Language Learning ISSN 0023-8333

Sensitivity to Violations of Gender Agreement in Native and Nonnative Spanish: An Eye-Movement Investigation

Gregory D. Keating

San Diego State University

This article reports the results of an eye-tracking experiment that investigated the effects of structural distance on readers' sensitivity to violations of Spanish gender agreement during online sentence comprehension. The study tracked the eye movements of native Spanish speakers and English-speaking learners of Spanish as they read sentences that contained nouns modified by postnominal adjectives located in three syntactic domains: (a) in the DP, (b) in the VP, or (c) in a subordinate clause. In half of the sentences in each condition, adjectives agreed with the noun in gender, and in half, they did not. The results indicate that gender agreement is acquirable in adulthood, contra the failed functional features hypothesis, and that the distance that separates nouns and adjectives affects the detection of gender anomalies in the second language. The findings support Clahsen and Felser's (2006a) shallow structure hypothesis, as it pertains to morphological processing.

Keywords eye-tracking; failed functional features hypothesis; full transfer full access hypothesis; gender agreement; sentence processing; shallow structure hypothesis; structural distance; Universal Grammar (UG); Spanish

Research in the field of second language acquisition (SLA) has produced conflicting evidence regarding how parameterized functional features are

This research was made possible with the generous support of a Dissertation Grant awarded by the Board of Directors of *Language Learning* and a Provost's Award for Graduate Research from the University of Illinois at Chicago. I am especially grateful to Bill VanPatten for his guidance and support during all phases of this research. I am also greatly indebted to Silvina Montrul, Jessica Williams, Luis López, Susanne Rott, Richard Cameron, Michael J. Leeser, Andrew P. Farley, and three anonymous reviewers for their insightful comments at various stages of this project. I thank Eniko Csomay for helpful statistical discussions. All errors are mine. An earlier version of this work was presented at the Second Language Research Forum in Seattle, WA (October 6–8, 2006).

Correspondence concerning this article should be addressed to Gregory D. Keating, Department of Linguistics and Asian/Middle Eastern Languages, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-7727. Internet: gkeating@mail.sdsu.edu

instantiated in the grammars of adult second language (L2) learners. An exemplary case in point is research on the acquisition of abstract grammatical gender. The crux of the debate concerns whether native speakers of English—a language lacking grammatical gender—acquire nativelike knowledge of gender agreement systems in L2s like Spanish, which require that nouns and modifiers match in gender. Some studies find that English speakers' gender representations in the L2 differ qualitatively from those of native speakers (for L2 Spanish, see Franceschina, 2001b, 2005 [chap. 5]; for L2 Dutch, see Sabourin, 2003; Sabourin, Stowe, & de Haan, 2006), whereas others evince no differences between native and nonnative representations (White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004). These and similar findings in other domains[1] fuel an ongoing polemic regarding the availability of Universal Grammar (UG) in SLA. Proponents of the failed functional features hypothesis (Franceschina, 2005; Hawkins & Chan, 1997), on the one hand, posit a critical period for the acquisition of gender and other parameterized functional features. On the other hand, advocates of the full transfer full access hypothesis (Schwartz & Sprouse, 1996) maintain that abstract categorical features available in childhood remain so in adult SLA.

Like much SLA research conducted to date, the aforementioned studies attempt to explain the divergent behavior of highly proficient L2 learners in terms of representational deficits (Klein & Martohardjono, 1999; Sorace, 2003). Hitherto unexplored is whether English speakers' nonnative behavior regarding grammatical gender is due, at least in part, to differences in grammatical processing. Research in L2 sentence processing using online experimental techniques finds that, in comparison to native speakers, L2 learners underutilize syntactic information during sentence comprehension (e.g., Felser & Roberts, 2007; Marinis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). To account for the observed differences between native and nonnative processing, Clahsen and Felser (2006a, 2006b) proposed the shallow structure hypothesis, which asserts that L2 learners compute less detailed syntactic representations during sentence comprehension compared to native speakers and that nativelike processing of morphosyntactic features is restricted to local domains. Does shallow processing characterize L2 learners' abilities to detect gender agreement errors in Spanish? If so, what light can learners' processing behavior shed on the controversy regarding the status of abstract gender in adult L2 grammars?

To answer these questions, the present study examines whether English-speaking learners of Spanish acquire nativelike knowledge of gender agreement and whether L2 learners achieve nativelike abilities to detect gender agreement

errors in Spanish. Specifically, this article reports the results of an eye-tracking experiment that investigates whether L2 learners of Spanish pattern like native Spanish speakers in their sensitivity to gender errors on postnominal adjectives that vary in distance from the nouns they modify.

Grammatical Gender Agreement in Spanish

Although English only marks gender in the pronoun system, all nouns in Spanish are specified for one of two genders: masculine or feminine. Per traditional analyses, the *–o* suffix usually designates masculine nouns (e.g., *libro* "book," *chico* "boy") and the *–a* suffix marks feminine nouns (e.g., *casa* "house," *chica* "girl"),[2] but exceptions abound (see Harris, 1991). Agreement is the criterion that determines whether a language has gender (Corbett, 1991). In Spanish, determiners and adjectives lack intrinsic gender, but they agree in gender (and number) with the nouns they modify. Masculine forms of the adjective usually end in *–o* (*pequeño* "small") and feminine forms in *–a* (*pequeña*), exceptions notwithstanding. So, in determiner phrases (DPs) like *la*/*una casa pequeña* (the/a$_{\text{D-Fem}}$ small$_{\text{Adj-Fem}}$ house$_{\text{N-Fem}}$) and *el/un libro pequeño* (the/a$_{\text{D-Masc}}$ small$_{\text{Adj-Masc}}$ book$_{\text{N-Masc}}$), all constituents agree in gender (and number).

Grammatical gender agreement in Spanish and similar languages poses a unique challenge to English speakers. First, abstract gender is not instantiated in English, so transfer from the first language (L1) is not possible. Second, the gender of inanimate nouns is arbitrary (i.e., lacks semantic value) and redundant (i.e., is encoded on multiple constituents). Research on L2 input processing shows that learners tend not to process forms that are low in communicative value, despite their availability in the input (VanPatten, 1996, 2004). A final and important point to consider is that nouns and adjectives in Spanish are not always juxtaposed; that is, agreement is both a local and a long-distance phenomenon. The data in (1) illustrate this point. In (1a), agreement is local because the noun and adjacent adjective appear in the DP (attributive adjective agreement). Agreement in (1b) and (1c) is a long-distance phenomenon in that modifying adjectives appear outside the DP; in the VP of the matrix clause in the case of predicative adjective agreement (1b) and in the VP of the subordinate clause in the case of (1c).

(1) a. [$_{\text{IP}}$ Una casa pequeña [$_{\text{VP}}$ cuesta mucho en San Francisco.]]
"A small house costs a lot in San Francisco."

b. [$_{\text{IP}}$ La casa [$_{\text{VP}}$ es bastante pequeña y necesita muchas reparaciones.]]
"The house is quite small and needs a lot of repairs."

c. [$_{IP}$ Una casa [$_{VP}$ cuesta menos [$_{CP}$ si [$_{VP}$ es pequeña y necesita reparaciones]]]].
"A house costs less if it is small and needs repairs."

From a processing perspective, the exemplars in (1) impose increasingly greater demands on the parser during sentence comprehension. Computing gender agreement in (1a) only requires a reader/listener to hold the gender feature of the head noun (*casa*) in working memory for a few hundred milliseconds before encountering the modifying adjective (*pequeña*). Computing agreement in (1b) and (1c) requires holding the noun's gender feature in working memory while processing additional propositional content and satisfying grammatical constraints such as theta role and case assignment (Pritchett, 1992). Thus, it is plausible that English-speaking learners of Spanish might show nativelike sensitivity to gender errors in (1a), but not in (1b) and (1c). This is what is predicted by the shallow structure hypothesis, to which I now turn.

