EARLY AND LATE BILINGUAL PROCESSING OF SPANISH GENDER, MORPHOLOGY, AND GENDER CONGRUENCY[*]

Irma Alarcón
Wake Forest University

ABSTRACT. The present study investigates whether advanced proficiency-matched early and late bilinguals display gender agreement processing quantitatively and qualitatively similar to that of native speakers of Spanish. To address this issue, a timed grammaticality judgment task was used to analyze the effects on accuracy and reaction times of grammatical gender, morphology, and gender congruency of the article and adjective within a noun phrase. Overall results indicated no significant statistical differences between the native speakers and the two bilingual groups. Both early and late bilinguals displayed similar grammatical gender knowledge in their underlying grammars. A detailed examination of the congruency effect, however, revealed that the native speakers, not the bilinguals, displayed sensitivity to gender agreement violations. Moreover, the native speakers and early bilinguals pattern together in accuracy and directionality of gender agreement processing: both were less accurate with incongruent articles than with incongruent adjectives, while the second language learners were equally accurate in both agreement domains. Despite having internalized gender in their implicit grammars, the late bilinguals did not show native-like patterns in real time processing. The present findings suggest that, for high proficiency speakers, there is a distinct advantage for early over late bilinguals in achieving native-like gender lexical access and retrieval. Therefore, age of acquisition, in conjunction with learning context, might be the best predictor of native-like gender agreement processing at advanced and near-native proficiency levels.

Keywords. language acquisition; second language learners; Spanish heritage speakers; gender agreement; morphology; gender congruency; language processing

RESUMEN. El presente estudio investiga si los hablantes bilingües tardíos y tempranos, que comparten un nivel avanzado de competencia lingüística, muestran en su español concordancia de género cuantitativa y cualitativamente similar a la de los hablantes nativos de español. Con este objetivo, se utilizó una prueba cronometrada de juicio de gramaticalidad para analizar los efectos en la exactitud y los tiempos de reacción del género gramatical, la morfología y la congruencia de género entre el artículo y el adjetivo dentro de un sintagma nominal. En general, los resultados no revelaron diferencias estadísticas significativas entre los hablantes nativos y los dos grupos de bilingües. Tanto los bilingües tempranos como tardíos mostraron conocimiento del género gramatical en sus gramáticas subyacentes. Sin embargo, un análisis detallado del efecto de la congruencia reveló sensibilidad a las violaciones en la concordancia de género solamente entre los hablantes nativos. En lo concerniente a la exactitud y direccionalidad del procesamiento de género, los hablantes nativos y los bilingües tempranos mostraron menor precisión con artículos incongruentes que con adjetivos incongruentes, mientras que los aprendices de segunda lengua mostraron igual precisión en ambas áreas de concordancia. A pesar de haber internalizado el género gramatical en sus gramáticas implícitas, los bilingües tardíos no

procesaron la concordancia de género en tiempo real del mismo modo que los hablantes nativos. Los resultados sugieren que, para hablantes con alto nivel de competencia, los bilingües tempranos gozan de una ventaja sobre los tardíos en el acceso léxico al género a nivel nativo.

Palabras clave. adquisición de lengua; aprendices de segunda lengua; hablantes de español como lengua heredada; concordancia de género; morfología; congruencia de género; procesamiento del lenguaje.

1. Introduction

A major issue in second language (L2) acquisition research is the extent to which adult L2 learners (late bilinguals) are able to fully develop both the implicit knowledge (representation) and the processing routines (computation) associated with grammatical features that are absent in their native language (L1). Since languages differ so greatly in their gender systems, the study of L2 acquisition of grammatical gender offers insights into this issue.

The L2 acquisition of Spanish has been a frequent topic in this literature. Spanish is a binary gendered language, in which every noun must be either feminine or masculine, and in which determiners and adjectives must agree with the gender of the noun they modify. Gender agreement (as in the feminine *la sociedad justa* 'the-FEM just-FEM society-FEM and the masculine *el hotel barato* 'the-MASC cheap-MASC hotel-MASC') is both a lexical and a morphosyntactic phenomenon, so investigating it enhances our understanding of the mechanisms underlying both lexical access and syntactic processing (Carroll 1989). For native speakers of Spanish, gender is acquired early in life (e.g., López-Ornat 1997; Mariscal 2009), but for adult English-speaking learners of Spanish, whose L1 system lacks grammatical gender, mastery of gender is one of the most challenging aspects of acquiring the language, even for high proficiency learners (e.g., Franceschina 2005; McCarthy 2008). Moreover, recent research with heritage language (HL) speakers (early bilinguals) of Spanish, those who grow up in a home where Spanish as a minority language is spoken, suggests that gender is a difficult grammatical feature for them as well. (Montrul 2016 presents a detailed account of this population.) Understanding why gender is problematic for both L2 and HL learners, and what this implies for L2 acquisition theories, has become a prominent theme in current L2 and HL research agendas (Foote 2011; Grüter et al. 2012; Montrul et al. 2014).

Various theoretical models have been proposed to account for both L2 grammatical knowledge and performance in real-time processing. These models, which will be discussed in the next section, have been tested in L2 studies of Spanish gender acquisition (e.g., Gillon Dowens et al. 2010; Keating 2009, 2010; Sagarra & Herschenson 2010, 2011), but their implications for HL processing have not yet been explored in sufficient detail. Suppose HL and L2 learners, who possess the same underlying grammars, were found to display no differences in their processing patterns. This would support full accessibility accounts of L2 acquisition, as opposed to deficit models, and the theoretical insights into L2 learning could largely apply to HL learning as well. Alternatively, if L2 and HL learners were found to exhibit different processing patterns, this could be attributed to the age of language onset (the key definitional distinction between the two groups) and/or the effects of learning context, naturalistic for HL learners and formal classroom instruction for L2 learners (Montrul 2010). Would computational deficit models, for example, apply equally to HL and L2 learners? Do early and late bilinguals use different types of knowledge

(implicit vs. explicit) during real time processing? Do they use the same mechanisms to access and store gender knowledge?

The present study tries to explore these questions by comparing the gender agreement processing of HL and L2 learners, and their ultimate attainment, analyzing the influences of noun gender, noun morphology, and gender congruency. Although research on L2 processing has been steadily increasing in the last decade, we still know very little about how early and late bilinguals differ during real-time comprehension and production tasks (Montrul 2009; Clahsen & Felser 2006a). To date, there have been few empirical studies of Spanish gender acquisition comparing these two bilingual populations (Alarcón 2011; Montrul et al. 2008; Foote 2011; Montrul et al. 2013; Montrul et al. 2014), and only two of them (Foote 2011; Montrul et al. 2014) used online tasks to examine gender agreement processing.

Consequently, the design of the present study was inspired by Montrul's calls for processing-oriented studies (2009, 2010) of high proficiency HL speakers (2013), as well as by Foote's observation that data derived from various tasks could provide "converging evidence that this type of participant does possess implicit knowledge of L2 inflectional morphology" (2011: 217). Given the intermediate status of HL learners between L2 learners and native speakers (see Montrul 2016), there is a need for online processing studies of the effects of age and context of acquisition on the ultimate attainment of both groups of language learners. To investigate these issues, the current study focuses on the processing of Spanish grammatical gender by comparing the linguistic behavior of proficiency-matched advanced L2 and HL learners, using native speakers as a baseline. A timed grammaticality judgment task is used to tap into implicit knowledge (Ellis 2005; Hopp 2006) during real-time processing of gender agreement in a noun phrase. In this type of psycholinguistic task, gender agreement processing is assessed as displaying sensitivity to gender errors during online comprehension or production (cf. Bates et al. 1996; Montrul et al. 2014)**.** For example, when presented with *La dieta ligero (incongruent), native speakers show longer response times than *La dieta ligera* (congruent). This sensitivity to gender errors as evinced in a slower response is referred to as the gender congruency or grammaticality effect. (It is also known as the gender interference effect, as in Schriefers 1993; Schriefers & Teruel 2000.) This effect has been reported in processing studies in both L1 (Barber & Carreiras 2005) and L2 (Sagarra & Herschensohn 2011), as well as in the few existing studies with both early and late bilinguals (Foote 2011).

2. Theoretical perspectives on morphosyntactic attainment in the L2 system: representation and computation

Representational (also known as UG or generativist) approaches to SLA focus on the L2 learner's interlanguage grammar, and how it compares with a native speaker's grammar. Computational models focus on how this implicit grammatical knowledge is processed in real-time production and comprehension. Both perspectives encompass both accessibility and deficit accounts, under which L2 learners can, or cannot, achieve native-like proficiency regarding a grammatical feature. This yields four distinct accounts of L2 morphosyntactic acquisition, and, consequently, four predictions for HL acquisition.

Proponents of representational deficit accounts, such as the Fundamental Difference Hypothesis (FDH) (Bley-Vroman 1989) and the Failed Functional Features Hypothesis

(FFFH) (Hawkins & Chan 1997), posit that only native speakers have access to Universal Grammar (UG), and thus only they are capable of fully acquiring grammatical features, such as gender. Adult L2 learners, who are past the critical period and therefore no longer have access to UG, cannot acquire grammatical features that are not present in their L1. Failure to achieve native-like proficiency is explained in terms of a deficit in implicit grammatical knowledge (representation). Under representational deficit frameworks, one could predict that HL speakers, who have acquired Spanish gender representation in early childhood, can attain native-like performance with grammatical gender.

Representational accessibility accounts, which include the Full Transfer/Full Access Hypothesis (FTFA) (Schwartz & Sprouse 1996) and the Missing Surface Inflection Hypothesis (MSIH) (Prévost & White 2000), suggest that L1 and L2 learners have similar access to UG. In this view, adult L2 learners are capable of fully acquiring grammatical features absent in their L1, including gender. The morphological errors L2 learners continue to make in their production, which might extend to the advanced level, are attributed to performance or mapping issues, rather than impaired underlying competence (Lardiere 2007). Applying full access models to HL learners might imply no differences between early and late bilinguals with respect to their implicit knowledge of gender.

Computational deficit models contend that native and non-native grammatical processing are fundamentally different. One example is Clahsen and Felser's (2006a, 2006b) Shallow Structure Hypothesis (SSH), which holds that L2 learners, compared with native speakers, rely more on lexical-semantic and pragmatic cues, and less on syntactic information, when parsing sentences. Consequently, the syntactic representations learners compute during L2 sentence comprehension are more "shallow" and less complex than those of native speakers. Applying Clahsen and Felser's (2006b) Continuity Parsing Hypothesis, that child and adult L1 parsing are essentially the same, could imply that HL learners of Spanish, who by definition acquired Spanish syntax in early childhood, would be able to use syntax-based parsing when processing gender agreement in Spanish.

Hopp (2006, 2010, 2013) presents the case for computational accessibility models, which argue that native and non-native processing systems are essentially the same. This approach highlights the role of proficiency in attaining native-like parsing patterns. At near-native proficiency levels, L2 learners converge on processing that is qualitatively identical to that of native speakers. In terms of L2 inflection, native-like attainment is possible in both underlying grammatical knowledge and processing, so differences between native and non-native performance are essentially quantitative. Although Hopp (2010) claims that the precise extent to which efficiency limitations can account for non-native ultimate attainment in L2 acquisition is still an open question, his computational accessibility perspective could be used to predict that high proficiency HL learners, just like their L2 learner counterparts, are capable of qualitatively native-like processing.

