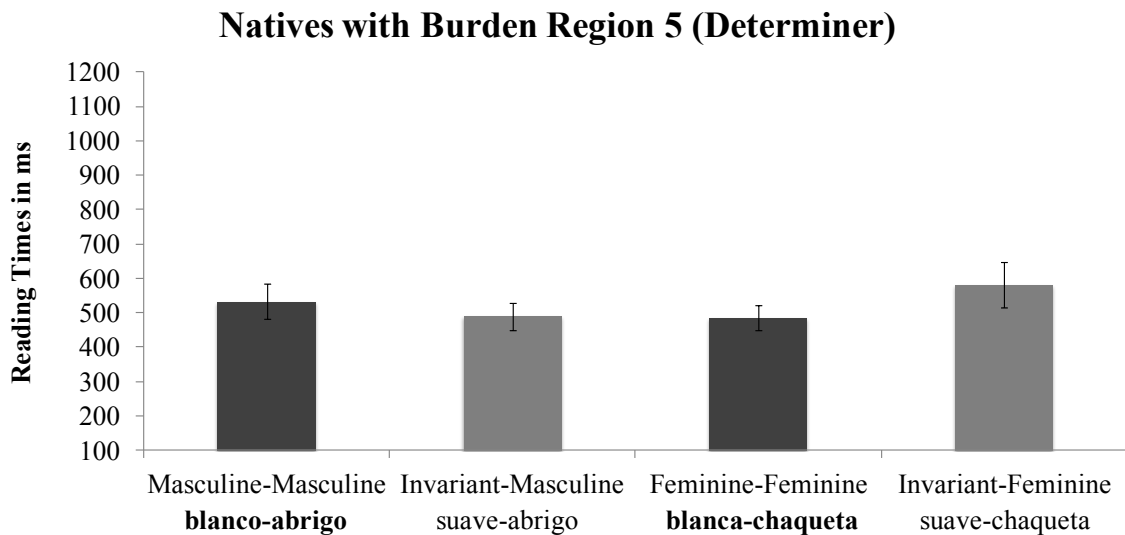


Figure 21. Mean RTs in Region 5 of grammatical gendered and invariant conditions in native speakers performing under processing burden.

Region 6: again, there were no effects by participants. By items, there was a main effect of Adjective Type ($F_2(1,35) = 4.21, p < .05$), indicating faster RTs in conditions with gendered adjectives than invariant adjectives.

Region 7: there were no effects by participants or items in this region.

6.6.3.3.1. Summary

The facilitation effect related to the marked feminine feature emerged for these native speakers in Region 5 (Determiner). Similar to the L2 learners, the facilitation effect was weaker than that found in the native speaker control group, but it was also in the hypothesized direction, with only the Feminine-Feminine condition showing a facilitation effect compared to the corresponding invariant condition. In addition, in Region 6 (Noun), conditions with gendered adjectives were read faster than conditions with invariant adjectives (by items), an effect that emerged both in L2 learners and the native speaker control group.

Overall, the results for the native speakers performing under processing burden were not robust. However, they showed qualitative similarity with L2 learners across analyses, providing some tentative support for the Computational accounts tested.

6.6.4. Acceptance Rates

Although the focus of the present study is on the online processing of gender agreement, the acceptance rates on the offline grammaticality judgments at the end of the target sentences were also examined. The same participants analyses performed on RTs were performed in these acceptance rates, separately for L2 learners and each of the native speaker groups. For the native speaker control group the results are the following:

In the Ungrammatical vs. Invariant comparison there was a main effect of Grammaticality, with invariant conditions having higher acceptance rates than ungrammatical conditions ($F(1,27) = 1254.99, p < .01$), as expected.

In the Ungrammatical vs. Gendered comparison there was also a main effect of Grammaticality reflecting higher acceptance of grammatical gendered conditions than

ungrammatical conditions ($F(1,27) = 1255.69, p < .01$). Neither the Ungrammatical vs. Invariant analysis, nor the Ungrammatical vs. Gendered analysis revealed any interactions. Thus, the native speaker controls were equally accurate in their distinction on grammatical versus ungrammatical sentences, regardless of the markedness of the initial adjective.

Finally, in the Grammatical Gendered vs. Grammatical Invariant comparison there was a marginal effect of Adjective Type indicating marginally higher acceptance of gendered conditions than invariant conditions ($F(1,27) = 3.53, p = .07$). Although weak, this is an interesting effect that could be related to the unexpected difficulty that L2 learners showed with invariant adjectives in their online processing.

For the L2 learners, the results are the following:

In the Ungrammatical vs. Invariant comparison there was a main effect of Grammaticality, with invariant conditions having higher acceptance rates than ungrammatical conditions ($F(1,15) = 720.43, p < .01$). This result shows robust sensitivity to the gender agreement violations in long-distance dependencies in L2 learners. This effect did not emerge in the RTs analysis, where L2 learners were slower reading the invariant conditions compared to the ungrammatical conditions. We can see here that learners were sensitive to the violations in their offline judgments.

In the Ungrammatical vs. Gendered comparison there was also a main effect of Grammaticality reflecting higher acceptance of grammatical gendered conditions than ungrammatical conditions ($F(1,15) = 496.20, p < .01$). Again, this result shows that learners were sensitive to the agreement violations in their offline judgments, and in this case the results parallel those obtained from the RTs analysis. Neither the Ungrammatical

vs. Invariant analysis, nor the Ungrammatical vs. Gendered analysis revealed any interactions. Thus, the L2 learners, like the native speaker controls, were equally accurate in their distinction on grammatical versus ungrammatical sentences, regardless of the markedness of the initial adjective.

In the Grammatical Gendered vs. Grammatical Invariant comparison there were no effects in this group. A weak facilitation effect driven by the marked feature emerged in the RTs analysis, which may have been too subtle to be reflected in acceptance rates.

Finally, the results for the native speaker group performing under processing burden are the following:

In the Ungrammatical vs. Invariant comparison there was a main effect of Grammaticality, with invariant conditions having higher acceptance rates than ungrammatical conditions ($F(1,21) = 275.47, p < .01$). In addition, there was an interaction of Grammaticality*Gender of Noun ($F(1,21) = 5.22, p < .05$), that was followed up with a one-tailed t -test, in order to test the hypothesis that there would be a larger difference between the Marked-Unmarked and the Invariant-Masculine conditions than between the Unmarked-Marked and Invariant-Feminine conditions. This was the pattern that emerged: the Marked-Unmarked condition had a significantly lower acceptance than the Invariant-Masculine condition ($t(21) = -17.86, p = 0$), and this difference was greater than for the other comparison, although the Unmarked-Marked condition had also a significantly lower acceptance than the Invariant-Feminine condition ($t(21) = -10.86, p = 1$), as would be expected. This interaction is consistent with the marked feature facilitating the detection of the agreement violation, compared to having

an unmarked feature in the adjective, in line with Wagers et al.'s (2009) results for number. This effect didn't emerge in the RTs analysis, nevertheless.

In the Ungrammatical vs. Gendered comparison there was also a main effect of Grammaticality reflecting higher acceptance of grammatical gendered conditions than ungrammatical conditions ($F(1,21) = 353.04, p < .01$). In addition, there was also an effect of Gender of the Noun ($F(1,21) = 4.67, p < .05$), which indicates higher acceptance of conditions with feminine nouns, compared to conditions with masculine nouns. This effect seems to be driven by the difference between Marked-Unmarked and Unmarked-Marked, although the effect was not strong enough to result in an interaction of Grammaticality*Gender of the Noun.

Finally, in this group no effects emerged in the Grammatical Gendered vs. Grammatical Invariant comparison.

6.6.4.1. Summary

The acceptance rates analyses in the native speaker control group revealed similar results to the analyses on RTs, in the comparisons involving agreement violations: this group showed robust sensitivity to gender agreement violations, and this sensitivity did not interact with the markedness of the initial adjectives. In the comparison including only grammatical sentences no effects emerged.

In the L2 learner group the acceptance rates results from the ungrammatical comparisons were interesting in that both comparisons (whether the baseline was the invariant or the gendered) revealed robust sensitivity to the violations. This suggests that the difficulty with invariant adjectives evident in the RTs analyses had disappeared by the time the learners provided the grammaticality judgment, thus allowing the effect of

sensitivity to the agreement errors to emerge. Like the native speakers, L2 learners were equally sensitive to all violation types, regardless of the markedness of the adjectives. In the grammatical comparison, no effects emerged.

Finally, the native speaker group performing under processing burden also showed robust sensitivity to the agreement violations in both ungrammatical comparisons. In addition, an interaction emerged in the comparison involving the invariant baseline that was compatible with the marked feature proposal by Wagers et al. (2009). Specifically, the violation condition with a marked feminine adjective had significantly lower acceptance rates, suggesting that the marked feature could have helped the detection of the violation. In the ungrammatical comparison with the gendered baseline the results also revealed higher acceptance of conditions with feminine nouns than masculine nouns. Finally, like the other groups, these native speakers didn't show any effects in the grammatical comparison.