Native and Nonnative Processing

The Shallow Structure Hypothesis

The shallow structure hypothesis (SSH) advanced by Clahsen and Felser (2006b) argues that L2 processing is fundamentally different from native processing in that adult L2 learners do not compute full syntactic analyses during sentence comprehension like native speakers do. Instead, L2 learners rely predominantly on lexical, semantic, and pragmatic information. In support of this position, they cited research on the processing of relative clause ambiguities and filler-gap dependencies that shows that end-state L2 learners fail to utilize the structure-based parsing strategies that natives use. Although the SSH is motivated by empirical evidence from L2 learners' processing of ambiguous and syntactically complex (but otherwise grammatical) sentences, an elaborated version of the hypothesis (Clahsen & Felser, 2006a) extends its scope to include the processing of agreement relationships, including violations of noun-adjective and subject-verb agreement. Shallow processing in this domain of the grammar is operationalized in terms of the distance that separates agreeing constituents. Without committing to a particular method of measuring distance, the SSH makes the general prediction that L2 learners can demonstrate nativelike sensitivity to agreement errors, but only in local domains, where local is loosely defined as "between closely adjacent constituents" (Clahsen & Felser, 2006a, pp. 111, 115, 120–121). If we take "closely adjacent" to mean within a phrase, the SSH would predict that highly proficient learners of Spanish can demonstrate nativelike abilities to detect gender errors on adjectives in sentences like

(1a), in which the concatenation of gender features occurs within the DP. If we take nonlocal to mean outside a phrase, the SSH would predict that even highly proficient learners of Spanish will not evince nativelike sensitivity to gender errors on adjectives in sentences like (1b) and (1c). These predictions are open to empirical investigation and lie at the heart of the present study.

Sources of Nonnative Processing

Empirical evidence that shows L2 learners of Spanish detect errors in gender agreement like native speakers in (1a) but not in (1b-c) would support Clahsen and Felser's (2006a) claim for shallow processing in this domain of the grammar. At issue, though, is whether shallow processing for the linguistic phenomenon under consideration is due to representational deficits or to deficits in processing. Clashen and Felser (2006a) claimed that the difference between the full parsing of natives and the shallow parsing of nonnatives lies in the grammar-parser relationship. They adopted the position that the grammar feeds the parser. Full parsing differs from shallow parsing in that the former is guided by the grammar whereas the latter is guided by knowledge of lexical-semantics, pragmatics, and so forth (2006a, p. 117). They delineated two potential sources of shallow processing in the L2: deficits in the L2 grammar and deficits in L2 parsing mechanisms. The first position—the one favored by Clahsen and Felser—claims that full parsing fails when the grammar that feeds the parser diverges in some way from the native grammar. This position coincides with the views of Bley-Vroman (1989) and advocates of the failed functional features hypothesis, who believe that L2 grammars are fundamentally different from native-speaker grammars. The second possibility is that full parsing fails due to the unavailability of appropriate parsing heuristics for processing the L2. This is the stance taken by VanPatten (1996, 2004), Pienemann (1998, 2005), and a host of UG researchers who view processing as a potential source of divergence between native and nonnative speakers (Carroll, 1996, 2001; Gregg, 1996, 2001; Harrington, 2001; Klein & Martohardjono, 1999; Sorace, 2006; Truscott & Sharwood Smith, 2004).

In the context of the current study, the issue boils down to the following question: If English-speaking learners of Spanish show nativelike sensitivity to gender agreement violations in (1a), must they also show sensitivity to gender agreement anomalies in (1b) and (1c) in order to claim that they have abstract gender in their L2 grammars? The position taken in this article is that nativelike sensitivity to gender errors in the DP provides sufficient evidence to rule out divergent representations. In order to demonstrate nativelike knowledge of abstract gender, learners must have (a) knowledge that Spanish adjectives

have gender features in need of checking and (b) a formal feature-checking mechanism capable of performing the operation. If L2 learners demonstrate nativelike sensitivity to gender errors on Spanish adjectives located in the DP, they must have acquired the gender features of Spanish adjectives and must possess the necessary mechanisms to check said features during comprehension. Conversely, failure to show nativelike sensitivity to gender errors in the DP suggests that abstract gender has not been acquired.

If shallow processing in this domain of the grammar is not the result of divergent representations, why might L2 learners fail to evince nativelike sensitivity to gender errors on adjectives located in nonlocal domains? A plausible explanation for the phenomenon under consideration is that nonnative processing is the result of computational complexity, an idea put forth by Sorace (2006), who asked whether shallow processing results when L2 learners' processing resources are taxed beyond their limits, especially in online tasks. As previously stated, what makes gender agreement progressively more difficult to compute in (1a)–(1c) is the number of intervening phrases learners must process—for propositional content and grammatical relationships—between the noun and modifying adjective. In (1b) the noun and adjective are separated by one grammatical boundary: the VP. Processing a verb is no small task, given that verbs have number features that must match those of their subjects, they determine the number and type of arguments in the sentence, and they are responsible for case assignment. Thus, learners have a number of grammatical relationships to juggle before they check the gender feature of the predicative adjective. In (1c) the noun and adjective are separated by three boundaries, two of which are VPs, thus compounding the number of grammatical relationships that need to be processed before the gender feature of the adjective can be checked. In (1a), there are no intervening phrases and therefore no additional propositional or grammatical information to process.

Evidence for Shallow Structures in Gender Agreement Processing

Only a handful of studies have investigated the role of distance in gender agreement processing during comprehension. The existing evidence supports shallow processing but debates its source. In an offline study, Myles (1995) administered written error correction tests to English-speaking learners of French to determine whether correction of gender-based co-reference errors decreased as distance between nouns and adjectives increased. Instructed learners of varying proficiencies read a text that contained gender agreement errors in four structural contexts: (a) within the NP, (b) within the VP, (c) within the CP, and (d) outside the CP. Learners corrected more gender errors in a-type

than b-type sentences and more gender errors in b-type than c-type sentences.[3] Myles attributed the results to a processing deficit. This conclusion is somewhat premature, given that this study did not measure the real-time processing of gender agreement. Furthermore, error correction is a metalinguistic task that may not speak to implicit knowledge of gender. These limitations aside, this study provides indirect evidence for shallow processing in that learners were more accurate on local versus nonlocal agreement.

In an online comprehension study, Sabourin (2003, chap. 6) recorded the event-related potentials (ERPs) of German-, Romance-, and English-speaking learners of Dutch as they read and judged the grammaticality of sentences containing gender agreement in three contexts: (a) between a definite determiner and an adjacent noun, (b) between an indefinite noun and an adjacent adjective, and (c) between a noun and an adjacent relative pronoun, the latter condition representing agreement outside the DP. Pertinent to the current study are the results of the English natives. The accuracy results found that the Dutch natives performed significantly better than the English natives, who performed below chance on each sentence type. The ERP data showed that the English natives did not exhibit nativelike P600 effects in response to gender errors in any condition. In line with the failed functional features hypothesis, Sabourin attributed the performance of the English natives to representational deficits.

In another ERP study involving grammaticality judgments, Tokowicz and MacWhinney (2005) examined whether beginning learners of Spanish (L1 English) were sensitive to gender errors on determiners in determiner-noun sequences (agreement within the DP) in sentences such as *Ellos fueron a *un/una fiesta* "They went to $^{*}a_{\text{D-Masc}}/a_{\text{D-Fem}}$ $\text{party}_{\text{N-Fem}}$." Although judgment accuracy was at chance level for these participants, they exhibited P600 effects in response to gender errors, a result the authors took to indicate that even beginning learners process gender agreement implicitly. However, in the absence of a native-speaker control group, it is difficult to determine to what extent the results are indicative of nativelike knowledge or processing.

In sum, the empirical evidence has only scratched the surface of how distance affects the processing of gender agreement.[4] No study known to this author has systematically investigated the effect of distance on L2 learners' sensitivity to gender errors in Spanish during online sentence comprehension. Given that Spanish nouns can be modified by adjectives located in various syntactic domains, the detection of gender agreement anomalies during online sentence comprehension is an ideal test case for the predictions of the SSH.

The Present Study

This study examines whether English-speaking learners of Spanish pattern like native Spanish speakers in their sensitivity to violations of grammatical gender agreement between Spanish nouns and postnominal adjectives during online sentence comprehension. An eye-tracking study was designed to determine whether L2 learners' processing of grammatical gender agreement is subject to shallow processing, where shallow processing means that L2 learners are sensitive to errors in local domains only. The study posed the following questions:

1. Do English-speaking learners of Spanish evince having acquired gender agreement by patterning like native Spanish speakers in their sensitivity to violations of gender agreement between nouns and modifying adjectives located within the DP?

In this study, agreement within the DP is the baseline criterion for establishing whether gender agreement is acquired. This study hopes to show that English-speaking learners of Spanish, like native speakers, are sensitive to violations of gender agreement when nouns and adjectives are juxtaposed. If this is the case, the assumption is that English speakers have acquired gender agreement, contra the failed functional features hypothesis.

2. If English-speaking learners of Spanish evince having acquired gender agreement, do they continue to pattern like native Spanish speakers in their sensitivity to violations of gender agreement between nouns and adjectives when modifying adjectives are located
    a. in the VP of the matrix clause?
    b. in a subordinate clause?