The present study will consider these accounts in examining whether advanced L2 and HL learners, both of whom possess similar Spanish gender representation, display gender agreement processing qualitatively and quantitatively similar to that of native speakers.

3. Background

3.1. The L2 acquisition of Spanish gender

Online psycholinguistic research is critically important because it can illuminate both implicit linguistic knowledge and automatic linguistic computation. Online experiments can precisely measure such phenomena as reaction times (RTs), eye movements, and event related potentials (ERPs), by tapping into mental processes as they are happening. Most of the existing online research on Spanish gender investigates adult L1 performance (e.g., Barber & Carreiras 2005; Domínguez et al. 1999; Wicha et al. 2004), but the online study of L2 processing of Spanish grammatical gender has increased in recent years. This research has centered on comprehension data, often mixing grammatical and ungrammatical phrases or sentences in order to assess L2 sensitivity to gender agreement violations. This sensitivity (the gender congruency or grammaticality effect) is manifested either in longer RTs or by a P600 effect, an ERP elicited when the brain detects a grammatical error during syntactic processing (e.g., Barber & Carreiras 2005). These findings tend to converge with representational access approaches by showing that L2 learners, even in the early stages of language learning, display sensitivity to a feature absent in their L1.

In a pioneering study on the L2 processing of gender, Tokowicz and MacWhinney (2005) used ERPs to examine the effect of different syntactic structures on grammaticality judgments of noun-determiner combinations. Testing English L1 beginning learners of Spanish, they found a strong grammaticality effect, suggesting that even low proficiency learners whose L1s do not include gender are implicitly sensitive to gender violations in comprehension. This sensitivity to ungrammaticality at the very early stages of language learning, however, has not been replicated in some other studies. Using a self-paced reading moving-window, Sagarra and Herschensohn (2010, 2011) found that intermediate learners, but not beginners, displayed sensitivity to noun-adjective gender agreement violations by exhibiting longer RTs for ungrammatical items. Nonetheless, these results all suggest that intermediate proficiency learners display grammatical gender knowledge and implementation that are qualitatively similar to that of native speakers, and are fully consistent with Hopp's (2006) computational access hypothesis that native-like attainment of L2 inflection is possible.

Using eye movement tracking methodology, Keating (2009, 2010) found a role for syntactic distance in L2 learners' sensitivity to gender agreement violations. Only advanced learners, not beginners or intermediates, and only in the noun phrase (NP), not in verb phases or subordinate clauses, displayed native-like performance. This is consistent with computational deficit accounts, since even the high proficiency L2 learners were not native-like in contexts outside the NP, suggesting that the noun-adjective agreement errors were attributable to deficits in processing, rather than to a lack of underlying grammatical knowledge. This conclusion is supported by a study of advanced L2 learners (Gillon Dowens et al. 2010), who exhibited native-like ERPs in the context of local domain noun-determiner gender agreement. But another investigation of advanced L2 Spanish learners with a non-gendered L1 (Alemán Bañón et al. 2014) found ERP evidence of native-like gender agreement across syntactic domains, which supports both representational and computational accessibility accounts.

Regarding RTs, only a few studies have explored the facilitative effects of gender cues on lexical access and retrieval during online grammatical sentence comprehension or production tasks. Existing findings suggest that L1 and L2 gender processing are fundamentally different, at least with respect to the predictive and effective use of gender related cues in real time processing. Such usage differences could be attributed to either representational (Bley-Vroman 1989) or processing (Clahsen & Felser 2006a) deficits. In a study of L2 gender agreement processing between a complex sentential subject and a predicate adjective during an online sentence completion task, Alarcón (2009) examined the effects on RTs of several linguistic variables, including noun morphology. Results showed that only the native speakers, not the L2 learners, were significantly faster when the head noun was overtly rather than non-overtly marked for gender, suggesting that morphology was a more salient cue for native than for non-native speakers of Spanish. These RT findings imply that L2 learners' automatized inflectional computations are slower, and their use of gender cues weaker, than those of Spanish monolinguals.

Using an eye-tracking procedure to investigate the use of morphosyntactic cues in spoken language processing by intermediate proficiency L2 learners, Lew-Williams and Fernald (2010) found that only their L1 controls took advantage of gender-marked articles to identify target referents. The native speakers were significantly faster than the learners in gender discord contexts. This lack of a gender facilitation effect was replicated in Grüter et al. (2012), which expanded the earlier study by including high proficiency learners and both comprehension and production exercises. Using the visual-world eye-tracking paradigm, they found that advanced L2 learners performed at ceiling in offline comprehension, but made errors in elicited production, and exhibited weaker than native-like use of gender-marked articles as a facilitative cue in online processing of familiar nouns. These findings suggest a difference between L1 and L2 gender retrieval and processing: native speakers use the gender information encoded by the determiner to facilitate processing of the upcoming noun, but L2 learners do not.

This claim has been challenged by other studies. Using the same experimental paradigm, but with a more difficult task because the relevant noun phrase was embedded in a complex sentence, Dussias et al. (2013) replicated the earlier finding that native speakers make use of gender marked articles to anticipate upcoming nouns during language processing. But they also found that highly proficient L2 learners were capable of this as well, using gender marked articles to more quickly recognize nouns in a comprehension task. These results, which suggest a role for L2 proficiency, were replicated by Halberstadt et al. (2018). Also using the visual world paradigm, but manipulating both cognate status and noun morphology, they found that both native speakers and L2 advanced learners were able to use gender information on articles to facilitate processing. The L2 learners, however, showed predictive gender processing only with overtly marked nouns.

Therefore, despite some discrepancies, online research generally supports the claim that L2 learners, at all proficiency levels, have implicit knowledge of Spanish gender in their developing systems, since they are capable of detecting gender agreement violations.

3.2. Age and context of acquisition of Spanish gender

In one of the first experimental studies comparing the Spanish gender agreement knowledge of HL and L2 learners, Montrul et al. (2008) examined oral production and written comprehension. They found that both groups of learners made gender agreement

errors (between 10% and 25%), while their native speaker controls performed at ceiling. Both experimental groups were more accurate with articles than adjectives, with overtly than non-overtly marked nouns, with masculine rather than feminine nouns, and exhibited a masculine default by using masculine forms in feminine contexts. The authors also found significant differences between the two groups: L2 learners were more accurate than HL learners in written comprehension, but less accurate in oral production. They concluded that the HL learners, despite their early exposure to Spanish and their more native-like oral production, had not fully acquired the Spanish gender system.

Alarcón (2011) replicated two of Montrul et al.'s experiments to examine if their asymmetrical findings also held for proficiency-matched advanced HL and L2 learners. Consistent with Montrul et al.'s findings, in both in written comprehension and oral production, the two groups were more accurate with overtly marked nouns and with masculine nouns. Contrary to Montrul et al., though, Alarcón found that the HL learners were equally accurate in gender comprehension and production, but that the L2 learners were significantly less accurate in production than in gender comprehension. In oral production, some of the HL learners, but none of the L2's, scored within the native speakers' range. Overall, these findings imply processing difficulties for advanced L2 learners. Their high comprehension scores indicate abstract gender representation, but their lower scores in production suggest struggles with the surface manifestations of that underlying knowledge. Alarcón concluded that age of acquisition is fundamental in predicting ultimate achievement in gender agreement performance.

To investigate both implicit and explicit knowledge, as well as the effect of noun morphology on gender agreement processing, and to control for modality, Montrul et al. (2014) designed three online word recognition tasks. Each task used aural stimuli consisting of the same Det Adj N phrases. The participants were intermediate and advanced HL and L2 learners, with native controls. Overall results showed significant grammaticality effects for all groups in both accuracy and RT's, but the word repetition task revealed this effect only for the HL and native speakers, not for the L2 group. The authors concluded that HL speakers show more native-like patterns than L2 learners in implicit tasks, which require aural comprehension and oral production. Findings from an elicited production task with pictures and aural stimuli administered to the same participants supported this conclusion (see Montrul et al. 2013).

Using a moving window word-by-word reading task, Foote (2011) investigated whether HL and L2 learners display integrated gender knowledge in Spanish, and whether age of acquisition plays a role, by examining sensitivity to gender agreement violations while reading for comprehension. Results indicated that age of acquisition did not affect the participants' automatic gender competence: both experimental groups were sensitive to noun-adjective agreement errors, and all participants, including native speaking controls, displayed significantly longer reading times in ungrammatical rather than grammatical contexts. Consistent with the above studies, this finding reveals that, despite occasional performance errors, L2 learners possess integrated knowledge (see Jiang 2007) of gender in their underlying grammars.

Findings from these studies support the following three claims. First, HL learners have an advantage over L2 learners in the oral production of gender agreement. Second, consistent with previous L1 and L2 findings (e.g., Franceschina 2001; Pérez-Pereira 1991),

both HL and L2 learners are more accurate with overtly marked nouns and masculine nouns, and display a masculine default strategy, which is more evident in oral than in written production. Third, HL speakers perform closer to the native-speaker range than L2 learners at the same level. These findings, taken together, can be explained by representational access accounts of morphosyntactic variability in developing grammars, such as the MSIH (Prévost & White 2000), implying that HL and L2 learners have acquired grammatical gender in terms of mental representation, despite occasional errors in the surface realization of its abstract features.

Two studies directly relevant to the present study have found a significant grammaticality effect during online comprehension. As noted above, Foote (2011) found that early and late bilinguals and native speakers were sensitive to gender agreement errors within the noun phrase. Her task, though, manipulated only adjectives, not articles. Similarly, Montrul et al. (2014) used, among others, an auditory grammaticality judgment task to reveal the same congruency effect, but the study included only noun phrases in which either both the article and the adjective were congruent, or both were incongruent, so there was no differentiation between incongruent articles and incongruent adjectives. The current study complements their work by using written stimuli in its timed grammaticality judgment task, and by manipulating articles and adjectives independently. Consequently, it adds another layer to the investigation of the gender interference effect in different agreement domains.

4. The current study

4.1. Research questions and predictions

Three research questions (RQ) guide the present study:

a) RQ1 - Are there any overall differences in NP gender agreement processing between native speakers and high proficiency L2 and HL learners?

An affirmative answer would highlight the issue of whether HL learners display more native-like gender processing behavior than L2 learners. Both the FDH (Bley-Vroman, 1989) and the SSH (Clahsen & Felser 2006a, 2006b) predict that this would be the case, the former due to L2 learners' lack of access to UG, and the latter due to shallow L2 parsing. Alternatively, if the answer were no, if L2 and HL learners were as rapid and as accurate as native speakers, and they patterned with native speakers, then we would have evidence that gender knowledge has been integrated (see Jiang 2007), and that L1 and L2 processing are identical (Hopp 2006, 2010, 2013; Sagarra & Herschensohn 2010, 2011). Based on the few online studies with HL learners (particularly, Montrul et al. 2014), one could expect early bilinguals to display an advantage over late bilinguals on the timed grammaticality judgment task due to their childhood exposure to the language.

b) RQ2 - Do noun gender (masculine or feminine) and noun morphology (overt or non-overt) affect processing?