6.6.5. Gender assignment task

The gender assignment task tested the learners' knowledge of the gender of the critical nouns in the target sentences. The mean accuracy of the learners selecting the determiner with the correct gender for each noun was 99.48%, the lowest accuracy rate in the group being 97.22%. These results are significant, given that this was a timed task. Thus, the L2 learners knew the target nouns and their gender very well.

7. Discussion

The present study set out to investigate three research questions with respect to the online processing of gender agreement in L2 learners and native speakers of Spanish. In order to

answer these questions, a self-paced reading study was conducted testing advanced L2 learners of Spanish and a control group of native speakers, as well as a group of native speakers performing the same task under processing burden. The target sentences tested gender agreement between a preposed adjective and a noun that were in a long-distance agreement dependency (across a CP). In these sentences, the grammaticality of the agreement dependency, the markedness status of the features involved, and the gender information provided by the preposed adjective were manipulated. The findings of the study are discussed below in response to each of the research questions investigated.

7.1. L2 sensitivity to gender agreement violations in long-distance dependencies

The first research question asked whether L2 learners are sensitive to gender agreement violations occurring across different syntactic phrases. This question was investigated by comparing RTs at critical regions in the conditions involving ungrammatical gender agreement between the preposed adjective and the noun, and the baseline conditions, in which the adjective was in its invariant form, therefore not providing any information on the gender of the upcoming noun. While the results for this analysis revealed a clear slowdown in ungrammatical conditions compared to the baseline in the native speaker control group, L2 learners showed slower RTs in the invariant baseline than the ungrammatical conditions. This result in learners seems to indicate some difficulty with invariant adjectives that nonetheless, learners seem to overcome by the end of the sentence, given that the invariant conditions were highly accepted as grammatical by this group, and there was a significant difference in acceptance rates compared to the ungrammatical conditions. Whatever the explanation for this unexpected difficulty with

invariant adjectives²⁰, it is possible that the learners did detect the violations in the ungrammatical conditions in their online processing as well, but the invariant conditions were read so slowly that the RTs masked the violation detection effect. In order to test this hypothesis, a second analysis was conducted in which the ungrammatical conditions were compared to grammatical conditions with gendered adjectives showing canonical endings. The results for this analysis confirmed the stated hypothesis, showing that both native speakers and L2 learners were sensitive to the gender agreement violations, with slower RTs in the ungrammatical conditions than the grammatical conditions. Even though the effect in L2 learners was somewhat weaker than in native speakers (marginal by participants, significant by items), this is a significant finding, given the fact that the effect emerged in the first critical region (Determiner region). The immediacy of this effect suggests that L2 learners used the gender cues on the adjectives to actively predict the gender of an upcoming noun. When the expected gender was not found in the determiner accompanying that noun, a slowdown in processing emerged. Thus, this finding not only indicates that L2 learners are in fact sensitive to gender agreement

²⁰ Due to the lack of studies examining the processing of invariant adjectives, we can only speculate about the source of this difficulty. One possibility is that it stems from the fact that invariant adjectives show non-canonical endings (*-e* or consonant), rather than the usual *-o/-a* gendered endings, making them harder/slower to process. Another related possibility is that given that many Spanish masculine nouns end in *-e*, the learners associate the *-e* ending in general with masculine gender. If this is the case, since this ending was present in eight out of the twelve invariant adjectives used, then learners would have to suppress this association when reaching feminine nouns in order to accurately judge the sentence as grammatical, thus resulting in longer RTs. However, this explanation would only account for the slowdown in the Invariant-Feminine condition, while both the Invariant-Feminine, and the Invariant-Masculine conditions were slower than the ungrammatical conditions. A third possibility is that since the invariant adjectives don't provide gender information about the upcoming noun, it takes the learners longer to integrate the gender information at the noun, compared to detecting the ungrammaticality of the conditions with the gender violation. This could be specifically related to the fact that participants were asked to provide a grammaticality judgment at the end of each sentence. It is possible that this task promoted greater focus on the form of the adjectives, thus making it easier to detect the ungrammaticality of the violations, in which the adjective was gendered, as compared to the conditions with the invariant adjectives, where no gender information was provided.

violations across long-distance dependencies, against the Shallow Structure Hypothesis (Clahsen & Felser, 2006; Clahsen et al., 2010), but it also suggests that learners engage in predictive processing in agreement dependencies. This result contrasts with Grüter et al.'s (2012) results, in that learners in that study only showed evidence of predictive processing with novel nouns. The learners in the current study clearly knew the nouns in the target sentences and their gender, as shown by the results of the gender assignment task, as well as the acceptance rates obtained from the grammaticality judgments. Nevertheless, while Grüter et al.'s learners were tested on grammatical sentences, the learners in this study showed the predictive effect in gender violations, the processing of which may not be as subtle. In addition, the current results may support Grüter et al.'s lexical learning hypothesis, if we take into account the characteristics of the learners tested in the current study. Recall that most of these learners were working towards graduate degrees in Spanish at the time of testing, and had extensive experience teaching Spanish at different levels. Therefore, it is possible that the group of learners tested here were able to develop strong gender nodes, and thus use gender information in a predictive manner, similar to the learners tested by Dussias et al. (2013), and Hopp (2013).

The emergence of a predictive effect of grammaticality in the current study is also inconsistent with the Representational Deficit Hypothesis (Hawkins, 2009), which posits a deficit in the L2 grammar for features not present in the L1. Specifically, for this theory English-speaking learners of Spanish would only show sensitivity to gender violations thanks to the use of default morphology combined with a frequency-based memorization strategy. This strategy should have emerged in the current results as an interaction driven by a greater slowdown in the ungrammatical condition where the

adjective is in its marked feminine form (e.g. **blanca-abrigo*). Furthermore, the slowdown in the interaction should be related to the frequency of the feminine forms stored in memory. Nevertheless, this interaction did not emerge in our results, and learners were equally sensitive to gender violations involving the marked form of the adjective (e.g. **blanca-abrigo*), and violations involving the unmarked form of the adjective e.g. (e.g. **blanco-blusa*). Thus, neither the Shallow Structure Hypothesis, nor the Representational Deficit Hypothesis can explain the results obtained here for L2 learners. In contrast, the Missing Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 2000) argues that L2 learners can show native-like performance, even with features not present in the L1, which is the pattern that emerged in the results. First, the L2 learners, like the native controls, showed sensitivity to long-distance gender violations, and the sensitivity was evident as early as the determiner region. Thus, the current study adds to the studies providing evidence of online processing of long-distance agreement dependencies (e.g. Alemán-Bañón et al., 2014; Gabriele et al., 2013), as well as predictive processing (e.g. Dussias, 2013; Hopp, 2013) in L2 learners. Second, the Missing Surface Inflection Hypothesis predicted the same interaction as the Representational Deficit Hypothesis, but in both native speakers and L2 learners, if markedness played a role in the processing of the gender violations. The interaction emerged only in the offline grammaticality judgments of the native speakers under processing burden, which may suggest the need for some decline in performance for this effect to emerge. In the case of the L2 learners, we may speculate that the self-paced reading task wasn't demanding enough for them for the interaction to emerge, since the learners were very advanced. Importantly, this interaction didn't emerge in the native

speaker control group either. Given the similarity between the L2 learners and the native controls in this respect, the absence of the interaction in these two groups provides support for the Missing Surface Inflection Hypothesis, rather evidence against it. The absence of the interaction may suggest that markedness does not play a role in the online processing of ungrammatical sentences, like those tested in order to answer the first research question. However, it is also possible that even if markedness plays a role, the ungrammaticality of those sentences obscured the markedness effects. Thus, a better scenario to test the role of markedness in the online processing of agreement is found in grammatical sentences, as those tested to answer the second research question.