This study seeks to show that English-speaking learners of Spanish fail to demonstrate nativelike sensitivity to anomalies in gender agreement when adjectives appear outside the DP, a result that would support the SSH. This study attributes such failure to a deficit in processing; namely, that the computational cost incurred to check gender features over distance outweighs available processing resources.

Eye-tracking Technique

The experiment reported here used the technique of recording eye movements during reading to examine participants' sensitivity to violations of gender

agreement. Similar to other online techniques, eye-tracking studies examine reading times on regions of interest (e.g., words or sentence segments) to make inferences about the cognitive processes involved in reading (for detailed accounts about the relationship between eye movements and cognitive processes, see, among others, Frazier & Rayner, 1982, and Rayner, Juhasz, & Pollatsek, 2005). One potential advantage that eye-movement recording holds over other online techniques is that it provides a finer-grained analysis of moment-to-moment language processing. Reading time on words can be divided into various subcomponents, such as the duration of the first fixation, refixations on a word, and regressions to words previously fixated (Carreiras & Clifton, 2004, p. 5). Given that adjectives appear postnominally in Spanish, eye-movement recording is an ideal method for examining the extent to which readers regress to nouns after reading modifying adjectives. If the participants in this study are sensitive to violations of gender agreement during online sentence comprehension, we expect to obtain different reading profiles for grammatical and ungrammatical adjectives—specifically, that ungrammatical adjectives will elicit longer reading times and higher incidences of regressive eye movements relative to grammatical controls.

Method

Participants

The participants were 18 native Spanish speakers and 44 English-speaking learners of Spanish divided into advanced ($n = 12$), intermediate ($n = 14$), and beginning ($n = 18$) groups. Participants belonged to the teaching and research communities at the University of Illinois at Chicago and were paid for their participation. Proficiency was determined by course placement and years of exposure to Spanish. The characteristics of each group are described next and summary statistics for the L2 learners are provided in Table 1.

Native Spanish Speakers

The native Spanish speakers were graduate students, teaching assistants, and lecturers at the university. The participant pool was limited to native speakers who (a) learned Spanish in a Spanish-speaking country outside of the continental United States, (b) received a high school diploma from a high school in a Spanish-speaking country, (c) moved to the United States after the age of 18 ($M = 26.22$ years), and (d) indicated using Spanish at home. The majority of the native speakers also completed their bachelor's degree in a Spanish-speaking country.

Table 1 Background information for the L2 learners of Spanish

Group	Years of high school Spanish	Years of college Spanish[a]	Years lived abroad
Advanced ($n = 12$)			
Mean	4.00	—	12 months
Range	3–5	4–7	1–48 months
SD	0.43	—	13 months
Intermediate ($n = 14$)			
Mean	3.46	—	—
Range	1.5–4	2–3 years	—
SD	0.93	—	—
Beginner ($n = 18$)			
Mean	2.42	1 year	—
Range	0–4	—	—
SD	1.06	—	—

[a]These data were not collected from participants but were estimated based on their current enrollment at the university.

Advanced L2 Learners

The advanced learners of Spanish were graduate students pursuing an M.A. or Ph.D. in Hispanic literature or linguistics, or an M.A. in Educational Leadership with emphasis in Spanish Teacher Education in the College of Education. Graduate students pursuing teaching and research careers in Spanish were chosen because they were most likely to be end-state learners. In order to participate in the study, an advanced learner had to (a) be a native English speaker, (b) speak no home languages other than English, (c) have a B.A. or equivalent in Spanish, and (d) have learned Spanish as an L2 in adulthood. All advanced participants reported immersion exposure in a Spanish-speaking country. The demographics of the available participant pool thwarted recruitment of a larger sample of advanced learners. Given the university's location in Chicago, a cosmopolitan polyglot area, the graduate programs attracted students of various language backgrounds. In addition, Chicago is home to a large Spanish-speaking population estimated at 26% Hispanic. Many potential participants were either (a) heritage Spanish speakers or (b) had studied another language (with or without gender) before studying Spanish. These participants were excluded in order to ensure that all advanced participants were native English speakers who learned Spanish as a *second* language. When high school and postsecondary (including graduate school) exposure were taken together, advanced learners had between 8 and 11 years of exposure to Spanish.

Intermediate L2 Learners

The intermediate learners were recruited from third-year Spanish courses, including introductory courses in Spanish linguistics, Spanish literature, and advanced composition and conversation. Gender agreement was not the focus of any lesson or activity in the third-year curriculum. Participation was limited to third-year students who (a) were native English speakers, (b) spoke no home languages other than English, (c) learned Spanish as a second language in adulthood, and (d) had not lived abroad in a Spanish-speaking country. Similar to the advanced learners, demographic variables limited the potential participant pool. First, the vast majority of third-year Spanish students were heritage Spanish speakers. Second, many of the available native English speakers had already studied abroad. Mixing students with and without exposure abroad was deemed unsuitable. Intermediate learners had approximately 5–6 years of exposure to Spanish (high school and college combined).

Beginning L2 Learners

Beginning learners were enrolled in Spanish 103, the third course in a four-course sequence that undergraduate students take to complete a general education requirement in foreign language. Gender agreement was not the focus of any lesson or activity in the 103 curriculum. Participants at this level were suitable for this study because they were completing their study of an introductory-level Spanish textbook and had been exposed to hundreds of examples of gender agreement. In order to be included in the present study, beginning learners had to meet the same criteria as the intermediate speakers. Beginners had approximately 3 years of exposure to Spanish (high school and college combined).

Materials and Design

The materials for the online reading task consisted of 36 sentences involving gender agreement between a noun located in a sentence-initial DP and a modifying adjective located in one of three syntactic domains, as illustrated previously in (1): (a) within the DP, immediately following the noun; (b) in the VP, preceded by the copular verb *es* "is" and the adverb *bastante* "quite, very"; and (c) in a subordinate clause introduced by *si* "if" or *cuando* "when." There were 12 sentences in each distance condition, half of which were grammatical (i.e., the noun and modifying adjective agreed in gender) and half were ungrammatical (i.e., the noun and modifying adjective did not agree in gender). Noun gender (masculine vs. feminine) was matched across conditions. The 36 critical sentences were mixed with 48 filler sentences that tested other aspects

of grammatical processing (tense and aspect choice). A complete list of the target stimuli is provided in the Appendix.

Several word-level considerations guided stimuli design. First, all critical nouns and adjectives ended in the suffixes commonly associated with gender (*–o* for masculine and *–a* for feminine). This was done to bias for the best: If learners process gender agreement during sentence comprehension, errors in gender agreement should be most salient when gender is marked morphologically on nouns and adjectives. Second, this study only included nouns with grammatical gender in order to eliminate confounding effects that might arise from the use of nouns coded for semantic and grammatical gender. Third, given that research in lexical access finds that L2 learners access cognates faster than noncognates (e.g., Dufour & Kroll, 1995), all target nouns and adjectives were noncognates to ensure that learners' eye movements were not influenced by the L1. Fourth, to ensure that eye movements were not the result of learners—particularly the beginners—not knowing the meaning of target words, a Spanish-English word translation test was administered to a random sample of Spanish 103 students prior to stimuli design. Stimuli were created using nouns and adjectives known by at least 90% of the participants, as well as nouns recently studied and tested in the 103 curriculum. Controlling for terminal morphology, type of gender (grammatical vs. semantic), cognate status, and word familiarity made it impossible to use adjectives of equal length.[5] Control was imposed by counterbalancing the placement of adjectives of different lengths across experimental conditions. Finally, the word selection criteria necessitated the recycling of nouns and adjectives to create 36 sentences.

Each critical and filler sentence was followed by an English comprehension sentence. Participants had to judge whether the English sentence expressed the same meaning as the Spanish sentence. Half of the English sentences were accurate renditions and half were not. Inaccurate renditions were created by substituting a word from the original sentence with a new word that changed its meaning but that did not draw unwanted attention to the critical nouns and adjectives. A sentence comprehension task was deemed more appropriate than a grammaticality judgment task because the nonnative Spanish speakers in this study were instructed learners who likely knew the explicit rules for making gender agreement. A grammaticality judgment task might alert participants to the target structure, despite the filler sentences. Accuracy results of the sentence comprehension task were used to ensure that participants were attending to the stimuli, but otherwise it had no bearing on the outcome of the study.

Apparatus and Procedure

While reading, participants' eye movements were recorded with an EyeLink II head-mounted tracking device designed by SR Research. Viewing was binocular, with eye movement recorded from the right eye. The tracking device was interfaced with a PC that controlled stimulus display and data storage. The sentences were presented individually on a high-resolution 17-in. display monitor. Sentences appeared in a single line in black against a light gray background using normal uppercase and lowercase letters.