L1 research suggests that children rely more on morphophonological and syntactic cues (i.e., noun endings) than on semantic information (i.e., sex of the referent) for recognizing

gender and establishing gender agreement (Mariscal 2009; Pérez-Pereira 1991). Moreover, both child and adult L1 Spanish speakers make use of gender cues to facilitate online processing (Lew-Williams & Fernald 2007), and are affected by gender agreement violations (Wicha et al. 2004). Consequently, morphosyntactic information (intralinguistic cues) can facilitate, or inhibit, gender processing. L2 findings on the effects of gender value (masculine or feminine) and morphology (overt or non-overt) on gender assignment and agreement predict that both early and late bilinguals would be more accurate and faster on gender agreement processing with masculine than with feminine nouns, and with overt rather than non-overt nouns (Alarcón 2011; Franceschina 2001; McCarthy 2008; Montrul et al. 2008; White et al. 2004). Interestingly, however, some research findings show that even highly proficient L2 learners, who perform at ceiling in offline comprehension tasks, are less efficient than native speakers in their use of morphosyntactic cues in lexical access (Lew-Williams & Fernald 2010; Grüter et al. 2012). Considering the intermediate status of HL learners between L2 learners and native speakers (Montrul 2016), it can be hypothesized that they would be more native-like than L2 learners in their access to and use of gender and morphological cues in gender agreement processing.

c) RQ3 - Does grammaticality affect NP gender processing? Are there processing differences between Art N *Adj and *Art N Adj sequences?

The effect of grammaticality (the congruency effect) on gender processing has been studied in both language comprehension (Wicha et al. 2004) and production (Schriefers & Teruel 2000). Previous L1 and L2 findings on sensitivity to Spanish gender agreement violations suggest that all three groups would be more accurate and faster when processing gender in grammatical rather than ungrammatical NPs (Alemán Bañón et al. 2014; Barber & Carreiras 2005; Caffarra et al. 2014; Foote 2011; Lew-Williams & Fernald 2007; Montrul et al. 2014; Sagarra & Hershensohn 2011). Under time pressure, ungrammaticality has been shown to lower accuracy scores and raise reaction times (Hopp 2006).

In terms of agreement domain, previous L1 and L2 research indicates that agreement is acquired earlier, and is more accurate, with articles than with adjectives (López-Ornat 1997; Bruhn de Garavito & White 2002). This suggests that both early and late bilinguals will process Art N *Adj and *Art N Adj differently. Although there has been little study of the grammaticality effect that differentiates between articles and adjectives within the same NP (cf. Gillon Dowens et al. 2011; Hopp & Lemmerth 2018), the existing findings suggest that all participants would be less accurate and slower when processing incongruent adjectives than incongruent articles.

4.2. Participants

Fifty-three participants from a small, private U.S. university took part in the study: 18 English-speaking learners of Spanish (L2), 18 heritage language Spanish speakers (HL), and 17 native Spanish speakers (NS), who served as a baseline for comparisons.

All of the L2 learners (7 males, 11 females; mean age: 21.2; range: 20-24 years) were raised as English monolinguals in English-speaking homes. Most of them (78%) started to learn Spanish around puberty (middle/high school), and had been studying Spanish for over seven years. All were undergraduate Spanish majors or minors, and all but one had studied

Spanish abroad. At the time of the data collection, all but one were enrolled in advanced Spanish courses at their university. Self-ratings showed that most of them considered their language abilities to be advanced: 89% for reading, 83% for listening, and 78% for both speaking and writing skills.

The HL learners (8 males, 10 females; mean age: 20.3; range: 18-27 years) were all students at the same university the L2 learners attended. Only 44% of them were Spanish majors or minors, and of the rest, only half had taken at least one undergraduate Spanish course. For most of them (78%), both parents were Latin American native Spanish speakers from various countries, including Cuba, Ecuador, Guatemala, Mexico, and Peru. 61% of the HL group were born in the U.S.: the rest were pre-pubescent when they first came to live in the U.S. (age-of-arrival range: 3-10). All were exposed exclusively to Spanish until at least age 5. They started learning English after their pre-school years, and their formal schooling was in English. According to self-ratings, 100% of them considered their own listening and reading skills as either native or advanced. The analogous percentage for speaking was 94%, and for writing 83%.

The NS of Spanish (5 males, 12 females; mean age: 42.5; range: 18-63 years) came from a variety of national backgrounds (Argentina, Chile, Colombia, Spain, Honduras, and Mexico), and were recruited from the same university community as the other participants. Eight were language instructors, and the rest were drawn from among their friends and relatives. All of them had received their K-12 schooling entirely in Spanish, and had arrived in the U.S. as adults. At the time of the study, these speakers had been living in the U.S. for an average of 13.1 years.

4.3. Procedure

All participants attended three separate sessions. At the first, each participant completed an extensive language background questionnaire, and took a vocabulary test. At the second meeting, a grammar test was administered. Finally, the experimental task, a timed grammaticality judgment task, was run during the third session.

To control for the influence of grammatical proficiency on the results of the experimental task, and to ensure that the L2 and HL learners were matched at an advanced level compared to the baseline, all participants, including the NS group, took the grammar test. This test was used in an earlier study (Alarcón 2011) and consisted of 60 multiple-choice questions presented in brief and familiar contexts, and covered structures including the copula *ser/estar* 'to be,' demonstratives, object pronouns, preterite vs. imperfect, and if-clauses. To verify knowledge of the meanings of the 60 target nouns included in the experimental task, a vocabulary test was administered. Participants were asked to provide the English equivalent or an English explanation of the meaning of each Spanish word, such as *talento* 'talent,' *país* 'country,' *lluvia* 'rain,' and *salud* 'health.' The results of these tests are presented in Table 1, with box plots in Figure 1. (Although the native controls also took the vocabulary test, their knowledge of English is irrelevant for the current study, as is the fact that three of their vocabulary test scores were outliers, as suggested in Table 1, and seen directly of Figure 1.)

Table 1. *Grammatical proficiency and vocabulary test results (percentage) by group*

	Grammatical Proficiency Test			Vocabulary Test		
	Mean	SD	Range	Mean	SD	Range
L2	86.4	4.49	75-92	97.1	2.02	93-100
HL	89.9	6.18	80-97	99.0	1.26	95-100
NS	97.7	2.02	93-100	91.4	20.02	24-100

All scores greater than the 75th percentile + 1.5 times the Interquartile Range (IQR), or less than the 25th percentile - 1.5 times the IQR were removed from the data. Since ANOVA assumptions were not fully satisfied, partly because the data were slightly skewed, and, even more important, because of the ceiling effects, particularly on the vocabulary test, the results were analyzed with non-parametric Kruskal-Wallis Rank Sum tests, and then with Dunn multiple comparison tests, with *p*-values adjusted by the Benjamini-Hochberg method.

For the grammar test, the Kruskal-Wallis indicated a significant difference between the three groups: chi-squared $= 27.407$, $df = 2$, $p = .000$. On the multiple comparisons, however, the NS-HL ($p = .000$) and NS-L2 ($p = .000$) comparisons were significant, but the HL-L2 comparison was not ($p = .121$). Kruskal-Wallis analysis of the vocabulary test results also showed a significant group difference: chi-squared $= 23.518$, $df = 2$, $p = .000$. All three group comparisons were significant: NS-HL ($p = .007$), NS-L2 ($p = .000$), and HL-L2 ($p = .030$).

There are two interesting points here. First, the raw scores on the vocabulary test were so close to ceiling that statistically significant group differences did not necessarily reveal practical differences. Second, the HL and L2 groups were not significantly different on the grammar test. Consequently, these two groups can be regarded as proficiency-matched.
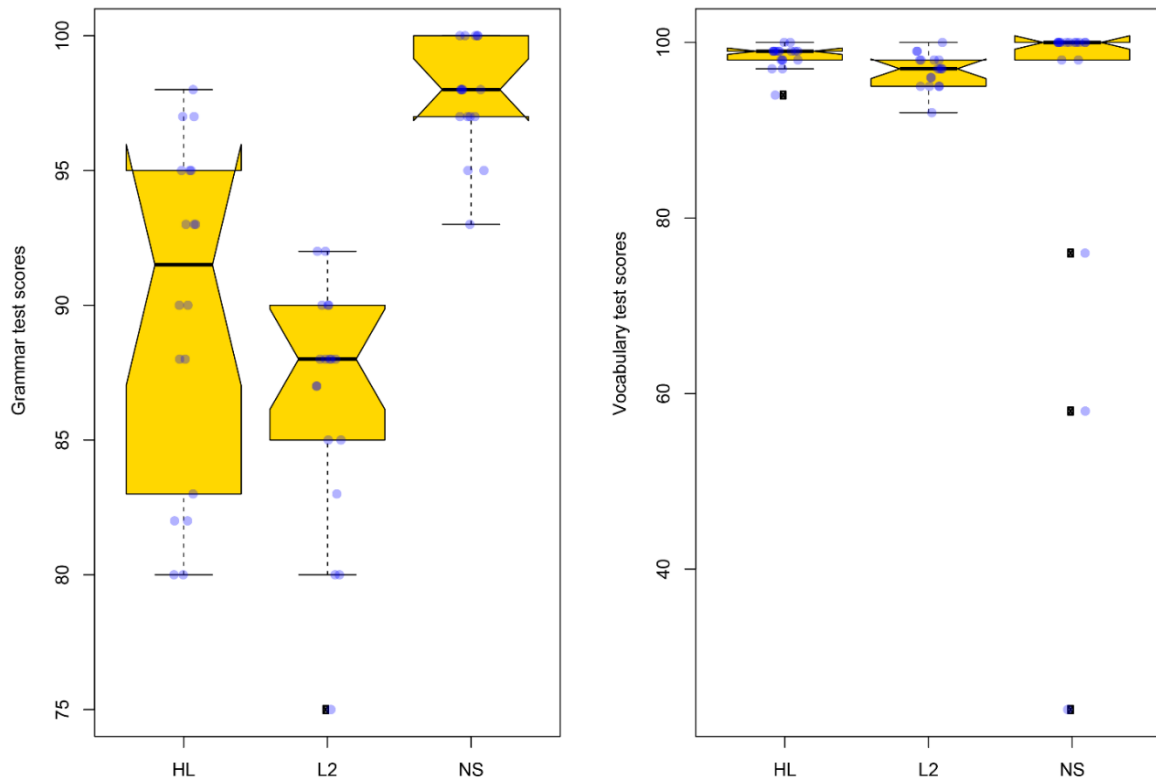
Fig. 1. *Individual results for the grammar and vocabulary tests by group*

4.4. Experimental timed grammaticality judgment task

4.4.1. Materials and design

The experimental task was a timed grammaticality judgment activity (GJT) consisting of a series of congruent and incongruent noun phrases displayed one at a time on a computer screen. All the words (except the articles) were drawn from the 1,000 - 3,000 frequency range in a frequency dictionary of core vocabulary for learners (Davies 2006), and were balanced by length: only two and three syllable words were used. To control for the influence of animacy (e.g., Sagarra & Herschensohn 2011) only inanimate nouns were included. The NPs were all of the form Def Art N Adj. There were 60 target nouns: 30 masculine and 30 feminine. Within each gender, half the nouns were overtly marked for gender (masculine nouns ending in –o or feminine nouns ending in –a) and the other half were not. Each noun was presented in three experimental conditions: congruent with both article and adjective, with a congruent article and an incongruent adjective, and with an incongruent article and a congruent adjective. All the adjectives were overtly marked for gender. Thus, there were 60 experimental nouns, each presented three times, for a total of 180 experimental items. The independent variables were noun gender, noun morphology, and condition. The dependent variables were accuracy and reaction times. See Table 2 for examples of the experimental conditions. (See the Appendix for the 60 grammatical noun phrases used in the task.)