7.2. Use of gender information in online processing

The second research question asked whether the marked status of a feature facilitates processing in L2 learners and native speakers. It has been argued for native speakers that marked features stay longer in the focus of attention than unmarked features (Wagers et al., 2009). Thus, in an agreement dependency, feature-checking at the critical region would happen faster when the feature involved is marked. This hypothesis is consistent with the asymmetrical representation of features proposed in Distributed Morphology for native speakers (Halle & Marantz, 1993; Harley & Ritter, 2002). For L2 learners, it has been proposed that the asymmetrical patterns observed in their agreement errors may actually reflect the same asymmetrical representation of features proposed for native speakers (Prévost and White, 2000). According to the Missing Surface Inflection Hypothesis, these errors emerge in learners (and not in native speakers) because the computational burden of processing the L2 makes it difficult to access target forms during online processing, resulting sometimes in agreement errors. Thus, the Missing

Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 2000) argues that even though these errors may emerge in L2 learners, their mental representation of features may be native-like. This hypothesis was tested by examining whether both L2 learners and native speakers show evidence of facilitation of agreement processing after a marked feature. Specifically, RTs at the critical regions in grammatical conditions with gendered adjectives were compared to RTs at the same regions in grammatical conditions with invariant adjectives. The results showed that both L2 learners and native speakers read the gendered conditions faster than the invariant conditions. This result indicates that in both groups the gender information provided by the gendered adjective made feature-checking at the critical determiner phrase more efficient than when the adjective provided no gender information about the upcoming noun. Crucially, in both groups this effect was driven by an interaction in the complementizer region. Although the interaction was weak in L2 learners, a planned follow-up revealed that in both groups there was a significant facilitation resulting from having a marked feminine feature in the preposed adjective, compared to having no gender information in the invariant counterpart. In both groups as well, having the unmarked masculine feature in the adjective did not result in facilitation, compared to having no gender information in the invariant counterpart. Thus, although the pattern was not as robust in L2 learners, it was in the same direction, and in the same region as in native speakers. Also importantly, the marked feature facilitation effect in the current study emerges in a comparison involving only grammatical sentences rather than the usual paradigm of violations, thus reflecting more natural grammatical processing.

The facilitation effect in the native speaker controls is consistent with the numerous attraction studies showing asymmetries between singular and plural features, and also with Akhutina et al.'s (1999) results finding a facilitation effect for the marked feminine feature in Russian in speech shadowing. Importantly, the present findings are in line with Wagers et al.'s (2009) results for the maintenance of singular and plural features in the focus of attention. Thus, the current finding supports Wagers et al.'s proposal that marked features stay longer in the focus of attention than unmarked features. In addition, it provides evidence of the applicability of the proposal to the processing of gender features in grammatical sentences, and importantly, in L2 learners as well. To the author's knowledge, the current study is the first to show a markedness-related facilitation effect in both native speakers and L2 learners in grammatical sentences, across a long-distance dependency. While Dussias et al. (2013) and Hopp (2013) showed evidence of predictive processing in grammatical sentences in advanced learners and native speakers, no markedness effects were found in these groups. Nevertheless, a lower proficiency group of Italian learners in Dussias et al.'s study showed anticipatory looks only in the case of the marked feature. This result is consistent with the findings in the present study, although Dussias et al. did not interpret it as a markedness effect. It is an open question, however, why the advanced learners and native speakers in that study didn't show the same effect.

In terms of the theories tested, the facilitation effect found for the marked feature in L2 learners speaks against the Shallow Structure Hypothesis, which does not predict any effects for L2 learners in a long-distance dependency. Instead, this finding supports the Missing Surface Inflection Hypothesis, but also the Representational Deficit

Hypothesis (Hawkins, 2009). Both theories make claims consistent with the facilitation effect found in L2 learners, although for very different reasons. The Missing Surface Inflection Hypothesis would argue that the facilitation effect in learners reflects a native-like representation of the gender feature. In contrast, the Representational Deficit Hypothesis makes the specific claim that while such facilitation in native speakers is related to markedness, in L2 learners it reflects a frequency-based strategy to compensate for a flawed representation of the gender feature. Specifically, the theory argues that in learners feminine forms are memorized with agreeing elements as they are encountered, as exceptions to the masculine default. Thus, when those feminine forms are accessed from memory for agreement purposes, the higher frequency noun-adjective pairs result in a facilitation effect, due to their faster access. Therefore, the prediction that the Representational Deficit Hypothesis makes for learners is that the amount of facilitation found for feminine noun-adjective pairs should be correlated with their frequency of co-occurrence. For native speakers, however, the theory argues that the facilitation effect is related to markedness, and therefore, should not be related to frequency. This hypothesis was tested in correlation analyses conducted separately for masculine and feminine items, in both native speakers and L2 learners. The results of these analyses revealed no significant correlations with frequency for either masculine or feminine items, in either of the groups. Thus, contrary to the Representational Deficit Hypothesis, the facilitation effect found in L2 learners is not linked to frequency. Overall, these results suggest that markedness plays a similar role in gender agreement in native speakers and L2 learners, and therefore are more consistent with the Missing Surface Inflection Hypothesis. In turn, these results support Prévost & White's (2000) theory relating morphological variability

in L2 learners to an asymmetrical representation of features as the one hypothesized in Distributed Morphology (Halle & Marantz, 1993; Harley & Ritter, 2002) for native speakers. Nevertheless, given the weakness of the effect found in learners, this issue should be revisited to search for converging evidence testing a larger sample of learners.

7.3. Native Speakers performing under processing burden

The third research question asked whether native speakers performing under processing burden would reveal similar patterns to L2 learners. This question builds on previous research showing similarities between the two groups (Hopp, 2010; López Prego & Gabriele, 2014), and supporting computational accounts of morphological variability. Under this logic, if learners' representations of features are target-like, and their variability in agreement stems from computational issues, then native speakers performing a task under a computational burden may show similar variability. This question was investigated by testing a native speaker group performing the same task as the native control group and the L2 learners, with an added memory load. The results for this group were not very robust, suggesting that the task may have been too taxing for the subtle differences investigated to emerge more strongly. Nevertheless, qualitatively, these results show some resemblance with the learner results. Specifically, like the L2 learners, this group of native speakers did not show an effect of grammaticality in the Ungrammatical vs. Invariant comparison. Rather, they showed a slowdown in the Invariant-Feminine condition, which interestingly, resembles the slowdown L2 learners showed for both invariant conditions. In addition, the offline grammaticality judgments of this group in this comparison showed an interesting interaction that is predicted under an asymmetrical representation of features. Specifically, the difference in acceptance

rates between the Marked-Unmarked and the Invariant-Masculine conditions was greater than between the Unmarked-Marked and the Invariant-Feminine conditions. This effect is predicted by Wagers et al.'s (2009) proposal for marked features, under which the marked feature in the Marked-Unmarked condition stays longer in the focus of attention than the unmarked feature in the Unmarked-Marked condition, thus facilitating the detection of the agreement violation in the earlier condition, but not in the latter. Interestingly, this pattern is the same that emerged in López Prego & Gabriele's (2014) study, for advanced L2 learners performing a speeded version of a grammaticality judgment task. In this version of the task the sentences were presented word by word at a very rapid pace, thus involving high processing demands to perform the task. Thus, the speculation raised earlier that a greater processing burden may be necessary for such subtle interaction to emerge gains some weight. This may also suggest that the L2 learners and the native speaker control group in the current study didn't show this interaction because the regular self-paced reading task they performed wasn't taxing enough. A possible follow-up to the present study would use the eye-tracking methodology, where the eye movements of the participants are tracked moment by moment as they read sentences, in order to obtain a more sensitive measure of online processing. In addition, this type of eye-tracking study examining the role of markedness in agreement has never been done before, and thus would fill an interesting gap in the field.

In the Ungrammatical vs. Gendered comparison, the burdened native speaker group, like the L2 learners in this comparison, did show an effect of grammaticality. The effect was robust, but in contrast to the native speaker control group, it emerged only in

the noun region. Thus, it seems that the processing burden of the task resulted in the restriction of the grammaticality effect. Finally, with regards to the grammatical comparison, the marked feature facilitation effect emerged for these native speakers in the determiner region. Again, like for L2 learners, the effect was not as robust as for the native speaker control group, but it was also in the hypothesized direction, with only the Feminine-Feminine condition showing a facilitation effect compared to the corresponding invariant condition. It is interesting that this effect emerged in the determiner region for the burdened native speakers, thus providing evidence of predictive processing, but not in the native control group, or the learner group, where the effect emerged in the complementizer region. It is possible that the increased demands of this task forced the native speakers to engage in more efficient processing, as in predictive processing.

In conclusion, although no strong claims can be made due to the weak effects in the burdened native speaker group, the qualitative similarities between the results of this group and the L2 learner group provide tentative support for computational accounts of morphological variability, thus adding to the evidence provided by Hopp (2010), López Prego & Gabriele (2014), and McDonald (2006).

7.4. Limitations of the study

The main limitation in the current study is the small sample size of the learner group, which only includes sixteen learners. A larger sample size would provide greater statistical power, likely boosting the effects that emerged in the study, and allowing for stronger conclusions. Data collection is still ongoing, and will allow to further explore the strength of the current results.

A second limitation of this study relates to the weak effects found in the native speaker group with the added processing burden. It is likely that the memory burden added to the self-paced reading task was too taxing for the participants, and thus suppressed some potentially interesting effects. The weak results in this group didn't allow for a clear-cut comparison with the learner group, or for firm conclusions. One possible follow-up to the study would involve reducing the numbers to be remembered in the secondary task. Another option would involve changing the secondary task altogether, using for example distracting background speech.

