

Early language experience facilitates the processing of gender agreement in Spanish heritage speakers*

SILVINA MONTRUL
University of Illinois at Urbana–Champaign
JUSTIN DAVIDSON
University of Illinois at Urbana–Champaign
ISRAEL DE LA FUENTE
Université Paris Diderot – Paris 7
REBECCA FOOTE
University of Illinois at Urbana–Champaign

(Received: August 10, 2012; final revision received: January 13, 2013; accepted: February 21, 2013; first published online 8 April 2013)

We examined how age of acquisition in Spanish heritage speakers and L2 learners interacts with implicitness vs. explicitness of tasks in gender processing of canonical and non-canonical ending nouns. Twenty-three Spanish native speakers, 29 heritage speakers, and 33 proficiency-matched L2 learners completed three on-line spoken word recognition experiments involving gender monitoring, grammaticality judgment, and word repetition. All three experimental tasks required participants to listen to grammatical and ungrammatical Spanish noun phrases (determiner–adjective–noun) but systematically varied the type of response required of them. The results of the Gender Monitoring Task (GMT) and the Grammaticality Judgment Task (GJT) revealed significant grammaticality effects for all groups in accuracy and speed, but in the Word Repetition Task (WRT), the native speakers and the heritage speakers showed a grammaticality effect, while the L2 learners did not. Noun canonicity greatly affected processing in the two experimental groups. We suggest that input frequency and reduced language use affect retrieval of non-canonical ending nouns from declarative memory in L2 learners and heritage speakers more so than in native speakers. Native-like processing of gender in the WRT by the heritage speakers is likely related to context of acquisition and particular experience with oral production.

Keywords: grammatical gender, Spanish, heritage speakers, processing, implicit/explicit tasks

Introduction

In the context of the United States, a heritage speaker is an individual who was exposed to a minority language at home with the family in early childhood, but as a young adult has become dominant in the majority language, in this case English. If English was acquired together with the minority language or later as a second language (L2), it became the primary language sometime in late childhood. The first language (L1), or one of the two first

languages in cases of simultaneous bilinguals, became the secondary and less dominant or weaker language. Some heritage speakers may be able to acquire their language fully in childhood, but many do not. Consequently, the heritage language resembles a L2, in the sense that it displays lexical and grammatical errors typical of earlier stages of language development and has not reached the full ultimate attainment of a L1 acquired in childhood. Although heritage speakers are a type of native speakers by virtue of having been exposed to the language since birth (Montrul, in press), several studies have shown that heritage speakers differ from fully fluent native speakers living in their home country or who immigrated in adulthood on many linguistic dimensions, including pronunciation (Au, Knightly, Jun & Oh, 2002), lexical repertoire and lexical access (Hulsen, 2000; Montrul & Foote, published online May 8, 2012; Polinsky, 1997, 2008), command of inflectional morphology and complex syntax (Albirini, Benmamoun & Saddah, 2011; Bolonyai, 2007; Montrul, 2002, 2004a, 2007, 2009; Montrul & Bowles, 2009; O’Grady, Kwak, Lee & Lee, 2011; O’Grady, Lee & Choo, 2001; Polinsky, 2011; Rothman, 2007), semantics (Montrul & Ionin, 2010), discourse and pragmatics (Otheguy, Zentella & Livert, 2007; Silva-Corvalán, 1994), among other features.

* This study was supported by internal funds from the University of Illinois, for which we are grateful. We thank all the students who participated in our experiments, as well as the undergraduate research assistants Celeste Larkin, Kayla Pennoyer, Adam Bethune and Rachel Pirovano for their help in setting up the experiments and running subjects. Earlier versions of this work were presented at the 2010 Hispanic Linguistics Symposium at Indiana University, at the 2011 Workshop on Heritage Languages at Harvard University, at *EUROSLA 2011* in Stockholm, at the Department of Spanish, Italian and Portuguese colloquium at the University of Illinois (November 2011), at the 2012 Heritage Language Summer Institute at UCLA and at the Psycholinguistics Supper series at the CUNY Graduate Center in October 2012. We are grateful to all the audiences in these venues for their helpful comments and feedback. We are also grateful to Carmen Silva-Corvalán and to the three anonymous reviewers who kindly provided very useful comments and suggestions.