Each Spanish test sentence and English comprehension item was preceded by a fixation target that occupied the position of the sentence initial character. Participants were instructed to look at the fixation target while pressing the "Advance" button on a hand-held game controller to display the sentence. Participants were asked to read the Spanish sentences for meaning and to indicate whether the English sentences expressed the same general meaning as the Spanish sentences by pressing YES or NO on the hand-held controller. Participants were told that English items were not word-for-word translations of the Spanish sentences. Trials timed out after 15 s without a response from the participant. Before moving to the test items, participants read the directions and two practice sentences with the headset in place. The stimuli were pseudorandomized into five different stimulus lists so that all participants saw the stimuli in one of five orders. Participants were randomly assigned to one of the five lists. The completion of paperwork, equipment setup and calibration, and the online sentence comprehension task took approximately 45 min.

Methods of Analysis

Although eye movements were recorded on every word in each sentence, analyses were limited to data obtained for the primary region of interest: the critical adjective. This study reports three standard reading time measures (in milliseconds): first fixation duration, first-pass reading time, and total reading time. *First fixation duration* refers to the duration of the first fixation made on the adjective. Readers often fixate a region more than once before leaving it. *First-pass reading time* (also called gaze duration) refers to the sum of all fixations made on the adjective from first entering it until first leaving it with a saccade in any direction. (If an adjective receives only one fixation, first-pass time equals first fixation duration.) *Total reading time* refers to the sum of all fixations made on the adjective, including fixations from second and subsequent passes (e.g., secondary fixations resulting from regressions back to the adjective).

In addition, this study reports the *probability of a first-pass regression*, which refers to the percentage of participants' first-pass fixations on the

adjective that resulted in a leftward movement out of the adjective (i.e., a regression initiated from the adjective, before fixating a region after the adjective). Analysis of participants' data files revealed that many regressions were launched from a point shortly after the critical adjective (usually one to three words afterward). These "delayed" regressions are likely due the fact that information about gender is encoded on the last letter of the adjective. Readers may process the gender of the adjective after the eyes have initiated ballistic movements to subsequent regions. As noted by two anonymous reviewers, regressions initiated from the critical adjective (first-pass regressions) and those launched from words after the adjective (delayed regressions) may be indicative of different processing behaviors and should be examined separately. Accordingly, this study reports the probability of a first-pass regression and the *probability of a delayed regression*. Furthermore, given that all groups made both types of regression and given that both types may be indicative of anomaly detection, this study also reports the probability of making either kind of regression (*total regression proportion*).

In addition to distinguishing regressions on the basis of launch site, this study was interested in the target or final landing site of regressions. Of particular interest was whether nouns were the target of first-pass and delayed regressions. Given that an initial regression is often part of a chain of successive backward eye movements, the regression target was determined to be the leftmost region fixated on a regression path (Braze, Shankweiler, Ni, & Conway Palumbo, 2002). First-pass regressions were subdivided into two categories: those with final landing sites on the head noun (possibly including the determiner) and those with final landing sites on a word appearing between the noun and the adjective, a possibility limited to sentences in the VP and CP conditions. The following are sample fixation sequences of each type:

- first-pass regression to noun: **Una casa pequeño casa pequeño cuesta* . . .
- first-pass regression to other word: **La casa es bastante pequeño bastante pequeño* . . .

Delayed regressions were subdivided into three categories: those with a final landing site on the critical adjective, those with a landing site on the critical adjective followed by a regression path to the noun (possibly including the determiner), and those that ended on the head noun (possibly including the determiner) without making an intermediate stop on the adjective. The following are sample fixation sequences of each type:

- delayed regression to adjective only: **Una casa pequeño cuesta pequeño cuesta mucho . . .*
- delayed regression to adjective and noun: **Una casa pequeño cuesta pequeño casa . . .*
- delayed regression to noun only: **Una casa pequeño cuesta casa pequeño cuesta . . .*

Results

Comprehension Accuracy

Responses to the comprehension sentences following each stimulus sentence in the eye-tracking study were tallied and mean scores were calculated for each participant. Failures to select a response and inappropriate responses (e.g., selecting the advance button instead of the YES/NO buttons) were scored as missing values and means were calculated on the remaining sentences. Such responses accounted for less than 1% of the data. Analysis of individual items revealed that 5 of the 36 items were missed by at least 50% of participants (range: 50–90%). These items were intended to be inaccurate renditions of the Spanish sentences, but the critical item in the English translation proved too similar in meaning to the Spanish word to elicit the intended NO response and therefore were excluded from analysis. Mean scores calculated on the remaining 31 items were high for all groups, an indication that participants attended to the task and understood the stimuli: native (85.0%); advanced (84.1%); intermediate (88.3%); beginning (83.5%). A univariate ANOVA on the mean accuracy scores revealed no significant differences among the groups, $F(3, 58) = 1.275$, $p = .292$, $\eta_p^2 = .062$.

Reading Times and Regression Proportions

Mean reading times and regression proportions were calculated for each participant on the basis of correctly answered trials. Error data (13.3%) and missing data due to track loss, word skipping, and sentence skipping (5.4%) accounted for 18.7% of the data. Each group's mean reading times on the critical adjectives in each condition appear in Table 2. Table 3 presents each group's regression frequencies broken down by launch site (first pass vs. delayed) and Table 4 provides each group's mean first-pass, delayed, and total regression proportions.

To determine whether there were differences in processing the experimental sentences among the groups, the data for each dependent variable were initially submitted to a mixed three-way ANOVA with the factors Distance (DP, VP, CP) and Grammaticality (G, UG) as within-subjects factors and Group

Table 2 Mean first fixation, first-pass, and total reading times (ms) and standard deviations (in parentheses) on the adjectives in each condition

		Adjectives in DP			Adjectives in VP			Adjectives in CP		
Measurement	N	G	UG	Diff.	G	UG	Diff.	G	UG	Diff.
First Fixation										
Native	18	202	212	10	201	225	24	212	204	−8
		(35)	(41)		(26)	(52)		(37)	(34)	
Advanced	12	223	254	31	251	256	5	230	215	−15
		(48)	(50)		(57)	(40)		(27)	(41)	
Intermediate	14	234	251	17	250	251	1	243	219	−24*
		(34)	(52)		(43)	(38)		(53)	(33)	
Beginning	18	300	303	3	301	276	−25	272	242	−30
		(102)	(108)		(91)	(70)		(73)	(51)	
First-Pass Time										
Native	18	247	262	15	219	274	55*	250	282	32
		(80)	(69)		(39)	(92)		(81)	(87)	
Advanced	12	305	303	−2	309	299	−10	295	277	−18
		(55)	(59)		(51)	(58)		(91)	(88)	
Intermediate	14	301	318	17	307	318	11	310	309	−1
		(56)	(65)		(68)	(64)		(70)	(85)	
Beginning	18	422	451	29	422	400	−22	356	310	−46
		(92)	(147)		(115)	(109)		(114)	(70)	
Total Time										
Native	18	398	702	304*	283	431	148*	355	552	197*
		(142)	(323)		(99)	(189)		(124)	(223)	
Advanced	12	647	903	256*	500	620	120	540	608	68
		(127)	(218)		(203)	(211)		(189)	(222)	
Intermediate	14	686	707	21	557	544	−13	541	686	145*
		(232)	(241)		(244)	(135)		(174)	(232)	
Beginning	18	917	884	−33	745	755	10	650	665	15
		(302)	(249)		(220)	(332)		(245)	(170)	

Note. Difference = reading times on UG adjectives minus reading times on G adjectives.
*Significant at the .05 level in a paired-samples *t*-test (two-tailed).