One final limitation of the current study concerns the generalizability of the results reported here, since most of the learners tested are pursuing graduate degrees in Spanish, and teach Spanish classes at the college level. In addition, some of them are married to Spanish native speakers. Given the specific characteristics of this learner group, it is an open question whether only learners in such specific circumstances can show native-like behavior like the one reported in the current study. What sets these learners apart from other learners is a question that will be explored in the near future, by examining the data collected on the individual differences measures administered to the learners in this study. Nevertheless, regardless of the generalizability of the present results, the current study provides some valuable knowledge of what is ultimately possible for L2 acquisition.

8. Conclusion

The current study is one of the first to test the role of markedness in the online processing of long-distance gender agreement dependencies, in both L2 learners and native speakers. When tested in ungrammatical sentences, the results revealed that against the Shallow

Structure Hypothesis, L2 learners show online sensitivity to gender agreement violations occurring across different phrases. A yet more revealing result is that when tested on grammatical sentences, both L2 learners and native speakers show facilitation of agreement processing when a marked feature is involved, compared to when the feature is unmarked. This finding in native speakers is in line with Distributed Morphology's proposed representation of features as asymmetrical hierarchies, as well as with Wagers et al.'s (2009) proposal that marked features stay longer than unmarked features in the focus of attention. The fact that this finding obtains for L2 learners as well is consistent with computational theories like the Missing Surface Inflection Hypothesis, and proposals that L2 learner agreement errors may reflect a native-like representation of features. In addition, the absence of a correlation between the marked facilitation effect and frequency differences within the marked, feminine items tested, runs counter to a frequency-based explanation for the facilitation effect in L2 learners, contra Representational Deficit Hypothesis. Finally, the similarities observed between the native speakers performing under processing burden and the L2 learners provide some promising results in support of computational accounts of morphological variability in L2 learners.

REFERENCES

- Acuña-Fariña, J.C., Meseguer, E., Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, 143, 108-128.
- Akhutina, T., Kurgansky, A., Polinsky, M., & Bates, E. (1999). Processing of grammatical gender in a three-gender system: Experimental evidence from Russian. *Journal of Psycholinguistic Research*, 28(6), 695-713.

- Alemán-Bañón, J., Fiorentino, R., and Gabriele, A. (2014). Morphosyntactic processing in advanced second language learners: An event related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Language Research* 30(3), 275-306.
- Alemán-Bañón, J., Rothman, J., & Miller, D. (2015): Using event-related potentials to examine the nature of L2 morphological variability: The role of morphological markedness and L1 transfer. Poster presentation at 28th CUNY Conference on Human Sentence Processing, University of Southern California, March 19-21, 2015.
- Anderson, R. T. (1999). Loss of gender agreement in L1 attrition: Preliminary results. *Bilingual Research Journal*, 23(4), 389-408.
- Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, 5, 1-25.
- Battistella, E. (1990). *Markedness: The evaluative superstructure of language*. Albany, NY: SUNY Press.
- Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: the acquisition of Dutch grammatical gender. *Second Language Research*, 24(3), 297-331.
- Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, 8(1), 57-99.
- Bock, K. & Cutting, J.C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99-127.
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, 23(1), 45-93.
- Brisk, M.E. (1976). The acquisition of Spanish gender by first grade Spanish-speaking children. In G.D. Keller, R.V. Teschner, & S. Viera (Eds.), *Bilingualism in the bicentennial and beyond* (pp. 143-160). New York: Bilingual Press.
- Bruhn de Garavito, J. (2003). Learners' competence may be more accurate than we think: Spanish L2 and agreement morphology. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference* (pp. 17-23). Somerville, MA: Cascadilla Press.
- Bruhn de Garavito, J., & White, L. (2002). The L2 acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. M. Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153-178). Dordrecht: Kluwer.

- Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(1), 107-126.
- Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, 60(1), 21-43.
- Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, 63, 544–573.
- Cowper, Elizabeth. (2005). A note on number. *Linguistic Inquiry*, 36, 441-455.
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35, 353–387.
- Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, 36(2), 147-164.
- Eberhard, K.M., Cutting, J.C., Bock, J.K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, 112, 531-559.
- Franceschina, F. (2005). Fossilized second language grammars: The acquisition of grammatical gender. Amsterdam: Benjamins.
- Gabriele, A., Fiorentino, R., & Bañón, J. (2013). Examining second language development using event-related potentials: a cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism*, 3, 213-232.
- Gilchrist, A. L., & Cowan, N. (2011). Can the focus of attention accommodate multiple, separate items?. *Journal of experimental psychology: learning, memory, and cognition*, 37(6), 1484.
- Greenberg, Joseph (1966). *Language universals*. The Hague: Mouton.
- 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.
- Guajardo, L.F. & Wicha, N.Y.Y. (2014). Morphosyntax can modulate the N400 component: Event-related potentials to gender-marked post-nominal adjectives. *NeuroImage*, 91, 262-272.

- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111-176). Cambridge MA: MIT Press.
- Harley, H. (1994). Hug a tree: deriving the morphosyntactic feature hierarchy. In A. Carnie, & H. Harley (Eds.) *MITWPL 21: Papers on phonology and morphology* (pp. 275-288). Cambridge, MA: MITWPL.
- Harley, H., & Noyer, R. (1999). Distributed morphology. *Glott International*, 4(4), 3–9.
- Harley, H., & Ritter, E. (2002). Structuring the bundle: A universal morphosyntactic feature geometry. In H. Simon & H. Wiese (Eds.), *Pronouns: Features and representation* (pp. 23-39). Amsterdam: Benjamins.
- 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.
- Hawkins, R. (2009). Statistical Learning and Innate Knowledge in the Development of Second Language Proficiency: Evidence From the Acquisition of Gender Concord. *Issues in second language proficiency*, 63.
- Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, 118(4), 595-612.
- Haznedar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 257–268). Somerville, MA: Cascadilla Press.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between nonnative and native speakers. *Lingua*, 120(4), 901–931. DOI: 10.1016/j.lingua.2009.06.004.
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33-56.
- Jakobson, R. (Ed.). (1975). *N. S. Trubetzkoy's letters and notes*. The Hague: Mouton.
- Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP Study. *Journal of Psycholinguistic Research*, 31, 165-193.
- Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59(3), 503–535. DOI: 10.1111/j.1467-9922.2009.00516.x

- Keller, F., Lapata, M., & Ourioupina, O. (2002). Using the web to overcome data sparseness. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing* (pp. 230-237). Association for Computational Linguistics.
- Keller, F., & Lapata, M. (2003). Using the web to obtain frequencies for unseen bigrams. *Computational linguistics*, 29(3), 459-484.
- Lardiere, D. (1998a). Case and tense in the ‘fossilized’ steady state. *Second Language Research*, 14(1), 1–26. DOI: 10.1191/026765898674105303
- Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research*, 14(4), 359–375. DOI: 10.1191/026765898674105303
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 102–129). Oxford: Blackwell.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447–464.
- López Prego, B., & Gabriele, A. (2014). Examining the impact of task demands on morphological variability in native and non-native Spanish. *Linguistic Approaches to Bilingualism*, 4(2), 192–221. DOI 10.1075/lab.4.2.03lop
- Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language*, 36(01), 143-171.
- McCarthy, C. (2007). Morphological variability in second language Spanish. Unpublished doctoral dissertation, McGill University, Montreal, Canada.
- 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
- McCarthy, C. (2012). Modeling morphological variation and development: Person and number in L2 Spanish. *Linguistic Approaches to Bilingualism*, 2(1), 25–53. DOI: 10.1075/lab.2.1.02mcc
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401. DOI: 10.1016/j.jml.2006.06.006

- Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research*, 21(3), 199–249. DOI: 10.1191/0267658305sr247oa
- Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers. *Studies in second language acquisition*, 25(03), 351-398.
- 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
- Natalicio, D. (1983). Native speaker intuitions as a basis for determining noun gender rules in Spanish. *Southwest Journal of Linguistics*, 6, 49-55.
- Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, 36(4), 569-587.
- Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research*, 29(1), 89-98.
- Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, 18(03), 571-590.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16(2), 110–133.
- Renaud, C. (2010). On the nature of agreement in English-French acquisition: A processing investigation in the verbal and nominal domains. Unpublished doctoral dissertation, Indiana University.
- Renaud, C. (2011). Processing gender: The case of pronouns and adjectives in L2 French. In J. Herschensohn & D. Tanner (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference* (pp. 121-134). Somerville, MA: Cascadilla Press.
- Renaud, C. (2012). An investigation of the role of gender in the resolution of pronouns in French. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 512-524). Somerville, MA: Cascadilla Press.
- Tagliaferri, B. (2005). Paradigm [Computer software]. Perception Research Systems, Inc. Retrieved from www.perceptionresearchsystems.com

- Trubetzkoy, Nikolaus (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
- Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, 1(1), 115-132.
- Vigliocco, G. & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, 40, 455-478.
- Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, 45(3), 368-390.
- Wagers, M., Lau, E., Stroud, C., McElree, B., & Phillips, C. (2009, March). Encoding syntactic predictions: Evidence from the dynamics of agreement. Paper presented at the CUNY Sentence Processing Conference 22, Davis, CA.
- White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(1), 105-133.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39(3), 483-508.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of cognitive neuroscience*, 16(7), 1272-1288.
- Wicha, N. Y., Orozco-Figueroa, A., Reyes, I., Hernandez, A., Gavaldón de Barreto, L., & Bates, E. A. (2005). When zebras become painted donkeys: Grammatical gender and semantic priming interact during picture integration in a spoken Spanish sentence. *Language and cognitive processes*, 20(4), 553-587.