Address for correspondence:

Silvina Montrul, Department of Linguistics and Department of Spanish, Italian & Portuguese, University of Illinois at Urbana–Champaign, 4080 Foreign Languages Building, MC-176, 707 South Mathews Avenue, Urbana, IL 61801, USA
montrul@illinois.edu

Many of the non-native patterns displayed by heritage speakers resemble the grammatical patterns typical of adult L2 learners who are either in the process of learning the L2 or have ceased development (and fossilized). These observations have generated an intense interest in understanding whether and how heritage speakers differ from L2 learners in their linguistic abilities, a question that carries important theoretical and practical significance.

According to several theoretical accounts of adult L2 acquisition, maturational effects (i.e., age of acquisition) explain fundamental differences between L1 acquisition by children, which under normal circumstances always results in native-like knowledge, and L2 acquisition by adults, which does not guarantee uniform success (Abrahamsson & Hyltenstam, 2009; DeKeyser, 2000; Long, 2007; Hawkins & Chan, 1997; Tsimpli & Dimitrakopoulou, 2007). It is often assumed that adult L2 learners are unable to reach native-like competence in the L2 due to the fact that they start to learn the L2 past puberty, when the linguistic and cognitive mechanisms involved in language learning in childhood are no longer operative or available. This in turn leads to differences in the nature of linguistic knowledge (Bley-Vroman, 2009), in patterns of language processing (Clahsen & Felser, 2006), and in degree of ultimate attainment (Abrahamsson & Hyltenstam, 2009). What is interesting about heritage speakers is that they were exposed to the (minority) language since birth, and presumably had access to the cognitive, linguistic, and processing language learning mechanisms assumed not to be available to adult L2 learners. Yet many heritage speakers and L2 learners display seemingly similar non-native patterns. By comparing the linguistic abilities of heritage speakers and adult L2 learners in their secondary, less dominant language, we can re-evaluate the role of age in bilingual language development as well as other potential experiential factors that may be confounded with age of acquisition.

As far as linguistic experience is concerned, there are important differences and similarities between heritage speakers and L2 learners (Montrul, 2008, 2010). To become a fluent native speaker, individuals must be born in a linguistic environment where the language is spoken, they must be exposed to the language most of the time, and they must use it in a variety of contexts and social situations on a daily basis. They also typically receive schooling in the native language. By contrast, both L2 learners and heritage speakers receive much less exposure to the language on a daily or even weekly basis, and the amount of exposure and frequency of use of the language can vary widely from speaker to speaker. In addition, language exposure and opportunities for language use in these two groups are restricted to particular environments and social contexts: the home and family, for heritage

speakers, and typically the foreign language classroom for L2 learners.

In this study we investigate whether, in addition to age of acquisition, early language experience brings advantages to Spanish heritage speakers in their knowledge of early-acquired aspects of morphosyntax when compared to L2 learners of Spanish, who initiated acquisition of the language around or after puberty. By advantage, we mean behavioral performance closer to native-speaker norms. This question was originally posed by Au et al. (2002), who conducted an experimental study of incipient L2 learners of Spanish and Spanish heritage speakers with receptive knowledge of Spanish (OVERHEARERS). Au et al. (2002) examined phonological and morphosyntactic abilities assessed through Voice Onset Time (VOT) measurements and a grammaticality judgment task, and concluded that early language experience gave heritage speakers an edge in phonology but not in morphosyntax. To date, however, findings from other recent studies have been inconsistent with respect to morphosyntax. Studies that controlled for proficiency have found advantages for heritage speakers over L2 learners in some grammatical areas (Bruhn de Garvito, 2002; Håkansson, 1995; Montrul, 2010), while others have found advantages for heritage speakers in certain tasks only (Alarcón, 2011; Bowles, 2011; Montrul, Foote & Perpiñán, 2008).