(native, advanced, intermediate, beginning) as a between-subjects factor. The result of interest is the Group and Grammaticality interaction, which, if significant, indicates that the grammaticality of the adjective affects reading times and regressions differently for the different groups. The Group and Grammaticality interaction was reliable in the analyses of the total reading times, $F(3, 58) = 8.785$, $p < .001$, $\eta_p^2 = .312$, the delayed regression proportions,

Table 3 Frequencies of first-pass and delayed regressions for each group and condition

	Adjectives in DP		Adjectives in VP		Adjectives in CP	
	G	UG	G	UG	G	UG
Native	16	60	16	39	33	56
First-pass	9 (56%)	16 (27%)	9 (56%)	12 (31%)	9 (27%)	17 (30%)
Delayed	7 (44%)	44 (73%)	7 (44%)	27 (69%)	24 (73%)	39 (70%)
Advanced	21	36	18	19	41	36
First-pass	7 (33%)	17 (47%)	5 (28%)	3 (16%)	14 (34%)	11 (31%)
Delayed	14 (67%)	19 (53%)	13 (72%)	16 (84%)	27 (66%)	25 (69%)
Intermediate	19	31	15	17	38	37
First-pass	7 (37%)	18 (58%)	4 (27%)	3 (18%)	10 (26%)	5 (14%)
Delayed	12 (63%)	13 (42%)	11 (73%)	14 (82%)	28 (74%)	32 (86%)
Beginning	38	49	46	46	55	43
First-pass	23 (61%)	29 (59%)	32 (70%)	24 (52%)	18 (33%)	12 (28%)
Delayed	15 (39%)	20 (41%)	14 (30%)	22 (48%)	37 (67%)	31 (72%)

$F(3, 58) = 7.032$, $p < .001$, $\eta_p^2 = .267$, and the total regression proportions, $F(3, 58) = 14.083$, $p < .001$, $\eta_p^2 = .421$. The interaction was not significant in the analysis of the first fixation durations, $F(3, 58) = 1.925$, $p = .136$, $\eta_p^2 = .091$, but approached significance in the analyses of the first-pass reading times, $F(3, 58) = 2.429$, $p = .074$, $\eta_p^2 = .112$, and the first-pass regression proportions, $F(3, 58) = 2.523$, $p = .067$, $\eta_p^2 = .115$. In the latter three cases in which the interaction failed to reach significance, the main effect of grammaticality also failed to reach significance, although it approached significance in the analysis of the first-pass regression proportions, $F(1, 58) = 3.288$, $p = .075$, $\eta_p^2 = .054$. Taken together, these results suggest that first fixation durations, first-pass reading times, and first-pass regressions were not sensitive enough to reveal grammaticality effects with these participants and stimuli.

Given that the three-way ANOVAs indicated interactions between Group and Grammaticality, subsequent analyses consisted of 4 (Group) × 2 (Grammaticality) ANOVAs conducted separately on the data for each of the three distance conditions. Considering that the primary aim of this study was to determine whether nonnative Spanish speakers are sensitive to gender agreement errors in each distance condition, all two-way ANOVAs were followed by paired-samples *t*-tests (one for each group) conducted as planned comparisons—that is, regardless of whether a significant Group and Grammaticality interaction obtained (although in most cases it did). Analyses were limited to total reading times, delayed regression proportions, and total regression proportions, as these

Table 4 Mean first-pass, delayed, and total regression proportions (expressed as %) and standard deviations (in parentheses)

		Adjectives in DP			Adjectives in VP			Adjectives in CP		
Measurement	N	G	UG	Diff.	G	UG	Diff.	G	UG	Diff.
First-pass										
Native	18	9	16	7*	9	16	7	9	20	11*
		(14)	(20)		(16)	(27)		(15)	(23)	
Advanced	12	10	28	18*	8	5	−3	21	19	−2
		(16)	(19)		(13)	(10)		(26)	(22)	
Intermediate	14	10	23	13	5	4	−1	13	6	−7
		(19)	(26)		(10)	(9)		(16)	(11)	
Beginning	18	28	30	2	31	26	−5	19	13	−6
		(21)	(19)		(31)	(23)		(24)	(20)	
Delayed										
Native	18	8	46	38*	7	31	24*	26	44	18*
		(17)	(21)		(9)	(24)		(22)	(31)	
Advanced	12	22	33	11	20	29	9	44	42	−2
		(14)	(23)		(14)	(24)		(26)	(34)	
Intermediate	14	18	17	−1	13	19	6	41	41	0
		(13)	(15)		(13)	(20)		(34)	(25)	
Beginning	18	18	22	4	15	24	9	41	34	−7
		(16)	(17)		(17)	(20)		(25)	(22)	
Total										
Native	18	18	63	45*	16	47	31*	35	64	29*
		(22)	(27)		(20)	(26)		(26)	(26)	
Advanced	12	33	60	27*	27	34	7	65	60	−5
		(22)	(27)		(20)	(25)		(26)	(35)	
Intermediate	14	27	40	13	18	24	6	54	47	−7
		(22)	(28)		(21)	(20)		(36)	(27)	
Beginning	18	46	51	5	46	50	4	59	47	−12
		(21)	(24)		(28)	(29)		(25)	(30)	

Note. Difference = regression proportions in UG sentences minus regression proportions in G sentences.

*Significant at the .05 level in a paired-samples *t*-test (two-tailed).

were the only measures for which the three-way ANOVAs revealed significant Group and Grammaticality interactions.

Adjectives in the DP

Analysis of the total reading times on adjectives in the DP yielded main effects of Group, $F(3, 58) = 10.578$, $p < .001$, $\eta_p^2 = .354$ and Grammaticality,

$F(1, 58) = 12.258, p = .001, \eta_p^2 = .174$, and a significant Group and Grammaticality interaction, $F(3, 58) = 5.022, p = .004, \eta_p^2 = .206$. Paired-samples *t*-tests indicated that ungrammatical adjectives incurred significantly longer reading times than grammatical adjectives for the native speakers, $t(1, 17) = 4.832$, $p < .001$, and the advanced learners, $t(1, 11) = 4.150, p = .002$.

The analysis of the delayed regression proportions revealed a marginally significant effect of Group, $F(3, 58) = 2.642$, $p = .058$, $\eta_p^2 = .120$, a reliable effect of Grammaticality, $F(1, 58) = 16.147$, $p < .001$, $\eta_p^2 = .218$, and a significant Group and Grammaticality interaction, $F(3, 58) = 8.556$, $p < .001$, $\eta_p^2 = .307$. Tests of simple effects revealed one difference: native speakers launched more delayed regressions in response to ungrammatical adjectives compared to grammatical controls, $t(1, 17) = 5.604, p < .001$.

Analysis of the total regression proportions yielded a significant main effect of Grammaticality, $F(1, 58) = 45.167, p < .001, \eta_p^2 = .438$, and a reliable Group and Grammaticality interaction, $F(3, 58) = 7.680$, $p < .001$, $\eta_p^2 = .284$, but no reliable effect of Group, $F(3, 58) = 1.700$, $p = .177$, $\eta_p^2 = .081$. Paired-samples *t*-tests revealed that native speakers and advanced learners initiated more total regressions in response to ungrammatical adjectives compared to grammatical adjectives: $t(1, 17) = 6.463, p < .001$ and $t(1, 11) = 2.769, p = .018$, respectively.

Summary

The results presented thus far indicate that advanced learners are sensitive to gender errors on adjectives located within the DP because they, like the native speakers, spend more time reading ungrammatical adjectives relative to grammatical adjectives and initiate more total regressions in response to ungrammatical adjectives compared to grammatical ones. Intermediate and beginning learners did not perform like the native speakers on any of the measures. To the extent that gender agreement processing within the DP is a valid litmus test for the acquisition of abstract gender, these data suggest that the advanced learners of Spanish have acquired gender, but the intermediate and beginning learners have not.

Adjectives in the VP

Analysis of total reading times on adjectives in the VP revealed main effects of Group, $F(3, 58) = 12.474$, $p < .001$, $\eta_p^2 = .392$, and Grammaticality, $F(1, 58) = 6.949$, $p = .011$, $\eta_p^2 = .107$, and a marginally significant Group and Grammaticality interaction, $F(3, 58) = 2.639, p = .058, \eta_p^2 = .120$. Planned-comparisons *t*-tests indicated that native speakers had longer reading

times on ungrammatical adjectives compared to grammatical adjectives, $t(1, 17) = 3.875$, $p = .001$. The difference for advanced learners was approaching significance, $t(1, 11) = 1.862$, $p = .089$.

Analysis of the delayed regression proportions yielded a main effect of Grammaticality, $F(1, 58) = 14.090$, $p < .001$, $\eta_p^2 = .195$, but no main effect of Group ($p = .500$) and no Group and Grammaticality interaction ($p = .181$). Planned-comparisons *t*-tests revealed a reliable difference between the ungrammatical and grammatical adjectives for the native speakers only, $t(1, 17) = 3.858$, $p = .001$.

Analysis of the total regression proportions yielded main effects of Group, $F(3, 58) = 4.575$, $p = .006$, $\eta_p^2 = .191$, and Grammaticality, $F(1, 58) = 16.882$, $p < .001$, $\eta_p^2 = .225$, and a significant Group and Grammaticality interaction, $F(3, 58) = 5.552$, $p = .002$, $\eta_p^2 = .223$. In the *t*-tests, only the native speakers launched significantly more total regressions in response to ungrammatical adjectives compared to grammatical adjectives, $t(1, 17) = 4.897$, $p < .001$.