APPENDIX I

Target stimuli sets

- 1 Como es blanca, he decidido que llevaré el monedero que compré en la boutique.
Como es blanco, he decidido que llevaré el monedero que compré en la boutique.
Como es suave, he decidido que llevaré el monedero que compré en la boutique.
Como es blanco, he decidido que llevaré la gorra que compré en la boutique.
Como es blanca, he decidido que llevaré la gorra que compré en la boutique.
Como es suave, he decidido que llevaré la gorra que compré en la boutique.
Since it is white/soft, I have decided that I will wear/carry the wallet/beany that I bought at the boutique.

- 2 Como es blanca, he decidido que compraré el sombrero que vi en las rebajas.
Como es blanco, he decidido que compraré el sombrero que vi en las rebajas.
Como es suave, he decidido que compraré el sombrero que vi en las rebajas.
Como es blanco, he decidido que compraré la almohada que vi en las rebajas.
Como es blanca, he decidido que compraré la almohada que vi en las rebajas.
Como es suave, he decidido que compraré la almohada que vi en las rebajas.
Since it is white/soft, I have decided that I will buy the hat/pillow that I saw on sale.

- 3 Como es blanca, he decidido que modificaré el abrigo que dejé en la entrada.
Como es blanco, he decidido que modificaré el abrigo que dejé en la entrada.
Como es suave, he decidido que modificaré el abrigo que dejé en la entrada.
Como es blanco, he decidido que modificaré la chaqueta que dejé en la entrada.
Como es blanca, he decidido que modificaré la chaqueta que dejé en la entrada.
Como es suave, he decidido que modificaré la chaqueta que dejé en la entrada.

- Since it is white/soft, I have decided that I will modify the coat/jacket that I left in the hall.
- 4 Como es negra, he decidido que usaré el marco que traje de la tienda.
 Como es negro, he decidido que usaré el marco que traje de la tienda.
 Como es azul, he decidido que usaré el marco que traje de la tienda.
 Como es negro, he decidido que usaré la pintura que traje de la tienda.
 Como es negra, he decidido que usaré la pintura que traje de la tienda.
 Como es azul, he decidido que usaré la pintura que traje de la tienda.
 Since it is black/blue, I have decided that I will use the frame/paint that I brought from the store.
- 5 Como es negra, he decidido que encargará el banco que señalé en el catálogo.
 Como es negro, he decidido que encargará el banco que señalé en el catálogo.
 Como es azul, he decidido que encargará el banco que señalé en el catálogo.
 Como es negro, he decidido que encargará la carpeta que señalé en el catálogo.
 Como es negra, he decidido que encargará la carpeta que señalé en el catálogo.
 Como es azul, he decidido que encargará la carpeta que señalé en el catálogo.
 Since it is black/blue, I have decided that I will order the bench/folder that I marked in the catalogue.
- 6 Como es negra, he decidido que traeré el carro que descubrí en la feria.
 Como es negro, he decidido que traeré el carro que descubrí en la feria.
 Como es azul, he decidido que traeré el carro que descubrí en la feria.
 Como es negro, he decidido que traeré la guitarra que descubrí en la feria.
 Como es negra, he decidido que traeré la guitarra que descubrí en la feria.
 Como es azul, he decidido que traeré la guitarra que descubrí en la feria.
 Since it is black/blue, I have decided that I will bring the car/guitar that I discovered at the fair.
- 7 Como es roja, he decidido que venderé el chaleco que hice el año pasado.
 Como es rojo, he decidido que venderé el chaleco que hice el año pasado.
 Como es gris, he decidido que venderé el chaleco que hice el año pasado.
 Como es rojo, he decidido que venderé la corbata que hice el año pasado.
 Como es roja, he decidido que venderé la corbata que hice el año pasado.
 Como es gris, he decidido que venderé la corbata que hice el año pasado.
 Since it is red/grey, I have decided that I will sell the vest/tie that I made last year.
- 8 Como es roja, he decidido que repararé el casco que rompí en el viaje.
 Como es rojo, he decidido que repararé el casco que rompí en el viaje.
 Como es gris, he decidido que repararé el casco que rompí en el viaje.
 Como es rojo, he decidido que repararé la maleta que rompí en el viaje.
 Como es roja, he decidido que repararé la maleta que rompí en el viaje.
 Como es gris, he decidido que repararé la maleta que rompí en el viaje.
 Since it is red/grey, I have decided that I will repair the helmet/suitcase that I broke during the trip.

- 9 Como es roja, he decidido que donaré el poncho que traje de las montañas.
Como es rojo, he decidido que donaré el poncho que traje de las montañas.
Como es gris, he decidido que donaré el poncho que traje de las montañas.
Como es rojo, he decidido que donaré la gabardina que traje de las montañas.
Como es roja, he decidido que donaré la gabardina que traje de las montañas.
Como es gris, he decidido que donaré la gabardina que traje de las montañas.
Since it is red/grey, I have decided that I will donate the poncho/raincoat that I brought from the mountains.
- 10 Como es amarilla, he decidido que coseré el paño que deshice el mes pasado.
Como es amarillo, he decidido que coseré el paño que deshice el mes pasado.
Como es lavable, he decidido que coseré el paño que deshice el mes pasado.
Como es amarillo, he decidido que coseré la alfombra que deshice el mes pasado.
Como es amarilla, he decidido que coseré la alfombra que deshice el mes pasado.
Como es lavable, he decidido que coseré la alfombra que deshice el mes pasado.
Since it is yellow/green, I have decided that I will sew the rag/carpet that I unravelled last month.
- 11 Como es amarilla, he decidido que cambiaré el florero que puse en la cocina.
Como es amarillo, he decidido que cambiaré el florero que puse en la cocina.
Como es verde, he decidido que cambiaré el florero que puse en la cocina.
Como es amarillo, he decidido que cambiaré la lana que puse en la cocina.
Como es amarilla, he decidido que cambiaré la lana que puse en la cocina.
Como es verde, he decidido que cambiaré la lana que puse en la cocina.
Since it is yellow/green, I have decided that I will return the vase/wool that I put in the kitchen.
- 12 Como es amarilla, he decidido que lavaré el columpio que dejé en el jardín.
Como es amarillo, he decidido que lavaré el columpio que dejé en el jardín.
Como es verde, he decidido que lavaré el columpio que dejé en el jardín.
Como es amarillo, he decidido que lavaré la manta que dejé en el jardín.
Como es amarilla, he decidido que lavaré la manta que dejé en el jardín.
Como es verde, he decidido que lavaré la manta que dejé en el jardín.
Since it is yellow/green, I have decided that I will wash the swing/blanket that I left in the garden.
- 13 Como es oscura, he decidido que barnizaré el barco que heredé de mi primo.
Como es oscuro, he decidido que barnizaré el barco que heredé de mi primo.
Como es resistente, he decidido que barnizaré el barco que heredé de mi primo.
Como es oscuro, he decidido que barnizaré la mesa que heredé de mi primo.
Como es oscura, he decidido que barnizaré la mesa que heredé de mi primo.
Como es resistente, he decidido que barnizaré la mesa que heredé de mi primo.
Since it is dark/sturdy, I have decided that I will varnish the boat/table that I inherited from my cousin.