The present study focuses on an aspect of Spanish morphosyntax – gender agreement – for three principal reasons. First, gender agreement is mastered by monolingual Spanish-speaking children before age three (see works cited in Montrul, 2004b). Second, adult native speakers of Spanish rarely make gender agreement errors, especially with words they know. Third, even very advanced post-puberty L2 learners of Spanish, including near-native speakers (Franceschina, 2001; Grüter, Lew-Williams & Fernald, 2012), rarely master gender agreement at the level of a native speaker, especially with nouns that are irregular or non-canonical (e.g., *el puente* “the bridge”, *la nariz* “the nose”; i.e., nouns that do not end in *-o* if masculine, or in *-a* if feminine). Finally, gender agreement is problematic for heritage speakers as well, even though they were exposed to Spanish in their early childhood. An important recent study comparing L2 learners and heritage speakers’ knowledge of gender in Spanish is Montrul et al. (2008). Montrul et al. showed that Spanish L2 learners and heritage speakers made more errors with gender agreement with feminine than with masculine nouns, and more errors especially with non-canonical ending nouns than with canonical ending nouns in written production, written comprehension, and oral production. Another main finding of this study was a task effect: the L2 learners were more target-like than the heritage speakers in the two written tasks, while the heritage speakers were more

target-like than the L2 learners in the oral production task. Considering the modality, type, and timing of required response, and the explicitness of each task (Bialystok & Ryan, 1985; Ellis, 2005), the untimed written tasks used by Montrul et al. (2008) may have tapped the L2 learners' explicit, and even metalinguistic, knowledge of gender. The oral task, by contrast, seems to tap into a more implicit type of knowledge. But because the explicitness or implicitness of the tasks was confounded with modality, it is not clear whether the heritage speakers were better at the implicit task than the L2 learners because the task elicited oral production, or because it was tapping into implicit grammatical knowledge of Spanish gender. Similarly unclear is whether the L2 learners did better than the heritage speakers in the more explicit tasks because the tasks were written, or instead because they were more controlled and tapped into explicit and metalinguistic knowledge of gender in Spanish. Since theoretical debates on the role of maturational effects in L2 acquisition specifically concern implicit knowledge, it is crucial to understand the types of implicit or explicit knowledge that different tasks tap into in L2 learners and in heritage speakers and, additionally, how the implicit/explicit dimension of the task interacts with the participants' age of acquisition.

The present study makes two important contributions. First, it addresses more directly whether early language experience related to age of acquisition confers an advantage to heritage speakers over L2 learners on implicit knowledge of gender agreement in Spanish. Unlike previous studies, we approach this question by controlling for modality and degree of explicitness of the tasks. Specifically, we focus on the automatic processing of gender agreement in auditory recognition and oral repetition, by using three on-line tasks with systematic variation on the type of response required in each task. Such manipulation elucidates the type of processing that occurs when there is more or less conscious attention on, or metalinguistic awareness of, gender. We provide evidence that heritage speakers display more native-like performance than the L2 learners, behaving like the native speakers in the most implicit task. This conclusion is further corroborated by the results of a related study with the same participants, reported in Montrul, de la Fuente, Davidson and Foote (2013).

The second important contribution of the present study is its additional focus on the canonicity of noun endings, since native-like knowledge of gender implies knowing the gender of both canonical and non-canonical ending nouns and being able to produce agreement correctly. This issue has not been properly addressed in L2 acquisition. Montrul et al. (2008) found that L2 learners and heritage speakers showed a strong sensitivity to the canonicity of the noun ending, but the vast majority of studies arguing that L2 learners are able to acquire and process gender

like native speakers (e.g., Sagarra & Herschensohn, 2011; White, Valenzuela, Kozłowska-Macgregor & Leung, 2004) have focused on regular ending nouns. Our study shows that these two types of nouns are processed differently by L2 learners and heritage speakers. We discuss how this finding relates to current theoretical models of regular and irregular morphology. Although age of acquisition remains a potential explanation for our overall findings, we consider how experiential factors may contribute to the observed processing differences between heritage speakers and L2 learners.

Gender agreement in Spanish

Spanish animate and inanimate nouns are arbitrarily classified into masculine and feminine in the lexicon, and the exponents of gender marking follow formal rules. Approximately 96.3% of feminine nouns end in the word marker *-a* (which can be construed as an inflectional morpheme as in *señor* “man.MASC” – *señora* “woman.FEM”, or the last vowel or word marker of a root as in *cara* “face.FEM”) and approximately 99.8% of masculine nouns end in the word marker *-o*, as in *hijo* “son”, *caballo* “horse” and *libro* “book” (Teschner & Russell, 1984). When new words enter the language, they abide by this regular pattern.