Summary

The data presented in this subsection support the claim that distance plays a role in gender agreement processing. When modifying adjectives are located in the VP of the matrix clause, the L2 learners of Spanish do not show the same sensitivity to violations of gender agreement that native speakers do.

Adjectives in the CP

Analysis of the total reading times on adjectives in the subordinate CP revealed main effects of Group, $F(3, 58) = 5.139$, $p = .003$, $\eta_p^2 = .210$, and Grammaticality, $F(1, 58) = 12.464$, $p = .001$, $\eta_p^2 = .177$, but no Group and Grammaticality interaction, $F(3, 58) = 2.081$, $p = .113$, $\eta_p^2 = .097$. Planned-comparisons *t*-tests revealed a reliable difference for the native speakers, $t(1, 17) = 3.772$, $p = .002$, and, unexpectedly, a reliable difference for the intermediate learners, $t(1, 13) = 2.508$, $p = .026$. A closer examination of the intermediate data revealed that a few participants had unusually large total reading times on the adjective in item 31. When this item was removed from analysis, the difference was still significant for the native speakers, $t(1, 17) = 2.321$, $p = .033$, but not for the intermediate learners, $t(1, 13) = .130$, $p = .899$.

Analysis of the delayed regression proportions revealed no main effects or interactions (*p*-values > .160). Planned-comparisons *t*-tests revealed a reliable difference between the ungrammatical and grammatical sentences for the native speakers, $t(1, 17) = 2.454$, $p = .025$.

Analysis of the total regression proportions yielded a significant Group and Grammaticality interaction, $F(3, 58) = 5.860, p = .001, \eta_p^2 = .233$, but no main effect of Group ($p = .455$) or Grammaticality ($p = .782$). A reliable difference between the ungrammatical and grammatical adjectives obtained for the native speakers only, $t(1, 17) = 4.365, p < .001$.

Summary

The results of gender agreement processing across clause boundaries parallel those found for gender agreement processing across phrase boundaries, in that Spanish natives continue to show sensitivity to gender agreement violations on adjectives located outside the matrix clause, but the L2 learners do not.

Landing Sites for Regressions

The distribution of landing sites for each group's first-pass and delayed regressions appear in Table 5. Of interest to this study is the extent to which readers' regressive eye movements result in refixations of the head noun in ungrammatical sentences. Data from the first-pass and delayed regressions suggest that the likelihood of regressing to the noun is an artifact of the distance conditions themselves, not an indication of a need to recheck the gender of the head noun to resolve an agreement anomaly. The data for the first-pass regressions show that when agreement occurs within the DP, 100% of participants' regressions result in refixations of the head noun. This is not surprising, given that the noun in DP-agreement sentences is the nearest region to the left of the critical adjective. However, when ungrammatical adjectives are located in the VP and the CP, all groups are more likely to regress to the word or two immediately preceding the adjective (67–100% of the regressions in the VP condition and 92–100% of the regressions in the CP condition). The pattern of results is similar for regressions made in grammatical sentences. These percentages should be interpreted with caution, given that the total number of regressions made by each group is, in some cases, quite small. Nevertheless, the data suggest that first-pass regressions out of the adjective are not likely to result in refixations of the head noun unless the noun and adjective are juxtaposed.

The data for the delayed regressions yielded a similar pattern, in that regressions to the noun were more likely in the DP condition compared to the VP and CP conditions. In the DP condition, 50–63% of participants' delayed regressions in ungrammatical sentences resulted in refixations of the head noun (Adj + N and Noun-only regressions combined). However, in both the VP and the CP conditions, participants were more likely to regress to the ungrammatical adjective only—that is, without continuing the regression path back to the noun

Table 5 Distribution of landing sites for the first-pass and delayed regressions in each condition

	Adjectives in DP		Adjectives in VP		Adjectives in CP	
	G	UG	G	UG	G	UG
First-Pass						
Native	9	16	9	12	9	17
Noun	9 (100%)	16 (100%)	3 (33%)	3 (25%)	2 (22%)	0 (0%)
Other	0 (0%)	0 (0%)	6 (67%)	9 (75%)	7 (78%)	17 (100%)
Advanced	7	17	5	3	14	11
Noun	7 (100%)	17 (100%)	2 (40%)	0 (0%)	6 (43%)	0 (0%)
Other	0 (0%)	0 (0%)	3 (60%)	3 (100%)	8 (57%)	11 (100%)
Intermediate	7	18	4	3	10	5
Noun	7 (100%)	18 (100%)	1 (25%)	1 (33%)	2 (20%)	0 (0%)
Other	0 (0%)	0 (0%)	3 (75%)	2 (67%)	8 (80%)	5 (100%)
Beginning	23	29	32	24	18	12
Noun	23 (100%)	29 (100%)	6 (19%)	4 (17%)	0 (0%)	1 (8%)
Other	0 (0%)	0 (0%)	26 (81%)	20 (83%)	18 (100%)	11 (92%)
Delayed						
Native	7	44	7	27	24	39
Adj. only	3 (43%)	22 (50%)	5 (71%)	22 (82%)	18 (75%)	28 (72%)
Adj. + N	1 (4%)	11 (25%)	0 (0%)	2 (7%)	2 (8%)	6 (15%)
Noun only	3 (43%)	11 (25%)	2 (29%)	3 (11%)	4 (17%)	5 (13%)
Advanced	14	19	13	16	27	25
Adj. only	4 (29%)	7 (37%)	8 (62%)	12 (75%)	15 (56%)	16 (64%)
Adj. + N	3 (21%)	7 (37%)	0 (0%)	0 (0%)	9 (33%)	3 (12%)
Noun only	7 (50%)	5 (26%)	5 (38%)	4 (25%)	3 (11%)	6 (24%)
Intermediate	12	13	11	14	28	32
Adj. only	4 (33%)	6 (46%)	9 (82%)	10 (72%)	20 (72%)	23 (72%)
Adj. + N	7 (59%)	1 (8%)	0 (0%)	1 (7%)	4 (14%)	5 (15%)
Noun only	1 (8%)	6 (46%)	2 (18%)	3 (21%)	4 (14%)	4 (13%)
Beginning	15	20	14	22	37	31
Adj. only	6 (40%)	8 (40%)	11 (79%)	14 (64%)	31 (84%)	26 (84%)
Adj. + N	4 (27%)	5 (25%)	3 (21%)	3 (13%)	5 (14%)	4 (13%)
Noun only	5 (33%)	7 (35%)	0 (0%)	5 (23%)	1 (2%)	1 (3%)

(64–82% of regressions in the VP condition and 64–84% of the regressions in the CP condition). The pattern was similar for regressions launched in grammatical sentences. In sum, regressions made from words after the adjective are more likely to result in refixations of the head noun when the noun and adjective are adjacent and more likely to result in refixations of the ungrammatical

adjective when the adjective and noun are not juxtaposed. Regressing to the noun may be a matter of convenience.

Discussion

The purpose of this study was (a) to determine whether English-speaking learners of Spanish show evidence of having acquired gender agreement and (b) to investigate whether L2 learners are sensitive to gender errors on postnominal adjectives in all of the contexts that native speakers are during online sentence comprehension. The results of the eye-tracking experiment are the following:

- Advanced learners of Spanish, like native Spanish speakers, are sensitive to gender agreement violations on Spanish adjectives located within the DP (a local domain). Therefore, gender agreement is acquirable in adult SLA.
- Beginning and intermediate learners of Spanish are not sensitive to gender errors on Spanish adjectives located within the DP, which suggests that gender agreement is acquired late.
- Regarding distance, native Spanish speakers are not affected by the distance that separates nouns and modifying adjectives, but advanced learners of Spanish are.

The results prompt the following conclusions: (a) Nativelike knowledge of abstract gender is acquirable in adult SLA and (b) nativelike processing of gender agreement violations in the L2 may not be achievable. The findings have implications for theories of representation and processing, as described in the following subsection.

Representation and Abstract Gender

The results of this study argue against the strong version of the failed functional features hypothesis proposed by Franceschina (2005), which claims that abstract gender cannot be acquired in adulthood if it is not present in the L1. If English natives are limited to the categorical representations of their L1 (i.e., a functional lexicon lacking gender), they should not pattern like Spanish natives in their sensitivity to gender errors on adjectives within the DP, regardless of their level of proficiency. Clearly, this is not the case, as the advanced learners display similar behavioral responses to gender errors that Spanish natives do. If parsing is guided by the grammar, the performance of the advanced learners is difficult to explain in the absence of abstract gender, especially considering the nature of the online task. In lieu of judging the grammaticality of the target sentences, participants read and responded to the critical sentences for their

propositional content. Despite being oriented to read for meaning, the advanced learners detected gender errors on adjectives in the DP during online sentence comprehension.