- 14 Como es oscura, he decidido que regalaré el cepillo que encontré en el mercadillo.
 Como es oscuro, he decidido que regalaré el cepillo que encontré en el mercadillo.
 Como es grande, he decidido que regalaré el cepillo que encontré en el mercadillo.
 Como es oscuro, he decidido que regalaré la cartera que encontré en el mercadillo.
 Como es oscura, he decidido que regalaré la cartera que encontré en el mercadillo.
 Como es grande, he decidido que regalaré la cartera que encontré en el mercadillo.
 Since it is dark/big, I have decided that I will give away the brush/wallet that I found at the flea market.
- 15 Como es oscura, he decidido que usaré el saco que traje para mi hermano.
 Como es oscuro, he decidido que usaré el saco que traje para mi hermano.
 Como es resistente, he decidido que usaré el saco que traje para mi hermano.
 Como es oscuro, he decidido que usaré la silla que traje para mi hermano.
 Como es oscura, he decidido que usaré la silla que traje para mi hermano.
 Como es resistente, he decidido que usaré la silla que traje para mi hermano.
 Since it is dark/sturdy, I have decided that I will use the sack/chair that I brought for my brother.
- 16 Como es dorada, he decidido que sacaré el plato que guardé en la vitrina.
 Como es dorado, he decidido que sacaré el plato que guardé en la vitrina.
 Como es resistente, he decidido que sacaré el plato que guardé en la vitrina.
 Como es dorado, he decidido que sacaré la cuchara que guardé en la vitrina.
 Como es dorada, he decidido que sacaré la cuchara que guardé en la vitrina.
 Como es resistente, he decidido que sacaré la cuchara que guardé en la vitrina.
 Since it is golden/sturdy, I have decided that I will take out the plate/spoon that I put in the glass cabinet.
- 17 Como es dorada, he decidido que esconderé el vaso que encontré en el baúl.
 Como es dorado, he decidido que esconderé el vaso que encontré en el baúl.
 Como es transparente, he decidido que esconderé el vaso que encontré en el baúl.
 Como es dorado, he decidido que esconderé la falda que encontré en el baúl.
 Como es dorada, he decidido que esconderé la falda que encontré en el baúl.
 Como es transparente, he decidido que esconderé la falda que encontré en el baúl.
 Since it is golden/transparent, I have decided that I will hide the glass/skirt that I found in the chest.
- 18 Como es dorada, he decidido que encargará el cuchillo que vi en el escaparate.
 Como es dorado, he decidido que encargará el cuchillo que vi en el escaparate.
 Como es original, he decidido que encargará el cuchillo que vi en el escaparate.
 Como es dorado, he decidido que encargará la tela que vi en el escaparate.
 Como es dorada, he decidido que encargará la tela que vi en el escaparate.
 Como es original, he decidido que encargará la tela que vi en el escaparate.
 Since it is golden/singular, I have decided that I will order the knife/fabric that I saw in the store window.

- 19 Como es rosada, he decidido que guardaré el pañuelo que encontré en la basura.
Como es rosado, he decidido que guardaré el pañuelo que encontré en la basura.
Como es lavable, he decidido que guardaré el pañuelo que encontré en la basura.
Como es rosado, he decidido que guardaré la bufanda que encontré en la basura.
Como es rosada, he decidido que guardaré la bufanda que encontré en la basura.
Como es lavable, he decidido que guardaré la bufanda que encontré en la basura.
Since it is pink/washable, I have decided that I will keep the handkerchief/scarf that I found in the trash.
- 20 Como es rosada, he decidido que tiraré el globo que traje de la feria.
Como es rosado, he decidido que tiraré el globo que traje de la feria.
Como es transparente, he decidido que tiraré el globo que traje de la feria.
Como es rosado, he decidido que tiraré la camisa que traje de la feria.
Como es rosada, he decidido que tiraré la camisa que traje de la feria.
Como es transparente, he decidido que tiraré la camisa que traje de la feria.
Since it is pink/transparent, I have decided that I will throw away the balloon/shirt that I brought from the fair.
- 21 Como es rosada, he decidido que traeré el lazo que dejé en tu casa.
Como es rosado, he decidido que traeré el lazo que dejé en tu casa.
Como es lavable, he decidido que traeré el lazo que dejé en tu casa.
Como es rosado, he decidido que traeré la cinta que dejé en tu casa.
Como es rosada, he decidido que traeré la cinta que dejé en tu casa.
Como es lavable, he decidido que traeré la cinta que dejé en tu casa.
Since it is pink/washable, I have decided that I will bring the ribbon/hairband that I left at your house.
- 22 Como es plateada, he decidido que tiraré el vestido que compré para el carnaval.
Como es plateado, he decidido que tiraré el vestido que compré para el carnaval.
Como es transparente, he decidido que tiraré el vestido que compré para el carnaval.
Como es plateado, he decidido que tiraré la blusa que compré para el carnaval.
Como es plateada, he decidido que tiraré la blusa que compré para el carnaval.
Como es transparente, he decidido que tiraré la blusa que compré para el carnaval.
Since it is silver/transparent, I have decided that I will throw away the dress/blouse that I bought for carnival.
- 23 Como es plateada, he decidido que retiraré el bolso que dejé en la mesilla.
Como es plateado, he decidido que retiraré el bolso que dejé en la mesilla.
Como es elegante, he decidido que retiraré el bolso que dejé en la mesilla.
Como es plateado, he decidido que retiraré la flauta que dejé en la mesilla.
Como es plateada, he decidido que retiraré la flauta que dejé en la mesilla.
Como es elegante, he decidido que retiraré la flauta que dejé en la mesilla.
Since it is silver (plated)/elegant, I have decided that I will remove the handbag/flute that I left on the coffee table.

- 24 Como es plateada, he decidido que cubriré el espejo que gané en el concurso.
Como es plateado, he decidido que cubriré el espejo que gané en el concurso.
Como es verde, he decidido que cubriré el espejo que gané en el concurso.
Como es plateado, he decidido que cubriré la copa que gané en el concurso.
Como es plateada, he decidido que cubriré la copa que gané en el concurso.
Como es verde, he decidido que cubriré la copa que gané en el concurso.
Since it is silver/green, I have decided that I will cover the mirror/cup that I won at the show.
- 25 Como es vieja, he decidido que restauraré el piano que guardé en el garaje.
Como es viejo, he decidido que restauraré el piano que guardé en el garaje.
Como es elegante, he decidido que restauraré el piano que guardé en el garaje.
Como es viejo, he decidido que restauraré la cámara que guardé en el garaje.
Como es vieja, he decidido que restauraré la cámara que guardé en el garaje.
Como es elegante, he decidido que restauraré la cámara que guardé en el garaje.
Since it is old/elegant, I have decided that I will restore the piano/camara that I put in the garage.
- 26 Como es vieja, he decidido que taparé el violonchelo que compré en los ochenta.
Como es viejo, he decidido que taparé el violonchelo que compré en los ochenta.
Como es frágil, he decidido que taparé el violonchelo que compré en los ochenta.
Como es viejo, he decidido que taparé la cafetera que compré en los ochenta.
Como es vieja, he decidido que taparé la cafetera que compré en los ochenta.
Como es frágil, he decidido que taparé la cafetera que compré en los ochenta.
Since it is old/fragile, I have decided that I will cover the violoncelo/coffeemaker that I bought in the eighties.
- 27 Como es vieja, he decidido que puliré el anillo que recibí de mi abuela.
Como es viejo, he decidido que puliré el anillo que recibí de mi abuela.
Como es elegante, he decidido que puliré el anillo que recibí de mi abuela.
Como es viejo, he decidido que puliré la pulsera que recibí de mi abuela.
Como es vieja, he decidido que puliré la pulsera que recibí de mi abuela.
Como es elegante, he decidido que puliré la pulsera que recibí de mi abuela.
Since it is old/elegant, I have decided that I will polish the ring/bracelet that I received from my grandma.
- 28 Como es nueva, he decidido que cuidaré el horno que compré para la casa.
Como es nuevo, he decidido que cuidaré el horno que compré para la casa.
Como es original, he decidido que cuidaré el horno que compré para la casa.
Como es nuevo, he decidido que cuidaré la cama que compré para la casa.
Como es nueva, he decidido que cuidaré la cama que compré para la casa.
Como es original, he decidido que cuidaré la cama que compré para la casa.
Since it is new/singular, I have decided that I will take care of the oven/bed that I bought for the house.
- 29 Como es nueva, he decidido que apartaré el termómetro que dejé en la estantería.