Despite these apparent regularities both masculine and feminine nouns can end in the vowel “a” or “o”, in the vowel “e”, or in a consonant, as shown in Table 1. According to Harris (1991), canonical *-o* masculine- and *-a* feminine-ending nouns form the “inner core”, or most prototypical cases, while non-canonical *-e* and consonant ending nouns form the “outer core”. Masculine nouns ending in *-a* and feminine nouns ending in *-o*, as well as other infrequent exceptional forms are the “residue”. Our study is exclusively concerned with gender agreement in inanimate nouns. We refer to masculine nouns ending in *-o* and feminine nouns ending in *-a* as canonical or transparent. All other endings (*-e*, consonant, opposite vowel) are referred to as non-canonical or non-transparent.

The specific morphological status of word markers is difficult to categorize. Words like *niño/niña* “boy/girl”, *perro/perra* “male dog/female dog”, and *abuelo/abuela* “grandfather/grandmother” have led linguists to treat the terminal elements *-a* and *-o* as actual inflectional morphemes with the meaning [\pm feminine] (Falk, 1978, p. 32). Such a “rule” seems to have psychological validity because native speakers perceive masculine and feminine words ending in the vowels “o” and “a” as regular, as opposed to words that end in the other non-transparent word markers (Frigo & McDonald, 1998). Nonetheless, this generalization fails to capture the lack of direct correspondence between form and meaning with many other words in the Spanish lexicon. In sum, word markers in Spanish are not full-fledged inflectional morphemes

Table 1. *Canonicity of Spanish inanimate nouns based on noun ending.*

	Canonical		Non-canonical	
	-a/-o	-e	Consonant	-a/-o
Masculine	<i>libro</i> “book”	<i>punte</i> “bridge”	<i>motor</i> “engine”	<i>planeta</i> “planet”
	<i>techo</i> “roof”	<i>diente</i> “tooth”	<i>papel</i> “paper”	<i>cometa</i> “comet”
Feminine	<i>mesa</i> “table”	<i>noche</i> “night”	<i>piel</i> “skin”	<i>mano</i> “hand”
	<i>casa</i> “house”	<i>nube</i> “cloud”	<i>nariz</i> “nose”	<i>foto</i> “photo”

like the past tense or plural, but they do share some partial predictability that must somehow be registered in the grammar.

Gender agreement is both a lexical property of nouns and a syntactic operation. All nouns are assigned gender in the lexicon, but must agree in gender and number with determiners and adjectives in noun phrases and in verb phrases. Thus, the masculine noun *techo* “roof” in (1) agrees with the masculine definite determiner *el* and the predicative adjective (participle) *dañado* “damaged”. In (2), the feminine noun *mesa* “table” agrees in gender with the definite determiner *la* and with the adjective *cuadrada* “square”.

- (1) El techo está dañado.
 the.MASC roof[+MASC] is damaged.MASC
 “The roof is damaged.”
- (2) La mesa cuadrada.
 the.FEM table[+FEM] square.FEM
 “The square table.”

Native-speaker knowledge of gender involves having a grammatical representation for the lexically determined gender feature [\pm feminine] (Carroll, 1989; Carstens, 2000) together with the psycholinguistic ability to access said feature and process gender agreement rapidly and efficiently in real time during speech production and comprehension. Native speakers of languages with gender rarely make gender assignment or agreement errors in production and comprehension, as we show in this study as well, suggesting that the linguistic and psycholinguistic mechanisms of gender agreement are intact and work together in mature and stable native grammars. By contrast, gender errors are very common in early stages of L1 acquisition (before age 3;00 in Spanish, see the review in Montrul, 2004b) and extremely prevalent and persistent in adult L2 acquisition, including quite advanced levels of proficiency in Spanish (Franceschina, 2001, 2005; Grüter et al., 2012). Bilingual school-age Spanish–English children also make gender errors in Spanish (Montrul & Potowski, 2007; Mueller Gathercole, 2002). Although gender errors go away in L1 acquisition, they persist in adult L2 acquisition and some cases of early bilingualism.