Table 2. *Congruent and incongruent conditions in the grammaticality judgment task*

Condition	Examples
1. Congruent	Article + Noun + Adjective:
	La crisis violenta 'The-FEM violent-FEM crisis-FEM'
	El sueño extraño 'The-MASC odd-MASC dream-MASC'
2. Incongruent adjective	Article + Noun + *Adjective:
	*La crisis *violento* 'The-FEM violent-MASC crisis-FEM'
	*El sueño *extraña* 'The-MASC odd-MASC dream-FEM'
3. Incongruent article	*Article + Noun + Adjective:
	El crisis violenta 'The-MASC violent-FEM crisis-FEM'
	La sueño extraño 'The-FEM odd-MASC dream-MASC'

4.4.2. GJT procedure

Participants were tested individually in the Cognitive Laboratory of the Psychology Department at their university. The stimuli were a series of Spanish noun phrases presented in written form, one at a time, in the center of a computer screen using E-Prime software (cf. Bowles 2011). For each participant, these phrases were presented in random order in black text on a white background. Participants were asked to indicate, as quickly and accurately as possible, whether they believed the NP stimulus was grammatical or ungrammatical by pressing the "V" key for grammatical and the "N" key for ungrammatical NPs. The stimulus remained on the screen until the participant pressed one of the two keys, which immediately brought the next item to the screen. This process continued until the entire task was completed. To familiarize the participants with the mechanics of the activity, a practice section with 18 items preceded the actual experiment. The software recorded both the accuracy (correct or incorrect) and the reaction time (in milliseconds) of each response.

4.5. Results

Descriptive results of the experimental GJT task are given by Group in Tables 3 (accuracy) and 4 (RT), by Condition in Tables 5 (accuracy) and 6 (RT), and are displayed visually in Figure 2. The latter two Tables also include C2 and C3, the two incongruent conditions, pooled together, to address part one of RQ3.

The most salient point is the high overall accuracy rates of all three groups. This was by design. Since the primary goal of the study was to explore differences in the levels of native-like automaticity in the gender processing of L2 and HL speakers, the experimental items had to be very simple, requiring as little conscious processing as possible. Consequently, any differences between the experimental groups would likely be exceedingly small and nuanced. Such limited variation renders statistical analysis more

perilous, by reducing the power of statistical tests and greatly attenuating effect sizes. Nonetheless, despite the intentionally conservative emphasis on avoiding Type II errors (false negatives), several statistically significant patterns were revealed.

Table 3. *Means and standard deviations for the grammaticality judgment task: Accuracy*

	L2 Learners		HL Learners		Native Speakers	
	M	SD	M	SD	M	SD
Overall Accuracy	97.50	4.25	96.93	4.93	98.15	3.80
Gender Accuracy:						
Masculine	97.84	3.77	97.68	4.30	97.78	4.05
Feminine	97.17	4.68	96.18	5.41	98.52	3.52
Morphology Accuracy:						
Overt	97.82	3.81	97.42	4.36	98.28	3.03
N-Overt	97.19	4.65	96.44	5.42	98.02	4.45
Condition Accuracy:						
Congruent	95.86	5.06	97.25	4.74	98.99	2.71
Incongruent w/adj	98.65	2.97	98.12	4.22	98.99	2.71
Incongruent w/art	98.00	4.00	95.40	5.43	96.49	4.99
Incongruent	98.83	3.53	96.76	5.04	97.74	4.19

Table 4. *Means and standard deviations for the grammaticality judgment task: Reaction times*

	L2 Learners		HL Learners		Native Speakers	
	M	SD	M	SD	M	SD
Grammaticality judgments (in milliseconds)						
Overall RTs	1625	629	1733	455	1550	558
Gender RTs:						
Masculine	1616	604	1719	442	1563	582
Feminine	1634	655	1746	470	1537	537
Morphology RTs:						
Overt	1496	484	1684	405	1588	604
Non-Overt	1754	726	1782	498	1512	509
Condition RTs:						
Congruent	1721	776	1715	361	1542	536
Incongruent w/adj	1566	525	1802	462	1634	622
Incongruent w/art	1588	553	1681	526	1474	508
Incongruent	1577	537	1742	497	1554	571

Table 5. *Means and standard deviations for the grammaticality judgment task by condition: Accuracy*

	Condition 1 (congruent)		Condition 2 (*adj)		Condition 3 (*art)		Condition 2-3 (*adj or *art)	
	M	SD	M	SD	M	SD	M	SD
Gender:								
Masculine	97.43	4.38	98.65	3.27	97.22	4.25	97.93	3.85
Feminine	97.24	4.61	98.51	3.50	96.05	5.50	97.28	4.76
Morphology:								
Overt	98.23	3.18	98.57	3.09	96.71	4.66	97.64	4.05
N-Overt	96.44	5.36	98.59	3.66	96.56	5.22	97.58	4.61

Table 6. *Means and standard deviations for reaction times for the grammaticality judgement task by condition: Reaction times*

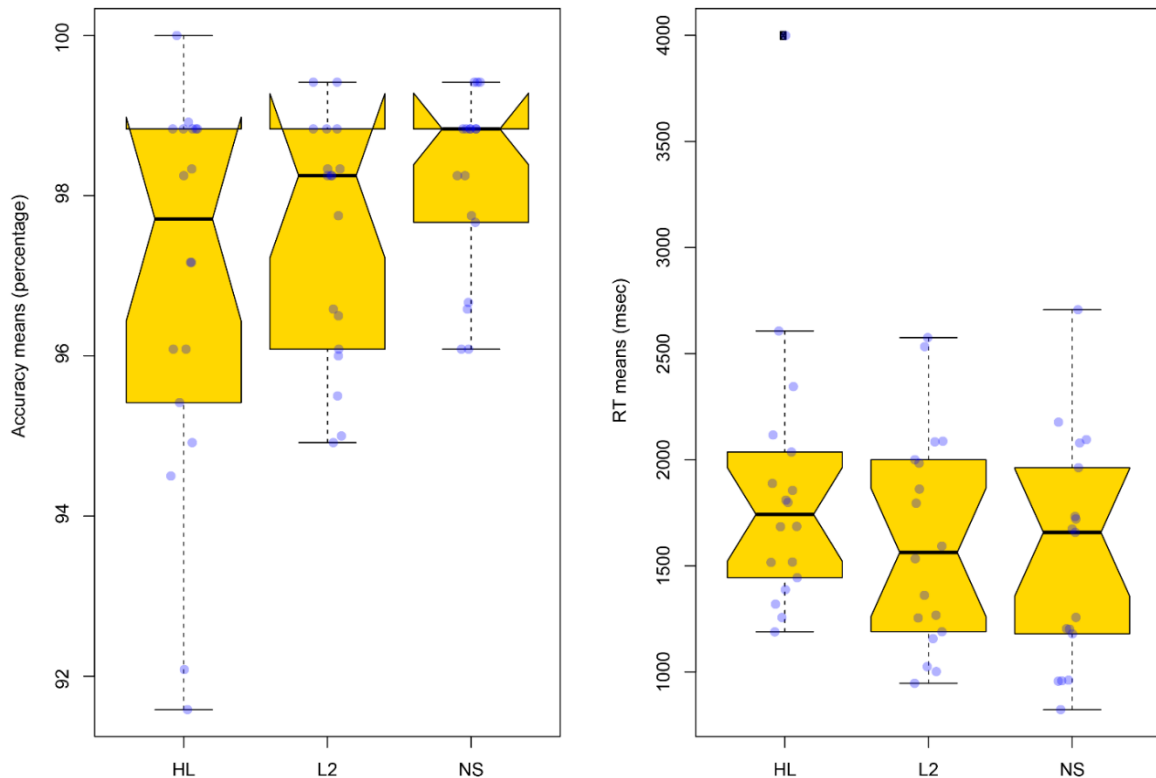|	Condition 1 (congruent)		Condition 2 (*adj)		Condition 3 (*art)		Condition 2-3 (*adj or *art)	
	M	SD	M	SD	M	SD	M	SD
Gender:								
Masculine	1689	571	1679	557	1529	511	1604	539
Feminine	1631	610	1652	537	1634	553	1643	544
Morphology:								
Overt	1564	475	1646	547	1553	500	1597	525
N-Overt	1757	676	1685	548	1610	566	1647	557

Fig. 2. *Individual results for the grammaticality judgment task by group*

For the analysis, outliers were defined as the Accuracy or RT data of any participant whose mean accuracy score or mean RT was more than 2.5 *sd*'s from the mean of that participant's group. The only deleted data were the RT results for one of the HL subjects.

For both accuracy and RT, the data were analyzed in a mixed-design repeated measures ANOVA, with Group (L2, HL, NS) as a between-subjects variable, and Gender (M, F), Morphology (Overt, non-Overt), and Condition (C1 = congruent article, congruent adjective; C2 = congruent article, incongruent adjective; C3 = incongruent article, congruent adjective) as within-subjects variables. The analysis was performed in R (R Core Team 2019), using the ezANOVA software package (Lawrence 2016), which automatically runs tests and corrections for ANOVA data assumptions. Main and interaction effects were considered significant if both their original and GG (Greenhouse-Geisser) adjusted *p*-values were below .05. In addition, significant results of post-hoc pairwise comparison tests are reported by the differences between the two means and the 95% CI surrounding those differences, with Tukey HSD *p*-value adjustments and Cohen's *d* effect sizes.