- Como es nuevo, he decidido que apartaré el termómetro que dejé en la estantería.
Como es frágil, he decidido que apartaré el termómetro que dejé en la estantería.
Como es nuevo, he decidido que apartaré la bandeja que dejé en la estantería.
Como es nueva, he decidido que apartaré la bandeja que dejé en la estantería.
Como es frágil, he decidido que apartaré la bandeja que dejé en la estantería.
Since it is new/fragile, I have decided that I will remove the thermometer/tray that I left on the shelf.
- 30 Como es nueva, he decidido que limpiaré el microscopio que ensucié por la mañana.
Como es nuevo, he decidido que limpiaré el microscopio que ensucié por la mañana.
Como es original, he decidido que limpiaré el microscopio que ensucié por la mañana.
Como es nuevo, he decidido que limpiaré la bicicleta que ensucié por la mañana.
Como es nueva, he decidido que limpiaré la bicicleta que ensucié por la mañana.
Como es original, he decidido que limpiaré la bicicleta que ensucié por la mañana.
Since it is new/singular, I have decided that I will clean the microscope/bicycle that I got dirty this morning.
- 31 Como es fea, he decidido que quemaré el cuadro que hice en el colegio.
Como es feo, he decidido que quemaré el cuadro que hice en el colegio.
Como es grande, he decidido que quemaré el cuadro que hice en el colegio.
Como es feo, he decidido que quemaré la figura que hice en el colegio.
Como es fea, he decidido que quemaré la figura que hice en el colegio.
Como es grande, he decidido que quemaré la figura que hice en el colegio.
Since it is ugly/big, I have decided that I will burn the painting/figure that I made at school.
- 32 Como es fea, he decidido que destrozaré el armario que encontré en el jardín.
Como es feo, he decidido que destrozaré el armario que encontré en el jardín.
Como es grande, he decidido que destrozaré el armario que encontré en el jardín.
Como es feo, he decidido que destrozaré la escultura que encontré en el jardín.
Como es fea, he decidido que destrozaré la escultura que encontré en el jardín.
Como es grande, he decidido que destrozaré la escultura que encontré en el jardín.
Since it is ugly/big, I have decided that I will destroy the closet/sculpture that I found in the garden.
- 33 Como es fea, he decidido que donaré el escritorio que hice para mi madre.
Como es feo, he decidido que donaré el escritorio que hice para mi madre.
Como es frágil, he decidido que donaré el escritorio que hice para mi madre.
Como es feo, he decidido que donaré la lámpara que hice para mi madre.
Como es fea, he decidido que donaré la lámpara que hice para mi madre.
Como es frágil, he decidido que donaré la lámpara que hice para mi madre.
Since it is ugly/fragile, I have decided that I will donate the desk/lamp that I made for my mom.

- 34 Como es antigua, he decidido que visitaré el monasterio que vi en la guía.
 Como es antiguo, he decidido que visitaré el monasterio que vi en la guía.
 Como es célebre, he decidido que visitaré el monasterio que vi en la guía.
 Como es antiguo, he decidido que visitaré la iglesia que vi en la guía.
 Como es antigua, he decidido que visitaré la iglesia que vi en la guía.
 Como es célebre, he decidido que visitaré la iglesia que vi en la guía.
 Since it is old/famous, I have decided that I will visit the monastery/church that I saw in the guide.
- 35 Como es antigua, he decidido que renovaré el colegio que apunté en la lista.
 Como es antiguo, he decidido que renovaré el colegio que apunté en la lista.
 Como es célebre, he decidido que renovaré el colegio que apunté en la lista.
 Como es antiguo, he decidido que renovaré la biblioteca que apunté en la lista.
 Como es antigua, he decidido que renovaré la biblioteca que apunté en la lista.
 Como es célebre, he decidido que renovaré la biblioteca que apunté en la lista.
 Since it is old/famous, I have decided that I will renovate the school/library that I wrote down on the list.
- 36 Como es antigua, he decidido que reconstruiré el instituto que seleccioné en la inspección.
 Como es antiguo, he decidido que reconstruiré el instituto que seleccioné en la inspección.
 Como es célebre, he decidido que reconstruiré el instituto que seleccioné en la inspección.
 Como es antiguo, he decidido que reconstruiré la plaza que seleccioné en la inspección.
 Como es antigua, he decidido que reconstruiré la plaza que seleccioné en la inspección.
 Como es célebre, he decidido que reconstruiré la plaza que seleccioné en la inspección.
 Since it is old/famous, I have decided that I will rebuild the high-school/plaza that I selected in the inspection.

Distracters

- 37 Cuando mi madre era niña, siempre estudiaba en el jardín de su casa.
 When my mom was a child, she would always study in the garden of her house.
- 38 Cuando el colegio era público, siempre salíamos al patio durante el descanso.
 When the school was public, we would always go out to the playground during recess.
- 39 Cuando mi padre era joven, siempre iba a la playa los domingos.
 When my dad was young, he would always go to the beach on Sundays.
- 40 Cuando los perros eran chachorros, siempre mordían los zapatos de los niños.

- When the dogs were puppies, they would always chew the kids' shoes.
- 41 Cuando la universidad era pública, siempre había muchos estudiantes en la biblioteca.
When the university was public, there would always be many students in the library.
- 42 Cuando los libros eran gratis, siempre teníamos más dinero para los otros materiales.
When the books were free, we would always have more money for the other supplies.
- 43 Cuando el cine era barato, siempre íbamos los domingos a ver alguna película.
When going to the movies was cheap, we would always go to watch some movie on Sundays.
- 44 Cuando la playa era accesible, siempre íbamos a bañarnos en las tardes de verano.
When the beach was accessible, we would always go swimming in the summer afternoons.
- 45 Cuando la casa era nuestra, siempre íbamos de excursión allí en el invierno.
When the house was ours, we would always go there on trips during winter.
- 46 Cuando las patatas eran abundantes, siempre preparábamos cocido los fines de semana.
When potatoes abounded, we would always make 'cocido' on the weekends.
- 47 Cuando el invierno era largo, siempre hacíamos hogueras en la chimenea de la sala.
When the winter was long, we would always light a fire in the living-room fireplace.
- 48 Cuando las materias eran difíciles, siempre estudiábamos en grupos de varias personas.
When the subjects were difficult, we would always study in groups of several people.
- 49 Cuando los días eran lluviosos, siempre saltábamos en los charcos de la calle.
When days were rainy, we would always jump in the puddles in the street.
- 50 Cuando la carne era cara, siempre comíamos muchas patatas con esas verduras.
When meat was expensive, we would always eat many potatoes with those greens.
- 51 Cuando los profesores eran buenos, siempre respondíamos a las preguntas en la clase.

- When the teachers were good, we would always answer the questions in the classroom.
- 52 Cuando los niños eran traviesos, siempre preparábamos ejercicios con mucha actividad física.
When the kids were naughty, we would always plan tasks with lots of physical activity.
- 53 Cuando el gato era pequeño, siempre arañaba los muebles de mi habitación.
When the cat was a kitten, he would always scratch the furniture in my bedroom.
- 54 Cuando las vacaciones eran cortas, siempre venían mis primas uno de los días.
When the holidays were short, my cousins would always come one of those days.
- 55 Cuando el lago era profundo, siempre *nadamos más cerca de la orilla.
When the lake was deep, we would always *swam closer to the shore.
- 56 Cuando el bosque era frondoso, siempre *jugamos en los árboles más altos.
When the forest was dense, we would always *played on the tallest trees.
- 57 Cuando la fruta era barata, siempre *compramos cerezas en los meses de verano.
When fruit was cheap, we would always *bought cherries in the summer months.
- 58 Cuando los festivales eran frecuentes, siempre *fuimos a los conciertos de artistas desconocidos.
When music festivals were frequent, we would always *went to the concerts of unknown artists.
- 59 Cuando las tardes eran calurosas, siempre *tomamos limonada con los hijos de los vecinos.
When the afternoons were hot, we would always *had lemonade with the neighbors' kids.
- 60 Cuando los vecinos eran famosos, siempre hicieron *fiestas en el salón de su casa.
When the neighbors were famous, they would always *held parties in their living-room.
- 61 Cuando las orquestas eran populares, siempre *bailamos en la plaza del pueblo.
When orquestas were popular, we would always *danced in the town square.
- 62 Cuando el pan era caro, siempre *cocimos varios bollos en nuestra cocina.
When bread was expensive, we would always *baked several rolls in our kitchen.
- 63 Cuando los artistas eran conocidos, siempre *pedimos autógrafos en la entrada del hotel.

- When the artists were well-known, we would always *asked for autographs in the hotel lobby.
- 64 Cuando las canciones eran repetitivas, siempre *salimos a la terraza del bar.
When the songs were repetitive, we would always *went out to the bar's terrace.
- 65 Cuando las reuniones eran importantes, siempre *llevamos comida para los padres de los niños.
When meetings were important, we would always * brought food for the children's parents.
- 66 Cuando las opciones eran pocas, siempre *inventamos nuevas estrategias durante el juego.
When options were limited, we would always *thought up new strategies during the game.
- 67 Cuando las clases eran obligatorias, siempre *asistimos a las sesiones de la tarde.
When classes were mandatory, we would always *attended the afternoon sessions.
- 68 Cuando el café era bueno, siempre *tomamos varias tazas con nuestros amigos.
When coffee was good, we would always *had several cups with our friends.
- 69 Cuando las noches eran frías, siempre *usamos la chimenea de la habitación principal.
When the nights were cold, we would always *used the fireplace in the main room.
- 70 Cuando los actores eran buenos, siempre *vimos las películas en una pantalla gigante.
When the actors were good, we would always *watched the movies in a huge screen.
- 71 Cuando las mañanas eran soleadas, siempre *paseamos por el paseo de la playa.
When the mornings were sunny, we would always *took a stroll down the beach boulevard.
- 72 Cuando el viento era fuerte, siempre *tendimos la ropa dentro de casa.
When the wind was strong, we would always *hung our clothes inside the house.