Non-canonical ending nouns present a particular challenge to language learners: one can only reliably determine the classification of these particular nouns by their endings because they are ambiguous. Instead, classification can only occur by the morpho-phonological form of the other items in the phrase that agree with the noun (i.e., determiners and adjectives, as in examples (1) and (2) above). Several studies of different languages have shown that gender assignment and agreement with non-canonical or non-transparent nouns take longer to learn and to process. Bates, Devescovi, Pizzamiglio, D’Amico and Hernández (1995) and Taraban and Kempe (1999) found slower processing of gender agreement with non-transparent nouns in Italian and Russian native speakers, while Taraban and Roark (1996) found similar difficulties in French native and non-native speakers. The child language errors in Spanish reported by Hernández Pina (1984) occurred with non-canonical ending nouns. In a study of L1 attrition in a Guatemalan adoptee, Montrul (2011b) also reported that the vast majority of errors produced by the adoptee occurred with non-canonical ending nouns.

The canonicity or transparency of the noun ending also poses significant difficulty for L2 learners and early bilinguals with a weaker command of their L1 than their L2, or Spanish heritage speakers as mentioned earlier (Montrul et al., 2008). Alarcón’s (2011) replication of the Montrul et al. (2008) study found the same patterns in written comprehension and oral production: the L2 learners and the heritage speakers were less accurate with non-canonical than canonical ending nouns. In another recent study, Montrul et al. (2013) administered an oral production task to the groups of Spanish heritage speakers and of L2 learners of Spanish tested in the present study. The two groups were more inaccurate with non-canonical ending nouns than with canonical ending nouns. The Spanish native speakers were not affected by the canonicity of the nouns, performing at ceiling across both types.

In this study we use psycholinguistic tasks to assess gender processing with canonical and non-canonical ending nouns in Spanish L2 learners and heritage speakers. In these types of tasks, native-like processing of gender agreement has been operationalized as showing sensitivity to gender cues on the determiner

and to agreement errors during on-line production or comprehension (Bates, Devescovi, Hernández & Pizzamiglio, 1996). For example, upon hearing or seeing phrases with gender errors (or mismatches) like **un casa*, or **un casa blanco* in on-line tasks, native speakers slow down and take longer to respond than if the gender of the noun and the determiner or adjective agree (or match). This sensitivity to gender violations manifested in slower response times is referred to as the gender incongruity or ungrammaticality effect. Other psycholinguistic studies have focused on detecting a facilitating effect of gender in either oral production or noun recognition (Lew-Williams & Fernald, 2007a, b). Native speakers are sensitive to the gender facilitation effect while late bilinguals or L2 learners seem not to be, as we discuss next and before presenting the details of our study.

Recent studies

Gender agreement is very hard to master in L2 acquisition, and the past few years have seen an increasing number of studies on this topic in monolingual and bilingual speakers of diverse languages. Due to space limitations, we present a selective overview of the recent research most relevant to the goals of our study.

Several recent studies have utilized on-line measures to investigate whether L2 learners of Spanish are actually able to process gender agreement as automatically and efficiently as native speakers in real time, and whether they show the same type of sensitivity to gender cues as native speakers (Alarcón, 2009; Grüter et al., 2012; Keating, 2009; Lew-Williams & Fernald, 2010; Sagarra & Herschensohn, 2011). Alarcón (2009), Keating (2009) and Sagarra and Herschensohn (2011) used written comprehension on-line tasks (a computerized matching task, an eye-tracking reading task, and a moving window reading task, respectively) to measure L2 learners' sensitivity to gender agreement violations. Although quite different in design, these studies demonstrate that psycholinguistic sensitivity to gender agreement develops steadily in L2 learners of Spanish whose L1 does not mark gender, and may be detectable at the intermediate or advanced proficiency level depending on the methodology.

Studies that have investigated L2 learners' sensitivity to gender cues during auditory comprehension and oral production have found different results, including those with very advanced speakers. Using the "look while listening" paradigm, Lew-Williams and Fernald (2010) presented an auditory stimulus containing noun phrases with determiners and nouns and asked participants to look at or find one of two pictures of objects on a screen. The experiment included a gender-match condition, where the two words presented auditorily had the same gender (*la casa* "the house.FEM", *la pelota* "the ball.FEM" or *el*

libro "the book.MASC", *el paquete* "the package.MASC"), and a gender-mismatch condition where the two nouns differed in gender (*el libro* "the book.MASC", *la pelota* "the ball.FEM"). Native speakers and young Spanish-speaking children responded faster when the two objects in the screen did not match in gender than when the two objects matched in gender. The interpretation of this response pattern is that there was a facilitation effect of gender: when native speakers heard the phrases, they used the gender information on the determiner to predict the gender of the noun. Lew-Williams and Fernald (2007a, b) found that L2 learners of Spanish showed the opposite pattern. They performed faster on same gender trials than on different gender trials, suggesting that there was no gender facilitation effect (i.e., they were not using the gender information on the determiner to predict the noun), as found for child and adult native speakers.