Regarding accuracy, the only significant main effect was for Condition: $F_{(2, 100)} = 12.284$, $p = .000$, $p(GG) = .000$, *ges* (generalized eta-squared) $= .038$. Pairwise comparisons revealed that this was due to greater accuracy with C2 than with C1: .012, (.003, .022), $p = .009$, $d = .314$, and greater accuracy with C2 than with C3: .019 (.010, .029), $p = .000$, $d = .461$. (See Figure 3.) There was a significant Group x Condition interaction: $F_{(4, 100)} = 6.753$, $p = .000$, $p(GG) = .000$, *ges* $= .042$. In C1, the native speakers were more accurate

than both the L2: .031 (.014, .048), *p* = .000, *d* = .764, and the HL groups: .017 (.000, .035), *p* = .048, *d* = .446. With C3, the L2 learners were significantly more accurate than the heritage speakers: .026 (.007, .045), *p* = .004, *d* = .544. This L2 advantage extended to the circumstance in which the two incongruent conditions, C2 and C3, were pooled: .016 (.004, .028), *p* = .006, *d* = .359. Looking within the groups, the L2 learners were more accurate in both of the incongruent conditions than in the congruent context: C1 x C2: .028 (.012, .044), *p* = .000, *d* = .673; C1 x C3: .021 (.005, .038), *p* = .006, *d* = .469, and, in the pooled context, C1 x C2, C3: .025 (.013, .036), *p* = .000, *d* = .602. Regarding the two incongruent conditions, the L2 group displayed no difference, but both the HL and NS groups were more accurate in C2 than in C3: .027 (.008, .046), *p* = .002, *d* = .560 for the heritage speakers, and .025 (.010, .040), *p* = .000, *d* = .623 for the natives. The native speakers performed similarly with C1 and C2, but still demonstrated a grammaticality effect with C1 x C2, C3: .013 (.001, .024), *p* = .026, *d* = .602.
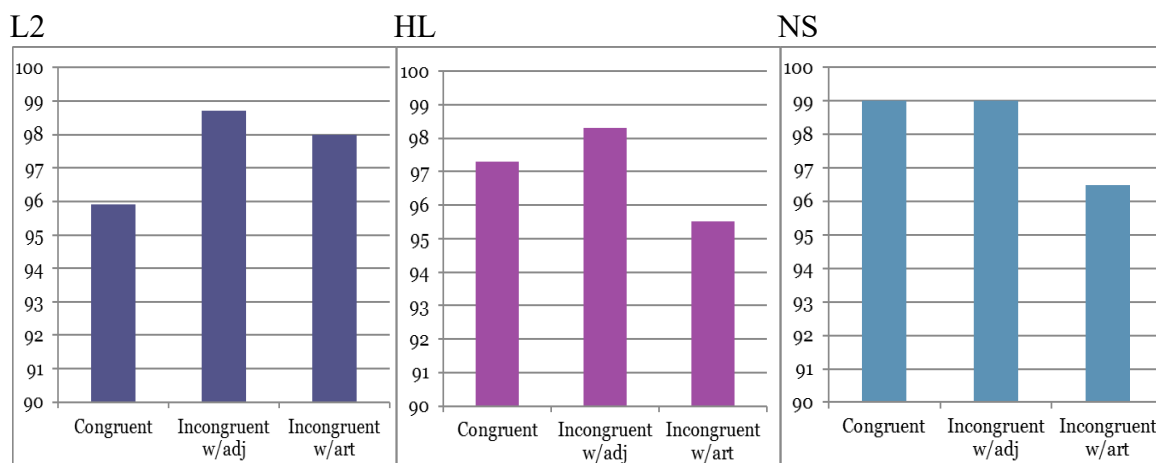
L2   HL   NS

Fig. 3. *Percentage accuracy for condition by group*

   The Group x Gender interaction was also significant: F(2, 50) = 3.737, *p* = .031, *ges* = .013. With feminine nouns, the native speakers were significantly more accurate than the heritage speakers: .023 (.008, .038), *p* = .001, *d* = .511. In addition, the heritage speakers were more accurate with masculine than with feminine nouns: .015 (.002, .028), *p* = .025, *d* = .307. A Morphology x Condition interaction was found, but it fell below significance with the GG correction.

   There was, however, a significant three-way interaction between Group, Morphology, and Condition: F(4, 100) = 2.821, *p* = .029, *p*(GG) = .039, *ges* = .017. This interaction was also found in the RT data, and was also reported by Montrul et al. (2014). For the present data, part of the explanation of this complex interaction can be seen in Figures 4 and 5. Only the L2 group, and only in the congruent condition, C1, displayed substantially lower accuracy rates, and substantially longer reaction times, with non-overt than with overt ending nouns.
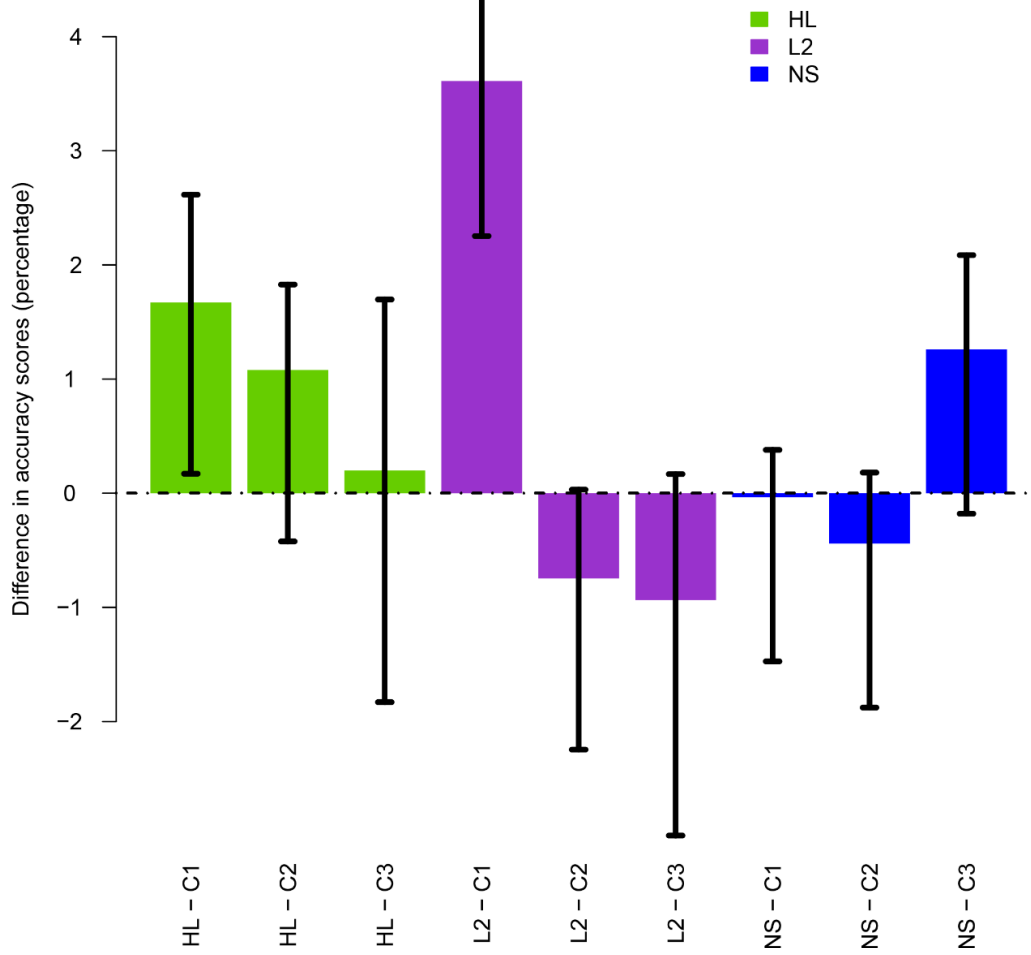
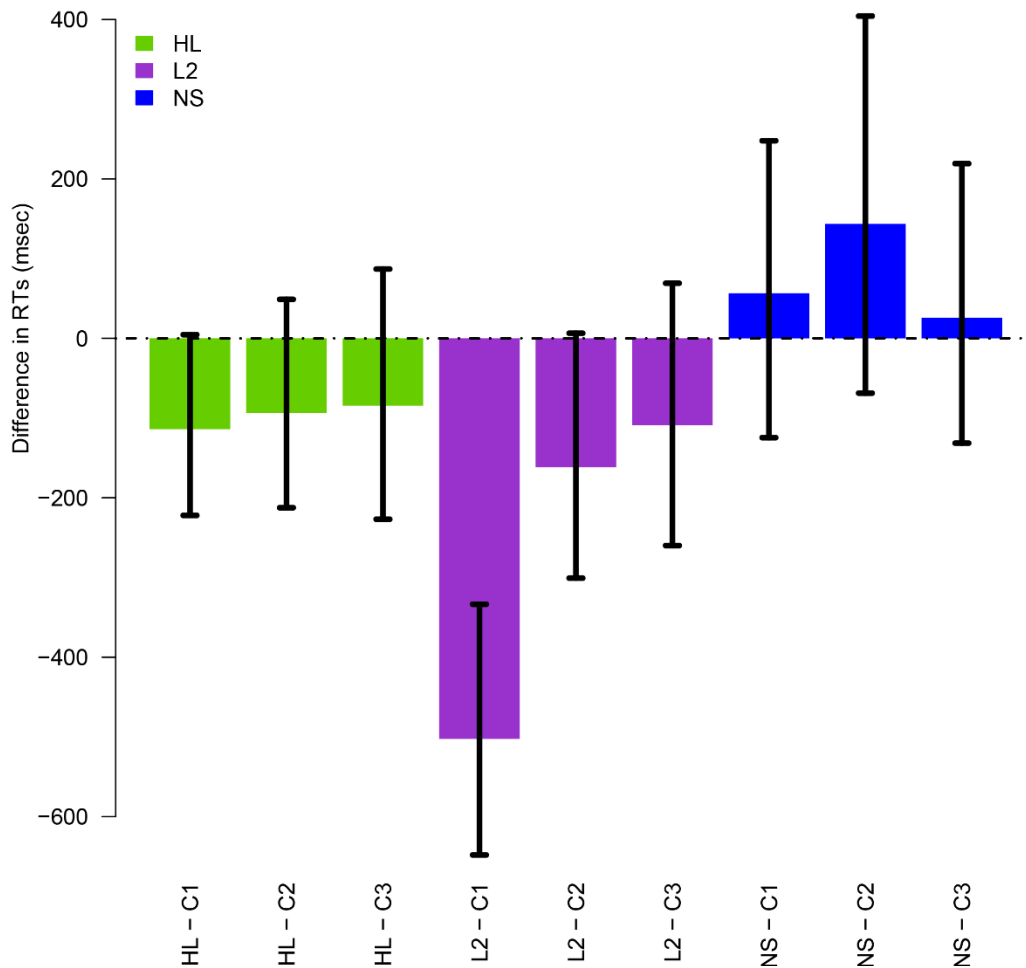Fig. 4. *Difference in accuracy between overt and non-overt noun by group and condition*

Fig. 5. *Difference in reaction times between overt and non-overt nouns by group and condition*

In the RT analysis, there was a main effect of Morphology: $F_{(1, 49)} = 10.822$, $p = .002$, *ges* = .007. Overall, participants were slower with non-overt than with overt nouns: 96.51 (8.903, 183.8), $p = .031$, $d = .173$.