APPENDIX II

Co-occurrence frequencies of target noun-adjective pairs

Masc. Noun + Masc. Adj.	Raw Frequency (Google tally)	Log10 Frequency
monedero blanco	9,890	3.995196292
sombrero blanco	199,000	5.298853076
abrigo blanco	116,000	5.064457989
marco negro	731,000	5.863917377
carro negro	86,300	4.936010796
banco negro	49,600	4.695481676
chaleco rojo	103,000	5.012837225
casco rojo	338,000	5.5289167
poncho rojo	52,600	4.720985744
pañó amarillo	13,700	4.136720567
florero amarillo	6,210	3.7930916
columpio amarillo	4,520	3.655138435
barco oscuro	8,660	3.937517892
cepillo oscuro	737	2.867467488
saco oscuro	8,420	3.925312091
plato dorado	10,900	4.037426498
vaso dorado	1,670	3.222716471
cuchillo dorado	2,010	3.303196057
pañuelo rosado	11,400	4.056904851
globo rosado	6,260	3.796574333
lazo rosado	49,400	4.693726949
vestido plateado	215,000	5.33243846
bolso plateado	20,400	4.309630167
espejo plateado	28,900	4.460897843
piano viejo	11,600	4.064457989
violonchelo viejo	133	2.123851641
anillo viejo	6,640	3.822168079
horno nuevo	106,000	5.025305865
termómetro nuevo	29,100	4.463892989
microscopio nuevo	9,770	3.989894564
cuadro feo	2,430	3.385606274
armario feo	729	2.862727528
escritorio feo	1,460	3.164352856
monasterio antiguo	6,930	3.840733235
colegio antiguo	12,400	4.093421685
instituto antiguo	3,750	3.574031268
Mean	62,903	4
SD	135834.3185	0.817793351

Masc. Noun + Inv. Adj.	Raw Frequency (Google tally)	Log10 Frequency
monedero suave	1,430	3.155336037
sombrero suave	6,220	3.793790385
abrigo suave	6,100	3.785329835
marco azul	122,000	5.086359831
banco azul	91,600	4.961895474
carro azul	63,900	4.805500858
chaleco gris	116,000	5.064457989
casco gris	33,300	4.522444234
poncho gris	7,750	3.889301703
pañó lavable	7,970	3.901458321
lazo lavable	3,440	3.536558443
pañuelo lavable	9	0.954242509
florero verde	11,200	4.049218023
columpio verde	3,790	3.57863921
espejo verde	34,700	4.540329475
barco resistente	7,750	3.889301703
saco resistente	2,390	3.378397901
plato resistente	3,240	3.51054501
cepillo grande	37,100	4.56937391
cuadro grande	253,000	5.403120521
armario grande	202,000	5.305351369
vaso transparente	36,300	4.559906625
globo transparente	19,300	4.285557309
vestido transparente	268,000	5.428134794
cuchillo original	10,900	4.037426498
horno original	8,980	3.953276337
microscopio original	536	2.72916479
bolso elegante	90,600	4.957128198
piano elegante	2,860	3.456366033
anillo elegante	32,300	4.509202522
violonchelo frágil	0	0
termómetro frágil	3	0.477121255
escritorio frágil	74	1.86923172
monasterio célebre	313	2.495544338
colegio célebre	681	2.833147112
instituto célebre	6,280	3.797959644
Mean	41,445	4
SD	69882.12775	1.303932307

Fem. Noun + Fem. Adj.	Raw Frequency (Google tally)	Log10 Frequency
gorra blanca	127,000	5.103803721
almohada blanca	26,900	4.42975228
chaqueta blanca	332,000	5.521138084
pintura negra	351,000	5.545307116
carpeta negra	27,100	4.432969291
guitarra negra	144,000	5.158362492
corbata roja	150,000	5.176091259
gabardina roja	14,700	4.167317335
maleta roja	116,000	5.064457989
alfombra amarilla	14,700	4.167317335
lana amarilla	17,200	4.235528447
manta amarilla	7,150	3.854306042
mesa oscura	23,600	4.372912003
silla oscura	2,070	3.315970345
cartera oscura	1,020	3.008600172
cuchara dorada	3,260	3.5132176
falda dorada	14,400	4.158362492
tela dorada	18,600	4.269512944
cinta rosada	164,000	5.214843848
bufanda rosada	10,100	4.004321374
camisa rosada	69,600	4.84260924
copa plateada	5,520	3.741939078
flauta plateada	6,040	3.781036939
blusa plateada	5,520	3.741939078
cámara vieja	30,500	4.484299839
cafetera vieja	2,250	3.352182518
pulsera vieja	2,680	3.428134794
cama nueva	195,000	5.290034611
bicicleta nueva	606,000	5.782472624
bandeja nueva	19,600	4.292256071
figura fea	2,880	3.459392488
escultura fea	1,130	3.053078443
lámpara fea	831	2.919601024
iglesia antigua	175,000	5.243038049
biblioteca antigua	72,000	4.857332496
plaza antigua	111,000	5.045322979
Mean	79,732	4
SD	127,326	1

Fem. Noun + Inv. Adj.	Raw Frequency (Google tally)	Log10 Frequency
gorra suave	2,750	3.439332694
almohada suave	41,800	4.621176282
chaqueta suave	20,900	4.320146286
pintura azul	277,000	5.442479769
carpeta azul	41,500	4.618048097
guitarra azul	54,900	4.739572344
corbata gris	40,300	4.605305046
maleta gris	5,540	3.743509765
gabardina gris	31,600	4.499687083
alfombra lavable	24,200	4.383815366
cinta lavable	3,040	3.482873584
bufanda lavable	738	2.868056362
lana verde	95,000	4.977723605
manta verde	15,500	4.190331698
copa verde	39,200	4.593286067
mesa resistente	18,700	4.271841607
silla resistente	63,500	4.802773725
cuchara resistente	3,840	3.584331224
cartera grande	214,000	5.330413773
figura grande	69,900	4.844477176
escultura grande	6,540	3.815577748
falda transparente	48,300	4.683947131
camisa transparente	44,000	4.643452676
blusa transparente	144,000	5.158362492
tela original	217,000	5.336459734
cama original	27,700	4.442479769
bicicleta original	76,400	4.883093359
flauta elegante	1,110	3.045322979
cámara elegante	38,700	4.587710965
pulsera elegante	52,800	4.722633923
cafetera frágil	9	0.954242509
bandeja frágil	84	1.924279286
lámpara frágil	276	2.440909082
iglesia célebre	6,610	3.820201459
biblioteca célebre	247	2.392696953
plaza célebre	438	2.641474111
Mean	48,003	4
SD	66095.64833	1.041407699

APPENDIX III

Practice sentences in self-paced reading tasks

1. Cuando es agradable, mi prima puede ser una persona muy divertida.
When he/she is nice, my cousin can be a really fun person.
2. Planté muchos, por eso he decidido donar los guisantes que están en esa cesta.
I sowed many, that is why I have decided to donate the peas in this basket.
3. Como tengo demasiados, he decidido que venderé los melones que planté el año pasado.
Since I have too many, I have decided that I will sell the melons that I sowed last year.
4. Como son plegables, creo que llevaré *la silla que usamos en la playa.
Since they are foldable_{pl}, I think I will bring *the chair that we use for the beach.
5. Antes de que pudiera empezar a *corregirla, las redacciones desaparecieron de la computadora.
Before I could start grading it, the compositions disappeared from the computer.
6. Cuando quise *retomarlas, la novela ya no estaba donde la había dejado.
When I went to resume *them, the novel wasn't where I had left it any more.
7. Como quería terminarlo, me llevé el libro para leer en el avión.
Since I wanted to finish it, I took the book with me to read on the plane.
8. Dos días después de comprarlas, las entradas bajaron de precio.
Two days after buying them, the tickets went down in price.
9. Como son inoxidable, he decidido que compraré *el tenedor que vi en la television.
Since they are stainless_{pl}, I have decided that I will buy *the fork that I saw on TV.
10. Como son desmontables, voy a encargar *el mueble que me recomendó tu cuñado.
Since they can be disassembled_{pl}, I am going to order *the cupboard that your brother-in-law recommended.