Grüter et al. (2012) expanded the Lew-Williams and Fernald (2007a, b) study and tested 19 very advanced L2 speakers of Spanish on off-line comprehension, using the same sentence–picture matching task as Montrul et al. (2008), an elicited production task (also based on the task used by Montrul et al., 2008), and a revised version of the eye-tracking during listening experiment developed by Lew-Williams and Fernald. Grüter et al. found that the L2 learners performed at ceiling in the off-line comprehension task, but were only around 80% accurate in the oral production task. The results of these two tasks replicated the findings of Montrul et al. (2008): the L2 learners were better in the written comprehension task than in the oral production task. Grüter et al. also found that the advanced L2 learners in their study did not show the same gender facilitation effects as the native speakers, replicating the results of Lew-Williams and Fernald (2007a, b). The general findings from all these studies suggest that L2 learners are quite accurate on off-line explicit tasks and in on-line tasks with visual stimulus presentation. Written language helps L2 learners in both on-line and off-line tasks (see also Foote, 2011). When it comes to oral production, however, advanced L2 learners are less likely to display native-like performance.

Another study showing a disadvantage for L2 learners in spoken word recognition and oral repetition is Guillelmon and Grosjean's (2001), which investigated whether age of onset of bilingualism (early vs. late) played a role in processing the gender feature in French. Very proficient early and late bilinguals (L2 learners) participated in the experiments. The early bilinguals started using French and English in childhood, before age 13 years (average 5;4), and the late bilinguals started speaking French at an average age of 15;11. The stimuli were noun phrases consisting of a determiner, an adjective, and a noun. Some noun phrases had correct agreement (*le joli bateau* "the.MASC pretty boat.MASC"), while others had incorrect agreement (**la jolie bateau* "the.FEM

pretty boat.MASC”), and others were grammatical and neutral (*leur joli bateau* “their pretty boat”). The neutral condition was the baseline condition, which was included to detect facilitation, when the determiner and the noun match in gender (grammatical), or inhibition, when the determiner and the noun do not match in gender (ungrammatical). Participants were asked to listen to each phrase presented to them over headphones and to repeat the last word of the phrase (the noun) as quickly as possible. The first experiment tested early bilinguals and monolinguals, and found a gender congruency and incongruency effect for both groups. The second experiment compared monolinguals to late bilinguals and found that the late bilinguals (L2 learners) did not show the same gender facilitation or inhibition effects in the gender-congruent and gender-incongruent conditions found in the monolinguals and in the early bilinguals. Guillelmon and Grosjean concluded that early bilinguals became sensitive to gender early in life and use gender cues in perception like monolinguals. The late bilinguals, by contrast, did not have access to gender features after a critical period. Their lack of sensitivity to gender congruency or incongruency is related to a late onset of bilingualism and maturational effects. Because Guillelmon and Grosjean only used a word repetition task, their study left open whether L2 learners would be sensitive to gender agreement violations in more controlled, metalinguistic tasks.

To summarize, although differences between early and late bilinguals (as found by Guillelmon and Grosjean (2001) or as studied by Montrul et al. (2008) for heritage speakers and L2 learners) are possibly related to age of acquisition, this explanation begs the question of why we find task effects in these two groups. In other words, why do L2 learners tend to demonstrate more native-like knowledge of gender in tasks that rely on written language whereas heritage speakers are more native-like in tasks involving spoken language? Our study attempts to provide an answer to this question.

Specific aims and hypotheses

The present study investigates the interaction between age of acquisition and the implicit/explicit nature of linguistic knowledge of gender in two ways: first, by controlling for modality, and secondly by implementing auditory tasks already used with native speakers (Bates et al., 1996) that might prove more efficient in tapping into the participants’ more automatic and implicit knowledge of grammatical gender than off-line written tasks. The tasks utilized are an aural/oral Word Repetition Task (WRT), an aural Gender Monitoring Task (GMT), and an aural Grammaticality Judgment Task (GJT). The tasks employ the same type of experimental materials but vary on degree of explicitness: the GMT and the GJT require subjects to pay attention to gender form directly (by deciding whether a word