The Group x Morphology interaction was also significant: $F_{(2, 49)} = 11.603$, $p = .000$, *ges* = .016. With non-overt nouns, native speakers were faster than both the L2 learners: 241.610 (49.724, 433.495), $p = .009$, $d = .384$, and the heritage speakers: 269.448 (74.841, 464.055), $p = .004$, $d = .535$. Additionally, the L2 learners were faster than the heritage speakers with overt nouns: 187.725 (23.889, 351.562), $p = .020$, $d = .420$, and they were faster with overt than with non-overt nouns: 257.387 (91.902, 422.873), $p = .002$, $d = .417$. These last two findings contributed to the significant Group x Morphology x Condition interaction: $F_{(4, 98)} = 2.844$, $p = .028$, $p(GG) = .045$, *ges* = .006, which was mentioned above, and which is illustrated in Figure 5.

Condition also exhibited a main effect: $F_{(2, 98)} = 3.390$, $p = .038$, $p(GG) = .048$, *ges* = .005 (see Figure 6), as well as, quite strikingly, significant interactions with each of the other independent variables: Group: $F_{(4, 98)} = 2.676$, $p = .036$, $p(GG) = .048$, *ges* = .008;

Gender: F(2, 98) = 9.256, *p* = .000, *p*(GG) = .000, *ges* = .004; and Morphology: F(2, 98) = 3.766, *p* = .027, *p*(GG) = .041, *ges* = .004.
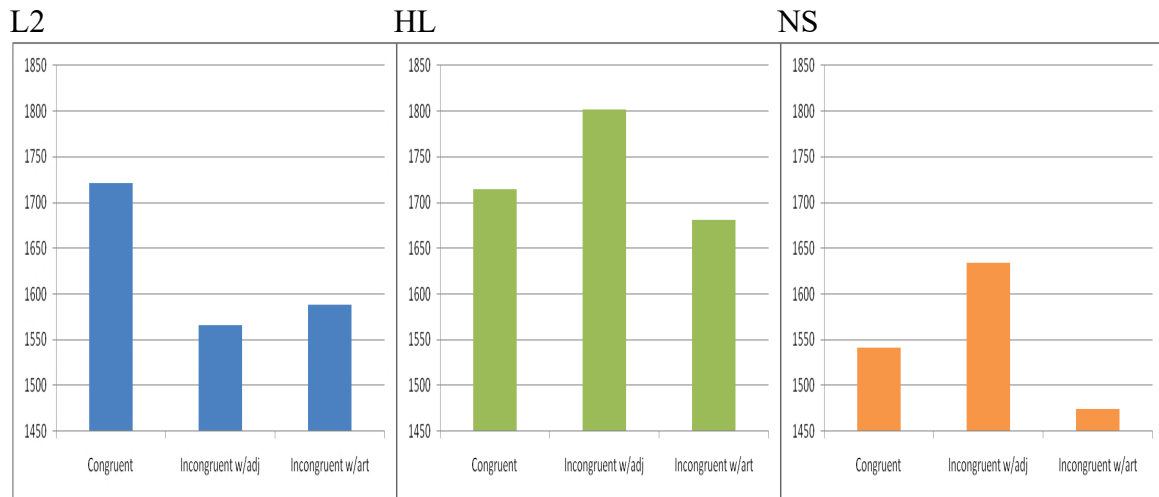

Fig. 6. *Reaction times for condition by group*

Although the Condition x Gender interaction was too weak for the post-hoc pairwise comparison tests to reveal significant differences, the tests did reveal that the Condition x Group interaction reflected that the L2 were faster than the HL speakers with incongruent adjectives (C2): 236.138 (20.538, 451.679), *p* = .028, *d* = .477. This L2 advantage also extended to the overall incongruent context (C2 and C3 pooled): 164.672 (13.942, 315.401), *p* = .028, *d* = .318. In the pooled context, the native speakers were also faster than the heritage speakers: 187.817 (34.950, 340.685), *p* = .011, *d* = .351. The Condition x Morphology interaction resulted from participants being faster with overt than with non-overt nouns in the congruent condition (C1): 192.633 (32.968, 352.298), *p* = .018, *d* = .330.

5. Discussion

The present study investigated the gender agreement behavior of early and late bilinguals through a written, timed grammaticality judgment task that considered the influence of noun gender, morphology, and gender congruency on Spanish gender agreement within a noun phrase.

RQ1. The first research question investigated whether there were any overall differences in NP gender agreement processing between native speakers and highly proficient L2 and HL learners. More specifically, the goal was to see whether early bilinguals displayed an advantage over late bilinguals on a timed grammaticality judgment task. Results indicated no such advantage. On both overall accuracy and RTs, there were no significant statistical differences between the three groups, all of which performed close to ceiling on the experimental task, with accuracy rates above 96%, and statistically indistinguishable RTs, averaging 1635.9 ms. Consequently, the RQ1 finding supports full representational access accounts of L2 acquisition (Schwartz & Sprouse 1996; Prévost & White 2000), since their relatively late age of first exposure to Spanish did not block the L2 learners from native-like attainment in their acquisition of Spanish gender, a feature not

instantiated in their L1. Both early and late bilinguals displayed similar grammatical gender knowledge in their underlying grammars. In addition, this finding also supports the Fundamental Identity Hypothesis (Hopp 2010), a computational access account, since it demonstrates native-like processing of inflectional grammar by L2 learners. Moreover, this result confirms previous offline findings by Alarcón (2011) with the same population, in which both HL and L2 learners were equally accurate on a written gender recognition task, and were indistinguishable from native speakers in their written comprehension of Spanish gender.

For at least three reasons, the high overall accuracy scores and similar reaction times were unsurprising. First, the experimental participants were all of advanced proficiency, which is positively correlated with both high accuracy and, more specifically, with native-like processing of L2 inflection (Hopp 2006). Second, regarding RTs, there is evidence that native speakers, compared with HL and L2 participants, are more prone to slowing down in their performance on timed tasks, perhaps because they are less accustomed to being experimental subjects (cf. Hopp 2010, Experiment 4). This could have had an equalizing effect on the RT results. A third consideration pertains to the ease of the task. The words used were all of high frequency, and the vocabulary test indicated that the participants were familiar with virtually all of them. Nonetheless, since existing L1 and L2 offline findings indicate that gender agreement is acquired later, and with less accuracy, with inanimate than animate nouns (Alarcón 2010; Pérez-Pereira 1991), an effort was made to add a layer of difficulty, and to simultaneously control for the influence of animacy, by including only inanimate nouns. Taking away the opportunity for participants to use biological gender cues in processing gender agreement also enhanced the capacity of the task to tap into exclusively linguistic implicit knowledge.

RQ2. The second research question focused on whether the gender or the overtness of the ending of a noun affected gender agreement processing. Based on previous findings (Montrul et al. 2013; White et al. 2004), it was predicted that all three groups would be more accurate and faster with masculine rather than feminine nouns, and with overt rather than non-overt ending nouns. Regarding accuracy, only the heritage speakers displayed a gender effect: they were less accurate with feminine than with masculine nouns. Surprisingly, however, the present study found that gender had no significant influence on reaction times. In both between and within group comparisons, all three groups were equally fast with both masculine and feminine nouns.

What could explain this almost complete absence of a gender effect? Given that the two experimental groups were of advanced proficiency, a likely part of the explanation relates to proficiency level, which is critical in L2 gender acquisition (McCarthy 2008). In both production and comprehension, White et al. (2004) found high accuracy even from their intermediate learners, and native-like results from their advanced learners. They also reported that, as proficiency increased, the gap between masculine and feminine gender accuracy rates decreased significantly, and that this effect was even more dramatic in the production of Det N Adj phrases: 38.4, 28.1, and 1.7 %, respectively, for low, intermediate, and advanced learners. Proficiency effects on L2 Spanish gender agreement have also been reported in grammaticality judgment studies using ERPs to track morphosyntactic processing (Gillon Dowens et al. 2011). The highest proficiency learners, but not the lowest, were sensitive to gender violations, implying that native-like sensitivity develops

with time and experience with the language (Alemán Bañón et al. 2018). In addition, some psycholinguistic studies have found no effect of gender on adult L1 native speakers of Spanish (Alarcón 2009; Igoa et al. 1999). The present findings suggest that when L2 learners, as well as HL speakers, achieve high levels of proficiency, their gender agreement performance, in both accuracy and reaction time, tends to converge with that of native speakers.

Similarly, there was no overall morphological effect on accuracy. Again, with both within and between group comparisons, all three groups were equally accurate with overt and non-overt noun endings. This could also be a reflection of the high proficiency of the participants, and the ease of the task. More in accordance with previous findings, however, the current study found a significant effect of noun ending on reaction times. Overall, participants were faster with overt than with non-overt nouns. This is largely attributable to the L2 group, since the heritage speakers, and the native speakers as well, were equally fast with both overt and non-overt noun endings. This is consistent with previous findings supporting a robust effect of overt gender morphology in agreement with L2 adults (Alarcón 2010; Franceschina 2001). The present finding that, with non-overt nouns, the native speakers were faster than both the heritage and L2 learners, is also supported by earlier research. Recent L2 studies using visual-world eye tracking tasks (Hopp 2013) offer strong evidence for an interaction effect between morphology and proficiency level. Within a visual-world paradigm, advanced L2 learners of Spanish were able to use the gender information encoded in definite articles to facilitate the processing of upcoming nouns, but only if the nouns were overtly marked for gender (Halberstadt et al. 2018). In this circumstance, high proficiency L2 learners displayed native-like predictive gender processing. The high proficiency HL and L2 learners in the present study performed virtually at ceiling in accuracy with both overt and non-overt noun morphology. But their reaction times were faster with overt morphology, thereby lending support to the predicted facilitatory effect of overtly gender marked nouns. Moreover, although the current design focused on written comprehension, the finding on morphology coincides with that of another study that also measured reaction times of early and late bilinguals using auditory comprehension tasks (Montrul et al. 2014).

RQ3. The third research question explored whether the un/grammaticality of the NP affected gender processing, and whether there were any differences in the processing of incongruent adjectives and incongruent articles. Based on previous findings on the gender congruency effect (Barber & Carreiras 2005; Foote 2011; Sagarra & Herschensohn 2010, 2011; Tokowicz & MacWhinney 2005), it was predicted that all three groups would be more accurate and faster with grammatical rather than with ungrammatical items. The present study had three conditions: congruent, incongruent w/adj, and incongruent w/art. To address the first part of the question, the two incongruent conditions were pooled (incongruent w/art or incongruent w/adj, but not both), so that overall sensitivity to congruence vs. incongruence could be measured. Results indicated that, with respect to both accuracy and RTs, the heritage speakers were insensitive to gender agreement violations. The reaction times of the L2 learners were also unaffected by congruency, but they were more accurate, not less, with ungrammatical rather than grammatical items. More predictably, the native speakers were more accurate in grammatical contexts, although their reaction times were not affected by grammaticality. These results contradict previous findings on Spanish by L1 (Barber & Carreiras 2005; Caffarra et al. 2014); L2 (Alemán

Bañon et al. 2014; Sagarra & Herschensohn 2010, 2011) and HL speakers (Foote 2011; Montrul et al. 2014, Experiments 1 and 2), all of whom displayed lower accuracy rates and/or longer reaction or reading times in incongruent gender contexts. The current results, and other similar findings (Foucart & Frenck-Mestre 2011; Guillelmon & Grosjean 2001), raise the question of why the present L2 and HL groups did not display native-like sensitivity to the gender grammaticality effect?