The results of the advanced learners contrast with those of prior research on advanced English-speaking learners of L2s with gender (Franceschina, 2001b, 2005; Sabourin, 2003; Sabourin et al., 2006). Prior studies found that English speakers perform significantly worse than native speakers on speech production tasks and online and offline comprehension tasks. One reason for the difference in results between the studies could be the types of learners studied. The advanced learners in this study were atypical language learners. In addition to having majored in Spanish and lived or studied in a Spanish-speaking country, most of the advanced learners were pursuing academic careers in Spanish. Their exposure to Spanish included extensive reading for academic purposes (e.g., literature and literary criticism) and frequent interaction with native speakers. Furthermore, acceptance into the graduate programs in Spanish required advanced writing and speaking skills. Much of the prior research examined naturalistic learners without formal instruction or classroom learners who did not have the amount and/or type of exposure to the L2 that the learners in this study had.

Looking at the three cross-sections of L2 learner data, the findings are more in line with the full transfer full access hypothesis, which asserts that the L1 grammar is the initial state of SLA—a claim that coincides with the initial state proposal of the failed functional features hypothesis—but that parameter-setting is ultimately possible in adulthood. The online reading behavior of the beginning and intermediate learners suggests they were not sensitive to gender agreement anomalies in the DP. This is to be expected if low-proficiency learners of Spanish transfer a DP that is unspecified for gender; hence the lack of gender features to check during language use. Crucially, the performance of the advanced learners indicates L2 learners do have access to abstract categorical features not instantiated in the L1 grammar.

Although the results of this study indicate that gender agreement is acquired late for most learners, the same conclusion as that of White et al. (2004) was reached; namely, that L2 learners acquire the gender features of Spanish adjectives and show gender agreement within the DP.

Gender Agreement and Shallow Processing

The results of the eye-tracking study indicate that distance is a key determinant of difficulty in the processing of noun-adjective gender agreement in L2 Spanish, just as it was for the English-speaking learners of French in Myles's (1995)

offline study. The present findings support the SSH in that the advanced learners were only nativelike in their sensitivity to gender errors in local domains. Under the assumption that learners who display nativelike sensitivity to gender violations within the DP have abstract gender in their L2 grammars, this article attributes nonnative sensitivity to gender agreement anomalies outside the DP to a deficit in processing, where *deficit* means that L2 learners may not have the processing resources necessary to hold information about gender in working memory while processing material that intervenes between nouns and adjectives. As an anonymous reviewer suggested, processing in the L2 might place burdens on the L2 learner that are not faced by the native speaker, one example being difficulties in lexical access. The reading time data in Table 2 support this claim. Although there is debate in the eye-tracking literature as to whether first fixation duration or first-pass reading time is the best indicator of lexical access (see Rayner, 1998), in either case, regardless of experimental condition, the native speakers had the fastest first fixation and first-pass reading latencies, followed by the advanced and intermediate learners, who were similar to each other, followed by the beginning learners.

Positing a processing deficit raises a number of interesting questions regarding ultimate attainment in the L2. Sorace (2006) asked whether L2 learners' reliance on shallow processing decreases over time and whether there is a qualitative difference between the processing of advanced and near-native speakers. Given that gender agreement is acquired late, we do not see a gradual decline in shallow processing across the three groups of L2 learners, as would be the case if the intermediate learners were sensitive to gender violations in the DP and the advanced learners were sensitive to gender errors in both the DP and the VP, for example. It seems necessary, as Sorace suggested, to compare the processing of advanced and near-native speakers of Spanish to examine this effect. In the absence of a proficiency test to distinguish advanced and near-native speakers, the present study cannot speak to such differences, but some research suggests that this might be an important distinction to make. In a study on the processing of subject-object ambiguities in German, Hopp (2006) found, contra the SSH, that near-native (but not advanced) learners of German were identical to native speakers not only in their knowledge of grammaticality but also in their processing behavior.

Finally, the results of this study indicate that nativelike processing was not achieved by the advanced learners in this study. Can English-speaking learners of Spanish ever show nativelike sensitivity to gender violations outside the DP? Further research is needed to answer this question, but individual data from the advanced learners in this study suggest that nativelike processing may

Table 6 Individual differences in mean total reading times on the adjectives (UG – G) in each distance condition for the advanced learners of Spanish ($n = 12$)

Research ID	Adjectives in DP	Adjectives in VP	Adjectives in CP
65	−24	27	−194
45	516	−139	−37
69	278	−145	−106
70	205	75	−12
72	125	39	−4
41	211	138	−76
42	258	146	−248
39	220	117	161
43	803	194	331
46	127	297	454
48	163	703	185
38	181	−11	356

be achievable. To determine whether some advanced learners were sensitive to gender errors outside the DP, mean total reading times on grammatical adjectives were subtracted from mean total reading times on ungrammatical adjectives in each distance condition for each participant. Total reading time was selected because it includes all fixations made on the adjectives. The mean differences for each participant and condition appear in Table 6. Of the 12 advanced learners, 11 had mean differences in the DP of at least 125 ms (range: 125–803). Of the 11, 6 also had mean differences in the VP of at least 117 ms (range: 117–703). Of these 6, 4 had mean differences in the CP of at least 161 ms (range: 161–454). The data also suggest an implicational hierarchy. With the exception of participant 38, all learners who had longer reading times on ungrammatical adjectives in the CP also had longer reading times on ungrammatical adjectives in the VP and in the DP. Whether the differences for individual learners are significant is unknown, and given that individual means were calculated on a small number of items per condition—in most cases less than six with error data removed—these differences must be interpreted with caution. However, the differences are in the right direction (ungrammatical > grammatical) and suggest that some advanced learners may achieve nativelike sensitivity to gender errors on adjectives located outside the DP.

Limitations and Future Research

It is clear from the present study that distance from the controller noun moderates L2 learners' sensitivity to gender errors on postnominal adjectives. What

is not clear is whether the distance effect is a purely linear phenomenon (i.e., a function of the number of words or discourse referents that intervene between a noun and adjective) or a structural phenomenon (i.e., a function of the depth of embedding of the modifying adjective relative to its controller noun) (for relevant discussion, see O'Grady, Lee, & Choo, 2003). Linear and structural distances were confounded in the current study, in that increases in the structural distance between nouns and adjectives resulted in concomitant increases in linear distance. Future research with different sentence types is needed to isolate the source of the distance effect.

Regardless of whether the distance effect proves to be a linear or a structural issue, inherent to all distance-related proposals is the assumption that long-distance dependencies are a burden to the language processor because they tax a comprehender's working memory capacity. To the extent that the individual data presented in Table 6 indicate that some advanced learners are sensitive to gender errors in nonlocal domains, research is needed to investigate whether individual differences in working memory capacity affect learners' sensitivity to gender errors. Although current research on L2 sentence processing has not found significant effects for individual differences in working memory, the research thus far has been limited to *wh*-gap processing (Felser & Roberts, 2007; Juffs, 2004, 2005).

In light of these avenues for future research, data collection is already under way on a new eye-tracking study that examines the effects of linear distance and working memory on the processing of gender errors by native Spanish speakers and advanced learners of Spanish.

Conclusion

The results of the advanced learners indicate that English-speaking learners of Spanish can acquire abstract gender in adulthood, contra the failed functional features hypothesis. Furthermore, the results show that learners' sensitivity to gender errors on adjectives is a function of the distance that intervenes between the adjectives and their controller nouns. In sum, this study finds that nonnative sensitivity to gender errors in nonlocal domains is due to deficits in processing, not underlying competence.

Revised version accepted 3 July 2008

Notes

1 See Juffs and Harrington (1995) and Hawkins (2001, chap. 7) for a review of conflicting evidence regarding knowledge of the constraints on Subjacency in adult SLA.