To begin to answer, consider a study relevant to the present research, which had opposing findings. Using a word-by-word sentence-reading task, Foote (2011) examined gender agreement between a noun and an adjective in both attributive and predicative positions by both early and late bilinguals and native speaker controls. Longer reading times in the incongruent conditions indicated that all three groups were sensitive to gender agreement violations. Why did the bilinguals in the current study not exhibit a congruency effect, as Foote's participants did? One explanation is the extensive formal experience of Foote's L2 participants. Foote herself invoked this argument to account for her results. With just one exception, all of her L2 participants were Spanish instructors who either already had, or were working toward, advanced professional degrees in the language, and had substantial formal instruction and ample practice in the L2. Therefore, their language experience differed considerably from the bilingual learners in the present study, all of whom were undergraduates. With one exception, all of the current L2 learners were Spanish majors, but only 44% of the HL group were either majoring or minoring in Spanish. Consequently, the difference in quantity and quality of linguistic input and learning experiences, including extensive L2 grammar training as teachers, might have made Foote's participants more sensitive to gender agreement errors than the current participants.

Nonetheless, the psycholinguistic literature on L2 gender processing in various languages displays mixed results on the congruency effect. Using a translation recognition task, Bobb et al. (2015) examined article-noun gender agreement by native German speakers and L1 English learners of German at the intermediate and advanced levels. Similar to the present findings on accuracy, only the native speakers showed the expected congruency effect. At both proficiency levels, the L2 learners' accuracy and reaction times were the same in gender congruent and incongruent conditions.

Results are also mixed on whether the presence or absence of L1 gender affects sensitivity to L2 gender violations. Sabourin and Stowe (2008) studied L2 Dutch with learners whose L1 was either similar to Dutch in its gender agreement structure (German) or dissimilar (a Romance language). Their ERP analysis showed a P600 effect only for the L1 German learners, which the authors attributed to L1 influence. Also using ERP's, Foucart and Frenck-Mestre (2011) found that L1 German learners of L2 French showed native-like sensitivity to gender congruence in Art N contexts, but not in N Adj or Adj N conditions. To explain this result, they cited L2 acquisition research findings that adjective agreement is less accurate and acquired later than article agreement (e.g., Bruhn de Garavito & White 2002). Their findings, though, are in line with research showing a gender congruency effect for Art N gender agreement, which suggests native-like processing by high proficiency L2 learners (Gillon Dowens et al. 2010). In a study including both early and late bilinguals, Guillelmon and Grosjean (2001) examined the congruency effect in noun phrases on a spoken word recognition task. Their results showed that both native

speakers and early bilinguals were sensitive to gender violations, but that late bilinguals were not. The authors concluded that, in addition to later age of acquisition and less frequent L2 exposure, L2 learners also have problems accessing gender information to facilitate the word recognition process (Carroll 1989).

This problem with automatic activation of gender knowledge in real time processing concurs with Grüter et al.'s (2012) claim that, in their mental lexicons, L2 learners have weaker associations between a noun and its gender node than L1 speakers. During online processing, then, L2 learners make less effective use of gender cues, such as the article and adjective, when producing and/or comprehending a noun phrase. Although this claim does not contradict L2 findings of native-like sensitivity to gender violations, it does emphasize the use of gender-marking cues within the noun phrase, such as the article and adjective, to facilitate retrieval of a noun, as a basic distinction between native and non-native speakers. Consequently, difficulties with rapid retrieval of gender information during real-time processing could explain why the experimental groups in the present study did not exhibit a gender congruency effect. Recall that these early and late bilinguals already possess native-like knowledge of Spanish gender, as indicated by an offline comprehension task (Alarcón 2011). But native-like knowledge of gender does not necessarily imply native-like processing. Although the overall accuracy and reaction times of the three current groups on the grammatical judgment task were similar, only the native speakers displayed sensitivity to gender agreement errors by being significantly less accurate in incongruent conditions.

The possibility that the HL and L2 groups might differ in their sensitivities to gender agreement errors with articles and with adjectives would not be revealed by the overall congruent/incongruent analysis, which pooled together the article and adjective errors. Therefore, the second part of the third research question, which distinguishes Art N and N Adj gender errors, might shed additional light on the participant groups' real online performance. Based on previous findings in L1 and L2 acquisition research (Bruhn de Garavito & White 2002; López-Ornat 1997), it was predicted that all three groups would be more accurate and faster in processing incongruent articles than incongruent adjectives. Regarding reaction times, however, none of the three groups displayed significant differences between incongruent articles and incongruent adjectives. In terms of accuracy, though, both the native and heritage speakers were less accurate with mismatched articles than with mismatched adjectives, while the L2 learners were equally accurate in both domains. Also, unexpectedly, the L2 subjects were more accurate in the incongruent conditions than in the congruent context, a reverse congruency effect. The native and heritage speakers patterned together in their sensitivity and directionality, while the L2 learners patterned differently.

These findings immediately raise at least two important questions. First, how to explain that the L2 learners were more accurate with incongruent rather than congruent conditions, and, within the incongruent conditions, were insensitive to domain, performing equally accurately with incorrect articles and incorrect adjectives? Second, why did the heritage group pattern so closely to the native speakers, rather than to the L2 learners?

To begin, it is not unprecedented to find that even advanced L2 learners display a reverse grammaticality effect. For example, Montrul et al. (2014) found that L2 learners, on a word repetition task, were faster in repeating ungrammatical than grammatical non-overt nouns. But, as Segalowitz observed, faster processing does not necessarily imply

automatic processing. So it could be that these L2 learners were exhibiting "faster *non-automatic* processing" (2003: 385), despite using explicit rather than implicit strategies.

Furthermore, there is evidence that L2 learners of Spanish differ from native speakers in their lesser use of morphosyntactic processing cues. For example, native speakers, both children and adults, use gender-marked articles as predictive cues to establish reference more quickly (Lew-Williams & Fernald 2007). The association between article and noun, as a unified unanalyzed chunk in L1 lexicon, is formed early in childhood, and is reinforced through distributional learning mechanisms (Aslin & Newport 2014). For post-pubescent L2 learners, this association might not become as strong in their L2 lexicon. Unlike native speakers, L2 learners do not use gender-marked articles to facilitate noun recognition (Hopp 2013; Lew-Williams & Fernald 2010). This difference between native and non-native predictive use of gender cues is manifested in slower gender retrieval in real time L2 production, and in less effective use of gender cues in real time L2 processing. (See Grüter et al. 2012 for a full discussion.) Furthermore, rather than storing gender as a fixed feature in the lexicon, as L1 speakers do, Bordag and Pechmann propose that for L2 learners gender is "computed each time anew, when needed, on the basis of various available pieces of information, for example the phonological form of the word" (2008: 156). This raises the question of whether L2 learners process gender qualitatively like native speakers. In order to use gender knowledge quickly and accurately during real time processing, this abstract knowledge needs to be stored and retrieved automatically. Automatic competence (or integrated knowledge) is the "ability to apply one's linguistic knowledge spontaneously in both the productive and receptive use of language" (Jiang 2007: 2). Important sources for the integration of knowledge, according to Jiang, are language exposure (input) and experiences (interaction).

This brings the HL speakers into the discussion, since the key definitional distinction between them and L2 learners is age of acquisition, which is highly correlated with both type of learning environment, naturalistic versus formal classroom instruction, and type of input, spoken versus written. In the current study, although neither group displayed the expected overall gender congruency effect (cf. Bobb et al. 2015; Guillelmon & Grosjean 2001; Sabourin & Stowe 2008; Foucart & Frenck-Mestre 2011), the heritage speakers, like the native speakers, and unlike the L2 learners, were less accurate with incongruent articles than with incongruent adjectives. A similar result, albeit for reaction times, was found by Montrul et al. (2014). Although their focus was on morphology, and their input was aural, the authors found that their HL and L2 participants performed equally on a gender monitoring and a grammaticality judgment task, but not on a word repetition task. The HL learners and native speakers repeated both congruent and incongruent non-overt nouns equally fast, but the L2 learners repeated incongruent non-overt nouns more quickly than congruent ones. One of the explanations the authors put forth for this anomalous result concerns differences in learning environment and in type of input. Montrul et al. concluded that heritage speakers show more native-like patterns than L2 learners in tasks requiring implicit knowledge of the L2, such as their word repetition task. One possible explanation for the current finding is that heritage speakers have integrated knowledge (Jiang 2007), which, due to their early acquisition, is stored in procedural memory, while L2 learner's knowledge is stored in declarative memory (see Morgan-Short et al. 2014). Similarly, one could argue that gender knowledge is implicit for heritage speakers, but explicit for L2

learners (Ellis 2005). All of this is consistent with Montrul's (2009) claim that early bilinguals are more successful than late bilinguals with tasks requiring automatized and integrated implicit knowledge (e.g., timed oral imitation), while late bilinguals are more successful on tasks demanding explicit metalinguistic knowledge (e.g., gender monitoring). Therefore, different learning experiences, including age of acquisition and mode of input (cf. Foote 2011), might partially account for the current results.

The present study used a timed grammaticality judgment task to measure implicit knowledge. If Montrul's proposed relationship between type of task and type of knowledge holds, heritage learners "may have more implicit knowledge of the language than the L2 learners," which suggests that early bilinguals "may have approached processing of input and the related language learning process through different learning mechanisms" (2009: 249-250). The present results differentiated between the L2 learners, who were equally accurate with incorrect articles and adjectives, and the heritage and native speakers, who were less accurate with incorrect articles than with incorrect adjectives (cf. Bowles 2011; Godfroid et al. 2015). Although gender information is available and accessible to L2 learners, they do not use it in the same way as heritage and native speakers do. There seems to be a disconnection between their implicit knowledge of gender and their ability to use this knowledge under time pressure, which shows that L2 gender processing is still challenging at the advanced proficiency level. This disconnection does not necessarily imply a representational deficit of gender in the L2 grammar, since it might involve lexical issues, such as lower activation of the gender information stored in the L2 mental lexicon (Schriefers & Jescheniak 1999). According to Costa et al. (2003), there are two main perspectives on gender retrieval from the lexicon. Activation-level dependency models (e.g., Schriefers 1993) contend that the efficiency with which gender is retrieved depends on its activation, which is conceptualized as the closeness of the association between a particular noun and its gender. Alternatively, the automatic gender-access model (e.g., Schiller & Caramazza 2003) posits that a noun and its gender are learned together, and therefore are indelibly linked, so that gender is directly and automatically retrieved with the noun itself. Both models suggest that the dissociation between implicit L2 gender knowledge and L2 processing involves lexical level issues that have consequences for predictive gender processing. Recall that the task in the present study investigated gender agreement at the phrasal (Art N Adj), not the sentential level, so concerns regarding structural and syntactic distance were not relevant (cf. Keating 2009). This strengthens the case that the L2 processing problems revealed in the current study are rooted in lexical access. We have already seen that the L2 results could be accounted for by less than native-like usage of morphosyntactic cues, and, more generally, by reduced automaticity in lexical storage and retrieval, especially as it relates to the development of article-noun pairs in the L1 and L2 lexicons. As Lew-Williams and Fernald maintain, "age- and experience-related differences suggest that L1 and L2 learners follow different trajectories in learning about grammatical gender, resulting in differences between L1 and L2 adults' knowledge and processing of Spanish gender agreement" (2010: 460).

6. Conclusion

The current findings suggest that high proficiency L2 and HL learners, who have similar implicit knowledge of grammatical gender, process gender agreement differently during online comprehension. The timed grammaticality judgment task revealed native-

like processing patterns by the early but not by the late bilinguals. The L2 learners did process gender, but possibly through entirely different mechanisms. The current findings and discussion suggest that differences in efficient processing might have resulted from several factors: age of acquisition, language exposure, and learning experiences, all of which influence how gender knowledge is integrated and automatized. Based on the current results, the L2 learners probably did not process gender agreement automatically. Despite having gender knowledge in their underlying grammars, they did not show native-like patterns in real time processing. So even advanced L2 learners might still need extended training and substantial practice to attain native-like automaticity in target gender use. Evolving from conscious, controlled behavior to automatic activity is itself a slow and gradual process. Perhaps native-like processing of gender agreement within the noun phrase is only possible after L2 learners have mastered the target gender system, and achieved near-native proficiency (see Dussias et al. 2013; Gillon Dowens et al. 2010; Gillon Dowens et al. 2011; Hopp 2013). Heritage language evidence supporting this claim is found not only in the present study, but in Foote (2011) and Montrul et al. (2014) as well.

Nonetheless, current findings must be seen as merely a beginning of online psycholinguistic investigation into the nuanced distinctions between early and late bilinguals of advanced proficiency. These distinctions presumably extend far beyond the narrow focus of the present study on gender agreement processing, which itself needs to be examined in greater depth, to include basic differences in the mechanisms used for processing a wide range of linguistic behavior. Consequently, such investigations could potentially advance our understandings of both L2 and L1 acquisition.

Irma Alarcón
Department of Spanish and Italian
Wake Forest University
1834 Wake Forest Rd.
Winston-Salem, NC 27109
alarcoi@wfu.edu
(+1)336-758-5194

References

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

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2018). Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PLoS ONE.* (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200791) https://doi.org/10.1371/journal.pone.0200791

Antón-Méndez, M.I. (1999). *Gender and number agreement processing in Spanish*. Unpublished doctoral dissertation. University of Arizona. Arizona.

Antón-Méndez, M.I., Nicol, L.J., & Garrett, M.F. (2002). The relation between gender and number agreement processing. *Syntax*, *5*(1), 1-25. https://doi.org/10.1111/1467-9612.00045

Aslin, R.N., & Newport, E.L. (2014). Distributional language learning: Mechanisms and models of category formation. *Language Learning*, *64*(2), 86-105. https://doi.org/10.1111/lang.12074

Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data. *Hispania*, *92*(4), 814–828.

Alarcón, I. (2010). Gender assignment and agreement in L2 Spanish: The effects of morphological marking, animacy, and gender. *Studies in Hispanic and Lusophone Linguistics*, *3*(2), 267-299.

Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, *14*(3), 332–350.

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

Bates, E., Devescovi, A., Hernández, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics, 58*(7), 992-1004.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (eds.), *Linguistic perspectives on second language acquisition*. Cambridge: Cambridge University Press, 41-68. https://doi.org/10.1017/CBO9781139524544.005

Bobb, S.C., Kroll, J., & Jackson, C. (2015). Lexical constraints in second language learning: Evidence on grammatical gender in German. *Bilingualism, Language, and Cognition*, *18*(3), 502-523. https://doi.org/10.1017/S1366728914000534

Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research*, *24*(2), 139-166. https://doi.org/10.1177/0267658307086299

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge. What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*, 247-271. https://doi.org/10.1017/S0272263110000756

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*. Dordrecht: Kluwer, 153-178. https://doi.org/10.1007/978-94-010-0291-2_6

Caffarra, S., Janssen, N., & Barber, H.A. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. *Neuropsychologia*, *63*, 124-134. https://doi.org/10.1016/j.neuropsychologia.2014.08.016

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

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3-42. https://doi.org/10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107-126. https://doi.org/10.1017/S0142716406060024

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181-200.

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

Domínguez, A., Cuetos, F., & Segui, J. (1999). The processing of grammatical gender and number in Spanish. *Journal of Psycholinguistic Research*, *28*(5), 485-498.

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.

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

Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, *32*, 187-220. https://doi.org/10.1017/S0142716410000342

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism, Language, and Cognition*, *14*(3), 379-399. https://doi.org/10.1017/S136672891000012X

Franceschina, F. (2001). Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. *EUROSLA Yearbook*, *1*, 143-158. https://doi.org/10.1075/eurosla.1.12fra

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

Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish: Evidence from event related potentials. *Neuropsychologia*, *49*(7), 1651-1659.

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

Godfroid, A., Loewen, S., Jung, S. Park, J-H., Gass, S., & Ellis. R. (2015). Time and untimed grammaticality judgements measure distinct types of knowledge. Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*, 269-297.

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.

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

Halberstadt, L., Valdés Kroff, J.R., & Dussias, P.E. (2018). Grammatical gender processing in L2 speakers of Spanish. The role of cognate status and gender transparency. *Journal of Second Language Studies*, *1*(1), 5-10. https://doi.org/10.1075/jsls.17023.hal

Hawkins, R., & Chan, Y. -H. C. (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis.' *Second Language Research*, *13*, 187-226. https://doi.org/10.1191/026765897671476153

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369-397. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/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. https://doi.org/10.1177/0267658312461803

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. Studies in *Second Language Acquisition*, *40*, 171-199.

Igoa, J.M., García-Albea, J.E., & Sánchez-Casas, R. (1999). Gender-number dissociations in sentence production in Spanish." *Rivista di Linguistica*, *11*(1), 163-196.

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

Keating, G. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (eds.), *Research in Second Language Processing and Parsing*. Amsterdam/Philadelphia: John Benjamins, 113-134. https://doi.org/10.1075/lald.53.05kea

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

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: A case study*. Mahwah, NJ: Erlbaum.

Lawrence, M.A. (2016). Ez: Easy analysis and visualization of factorial experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

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*, 447-464.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193-198.

López-Ornat, S. (1997). What lies in between a pre-grammatical and a grammatical representation? Evidence on nominal and verbal form-function mappings in Spanish from 1;7 to 2;1. In A.T. Pérez-Leroux & W.R. Glass (eds.), *Contemporary perspectives on the acquisition of Spanish, Vol. 1: developing grammars*. Somerville, MA: Cascadilla, 3-20.

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

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

Montrul, S. (2009). Reexamining the fundamental difference hypothesis. What can early bilinguals tell us? *Second Language Acquisition Studies*, *31*, 225-257.

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, *30*, 3-23. https://doi.org/10.1017/S0267190510000103

Montrul, S. (2013). How "native" are heritage speakers? *Heritage Language Journal*, *10*(2), 153-177.

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139030502

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2013). The role of experience in the acquisition and production of diminutives and gender in Spanish: Evidence from L2 learners and heritage speakers. *Second Language Research*, *29*(1), 87-118.

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilities the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition*, *17*(1), 118-138.

Montrul, S., Foote, R., & Perpiñan, 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.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K.A., Carpenter, H., & Wong, P.C.M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition, 17*(1), 56-72. https://doi.org/10.1017/S1366728912000715

Neary, L.C. (2001). *Lexical Representation and Processing in Second Language Learners*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, *18*(3), 571-590. https://doi.org/10.1017/S0305000900011259

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

R Core Team. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. URL https://www.R-project.org/

Sabourin, L., & Stowe, L.A. (2008). Second language processing: When are the first and second languages processed similarly. *Second Language Research*, *24*(3), 397-430. https://doi.org/10.1177/0267658308090186

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

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, *61*(1), 80-116. https://doi.org/10.1111/j.1467-9922.2010.00588.x

Segalowitz, N. (2003). Automaticity and second languages. In J. Doughty, and M.H. Long (eds.), *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing, 382-408. https://doi.org/10.1002/9780470756492.ch13

Schiller, N.O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*, 168-194. https://doi.org/10.1016/S0749-596X(02)00508-9

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 841-850. https://doi.org/10.1037/0278-7393.19.4.841

Schriefers, H., & Jescheniak, J.D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research,* *28*(6), 575-600. https://doi.org/10.1023/A:1023264810403

Schriefers, H., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1368-1377. https://doi.org/10.1037/0278-7393.26.6.1368

Schwartz, B. D., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*, 40-72. https://doi.org/10.1177/026765839601200103

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition*, *27*, 173-204. https://doi.org/10.1017/S0272263105050102

Wicha, N.Y.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. https://doi.org/10.1162/0898929041920487

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

Appendix
Experimental noun phrases (congruent condition)

Masculine Nouns: Over & Non-overt		Feminine Nouns: Overt & Non-overt	
El trabajo digno The dignified job	*El país pequeño* The small country	*La idea concreta* The concrete idea	*La verdad profunda* The deep truth
El progreso lento The slow progress	*El informe técnico* The technical report	*La historia pasada* The past history	*La versión moderna* The modern version
El consejo práctico The useful advice	*El hospital viejo* The old hospital	*La noticia mala* The bad news	*La virtud sencilla* The simple virtue
El sueño extraño The strange dream	*El bosque lejano* The distant wood	*La vivienda antigua* The old house	*La bondad inmensa* The big kindness
El dinero perdido The lost money	*El crimen distinto* The different crime	*La iglesia típica* The typical church	*La crisis violenta* The violent crisis
El talento único The unique talent	*El humor clásico* The classic humor	*La escena íntima* The intimate scene	*La emoción perfecta* The perfect emotion
El premio famoso The famous award	*El ataque directo* The direct attack	*La lluvia intensa* The intense rain	*La ilusión compleja* The complex illusion
El asiento cómodo The cozy seat	*El hotel barato* The cheap hotel	*La playa privada* The private beach	*La situación previa* The prior situation
El rostro curioso The curious face	*El factor humano* The human factor	*La dieta ligera* The light diet	*La sociedad justa* The just society
El viento tremendo The huge wind	*El valle hermoso* The beautiful valley	*La conducta rara* The odd behavior	*La razón correcta* The correct reason
El conflicto interno The inner conflict	*El origen secreto* The secret origin	*La venta segura* The certain sale	*La pasión eterna* The eternal passion
El mercado limpio The clean market	*El viaje futuro* The future trip	*La infancia corta* The brief childhood	*La señal opuesta* The opposite sign
El castigo duro The hard abuse	*El límite preciso* The precise limit	*La revista diaria* The daily magazine	*La salud pública* The public health
El delito crítico The critical felony	*El reloj exacto* The accurate clock	*La promesa falsa* The false promise	*La dosis completa* The complete dose
El barrio diverso The diverse block	*El ambiente fino* The fine atmosphere	*La cerveza clara* The clear beer	*La juventud unida* The united youth