2 The singular is not marked morphologically, but plural nouns ending in vowels typically end in *–s* (e.g., *libros* "books," *casas* "houses") and nouns ending in consonants take *–es* (e.g., *jardines* "gardens").
3 Accuracy was highest on d-type sentences, but these consisted of agreement with a pro-NP and thus were different from those in (1a)–(1c).
4 In addition to the comprehension data reported here, speech data from studies conducted in the Processability and UG paradigms show that accuracy on attributive adjectives is superior to that of predicative adjectives, even when the L1 has gender (Bartning, 2000; Bruhn de Garavito & White, 2002; Dewaele & Véronique, 2001; Di Biase & Kawaguchi, 2002; Glahn et al., 2001). Although distance was not the focus of her study, Franceschina (2001a) documented examples of gender-agreement errors between local and nonlocal constituents in the speech of advanced English-speaking learners of Spanish. Examples 16, 17, and 19 from her data (pp. 155–156) illustrate cases of gender errors in nonlocal domains.

16. *Todos$_{\text{N-Masc}}$ están ya pegadas$_{\text{Adj-Fem}}$ (talking about *etiquetas* "stickers")
"*They$_{\text{N-Masc}}$ are all stuck$_{\text{Adj-Fem}}$ already"
17. . . . sacar todas$_{\text{N-Fem}}$, de a *uno$_{\text{N-Fem}}$ (talking about *revistas* "magazines")
"to take them$_{\text{N-Fem}}$ all out, *one$_{\text{N-Fem}}$ by one"
19. Sí, tiene las patas$_{\text{N-Fem}}$ con una forma muy extraña [. . .] *los$_{\text{N-Masc}}$ tienen más largas$_{\text{Adj-Fem}}$ aun
"Yes, it has legs$_{\text{N-Fem}}$ of a strange shape [. . .] they have still longer *ones$_{\text{N-Masc}}$"

5 I also attempted to control for word frequency, a variable known to affect eye fixations (Rayner & Duffy, 1986), by consulting frequency dictionaries available at the time of the study (e.g., Alameda & Cuetos, 1995). Frequency dictionaries proved unsuitable for selecting words for a study that included L2 learners, particularly beginning and intermediate learners, because they do not account for a classroom learner's exposure to the language. Many words that are frequent for classroom learners (e.g., *mochila* "backpack") are not high-frequency words by dictionary counts, and the reverse is also true. The beginning and intermediate learners in this study had no immersion experience and did not use the L2 out of class. As such, they were not exposed to the quantity and variety of texts that serve as the basis for deriving the frequency counts reported in frequency dictionaries.

References

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Universidad de Oviedo.

Bartning, I. (2000). Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica, 54*, 225–237.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41–68). Cambridge: Cambridge University Press.

Braze, D., Shankweiler, D., Ni, W., & Conway Palumbo, L. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31*, 25–44.

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

Carreiras, M., & Clifton, C., Jr. (2004). On the on-line study of language comprehension. In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 1–14). New York: Psychology Press.

Carroll, S. E. (1996). Parameter-setting in second language acquisition: Explanans and explanandum. *Brain and Behavioral Sciences, 19*, 720–721.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: Benjamins.

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

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

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

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

Di Biase, B., & Kawaguchi, S. (2002). Exploring the typological plausibility of processability theory: Language development in Italian second language and Japanese second language. *Second Language Research, 18*, 274–302.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory and Cognition, 23*, 166–180.

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

Franceschina, F. (2001a). Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. In S. H. Foster-Cohen & A. Nizegorodcew (Eds.), *EUROSLA Yearbook* (Vol. 1, pp. 143–158). Amsterdam: Benjamins.

Franceschina, F. (2001b). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research, 17*, 213–247.

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

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Glahn, E., Håkansson, G., Hammarberg, B., Holmen, A., Hvenekilde, A., & Lund, K. (2001). Processability in Scandinavian second language acquisition. *Studies in Second Language Acquisition, 23*, 389–416.

Gregg, K. R. (1996). The logical and developmental problems of second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The handbook of second language acquisition* (pp. 49–81). San Diego, CA: Academic Press.

Gregg, K. R. (2001). Learnability and second language acquisition theory. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 152–180). Cambridge: Cambridge University Press.

Harrington, M. (2001). Sentence processing. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 91–124). Cambridge: Cambridge University Press.

Harris, J. W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry, 22*, 27–62.

Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.

Hawkins, R., & Chan, C. Y.-h. (1997). The partial availability of universal grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research, 13*, 187–226.

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

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 199–225.

Juffs, A. (2005). The influence of first language on the processing of *wh*-movement in English as a second language. *Second Language Research, 21*, 121–151.

Juffs, A., & Harrington, M. (1995). Parsing effects in L2 sentence processing: Subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition, 17*, 483–512.

Klein, E. C., & Martohardjono, G. (1999). Investigating second language grammars: Some conceptual and methodological issues in generative SLA research. In E. C. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 3–34). Amsterdam: Benjamins.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*, 53–78.

Myles, F. (1995). Interaction between linguistic theory and language processing in SLA. *Second Language Research, 11*, 235–266.

O'Grady, W., Lee, M., & Choo, M. (2003). A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition, 25*, 433–438.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 25*, 501–528.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.

Pienemann, M. (2005). *Cross-linguistic aspects of processability theory*. New York: Benjamins.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14*, 191–201.

Rayner, K., Juhasz, B. J., & Pollatsek, A. (2005). Eye movements during reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 79–97). Oxford: Blackwell.

Sabourin, L. (2003). *Grammatical gender and second language processing: An ERP study*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen, Groningen Dissertations in Linguistics 42.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*, 1–29.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*, 40–72.

Sorace, A. (2003). Near-nativeness. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 130–151). Oxford: Blackwell.

Sorace, A. (2006). Possible manifestations of shallow processing in advanced second language speakers. *Applied Psycholinguistics, 27*, 88–91.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*, 173–204.

Truscott, J., & Sharwood Smith, M. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition, 7*, 1–20.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5–31). Mahwah, NJ: Lawrence Erlbaum.

White, L., Valenzuela, E., Kozlowska-Macgregor, M., & Leung, Y.-k. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25*, 105–133.

Appendix

Stimuli

(underlining not present in the experiment)

Adjectives in the DP

Grammatical

1. Un sombrero nuevo cuesta mucho en tiendas como Bloomingdales.
2. Un libro aburrido es más difícil de leer que un libro interesante.
3. El pollo frío tiene tantas calorías y proteínas como el pollo caliente.
4. Una naranja negra no se debe comer porque puede tener bacterias.
5. Una película aburrida muchas veces trata temas poco originales.
6. Una camisa roja combina muy bien con pantalones negros.

Ungrammatical

7. Un libro larga generalmente no se puede leer en un par de horas.
8. Un trabajo aburrida es ideal para alguien que no tolera el estrés.
9. Un refresco fría es preferible al café cuando hace sol y calor afuera.
10. Una casa blanco no se vende rápido en esta zona de la ciudad.
11. Una fiesta pequeño es ideal para una persona tímida o introvertida.
12. Esa biblioteca alto ofrece una vista increíble de todo el campus.

Adjectives in the VP

Grammatical

13. El pollo es bastante bueno cuando se sirve con ensalada y vino.
14. Un libro es bastante pequeño cuando tiene sólo treinta páginas.
15. El queso es bastante malo cuando está verde o tiene un olor fuerte.
16. Una pregunta es bastante buena cuando hace pensar a la gente.
17. Una fiesta es bastante aburrida cuando no hay ni música ni comida.
18. Una casa es bastante alta cuando tiene más de tres o cuatro pisos.

Ungrammatical

19. Un vuelo es bastante larga cuando hay que hacer escala en varias ciudades.
20. Un vestido es bastante barata en las tiendas grandes como Old Navy.
21. Un trabajo es bastante mala cuando no ofrece vacaciones o días libres.
22. Una película es bastante largo cuando dura más de tres horas.
23. Una casa es bastante pequeño cuando tiene sólo una habitación.
24. Una biblioteca es bastante alto cuando tiene más de veinte pisos.

Adjectives in the Subordinate CP

Grammatical

25. Un libro recibe bastante atención cuando es nuevo y popular.
26. Un sombrero no llama la atención cuando es pequeño y poco original.
27. Un libro no se debe leer cuando es malo y poco interesante.
28. Una pregunta causa bastante confusión cuando es larga y difícil.
29. Una naranja no se debe comer cuando es pequeña y está verde.
30. Una casa es muy difícil de vender cuando es roja y verde.

Ungrammatical

31. Un trabajo no le gusta a alguien cuando es mucha y poco interesante.
32. Un libro no se lee rápidamente cuando es aburrida y difícil.
33. Un refresco tiene muy buen sabor cuando está fría y no caliente.
34. Una casa se vende bastante bien cuando es nuevo y muy grande.
35. Una biblioteca no tiene computadoras cuando es pequeño y falta dinero.
36. Una película no genera dinero cuando es malo y poco popular.