is masculine/feminine or grammatical/ungrammatical) or indirectly (by judging whether phrases are grammatical or ungrammatical), and are therefore more explicit, while the WRT taps into more implicit and automatic knowledge and use of gender. Because heritage speakers have more experience with spoken language than with written language, and because they acquired the language at an earlier age, we could predict an overall advantage for heritage speakers in the three tasks. However, if the degree of explicitness of the task (or reliance on metalinguistic knowledge) enhances the performance of the L2 learners (Bowles, 2011; Rebuschat & Williams, 2011) despite differences in age of acquisition between the two groups, the L2 learners should perform more native-like on the GMT and the GJT than on the WRT. Heritage speakers should perform more native-like than the L2 learners on the WRT task in comparison to the other two more explicit tasks.

If canonicity of the noun ending plays a role in gender assignment and agreement during processing, due to these nouns’ infrequency and irregularity in the input, then we expect non-canonical ending nouns to be processed less accurately and slower than canonical ending nouns in general, but particularly by the L2 learners and the heritage speakers since they have received less input and used the language less frequently than native speakers.

Method

Participants

A group of 23 Spanish native speakers, a group of 29 Spanish heritage speakers, and a group of 33 L2 learners of Spanish whose native language was English participated in the three experiments. The native speakers (mean age 30.5) were all born and raised in a Spanish-speaking country. They were all graduate and undergraduate students at the American university where the study took place. Their length of residence in the United States ranged from two months to 10 years (mean 3.4). The heritage speakers and the L2 learners were recruited from advanced Spanish classes at the same university. The L2 learners (mean age 23.5) were all born in the United States to English-speaking parents. They started learning Spanish as a second language predominantly in instructed settings between the ages of 12 and 20, in middle school, high school or college. Their mean age of onset of L2 learning was 15.1 years. As for the heritage speakers, 22 were born in the United States to Spanish-speaking families and began exposure to English before age four. Six heritage speakers were born in Mexico and one in Argentina, and immigrated to the United States before that age. All the heritage speakers were schooled in English in the United States. The mean age of the heritage speakers at time of testing was 21.7. More information on the heritage

Table 2. Information about the heritage speakers and the L2-learner participants.

	Spanish heritage speakers (n = 29)	L2 learners of Spanish (n = 37)
Age	23.6	21.5
Age of acquisition of Spanish native language	Birth	13.2
Language(s) spoken at home in early childhood	Spanish 90%	English 100%
	Spanish/English 10%	
Language of schooling at the elementary level	Spanish 51.7%	English 100%
	Spanish/English 48.3%	
Language of schooling in middle and high school	English 76%	English 100%
	some Spanish 24%	
Level of Spanish classes taking at the university	English 100%	English 100%
	Intermediate and advanced	Advanced
Travelled to Spanish-speaking country	Yes 97% (since childhood) No 3%	Yes 100% (in college)
Length of stay in Spanish-speaking country	2 weeks – 9 months	2 weeks – 6 months

speakers and the L2 learners is displayed in Table 2. Thirteen L2 learners and nine heritage speakers reported beginner or intermediate knowledge of other languages (Chinese, ASL, Japanese), some of them with gender in nouns (Italian, Portuguese, French, German, Polish, Hindi). If knowledge of languages with gender affected the results of the WRT as a reviewer suggests, it affected the two groups similarly given the amount of subjects in each group who had knowledge of another language.

In previous studies of L2 learners and heritage speakers we have used a written test (cloze test and vocabulary task) to assess proficiency in the two groups (e.g., Montrul et al. 2008), but since the present study did not include written language, we opted for a measure of oral proficiency instead. We followed the proficiency/dominance assessment proposed and developed by O'Grady (2009) for Korean heritage speakers because we think it is more informative than speech rate in the weaker language only, as used by Polinsky (2008). Even though the focus of our study was oral production, we also administered a written proficiency test to the two experimental groups, the same test used in Montrul et al. (2008). The maximum score on this test was 50, and the two groups scored in the range of 30–48 (intermediate and advanced). The mean for the heritage speakers was 41.51 ($SD = 4.57$) and the mean for the L2 learners was 38.21 ($SD = 4.57$), which were significantly different on an independent samples t -test ($t(64) = 2.54, p < .013$) because the distribution of scores was slightly different in the two groups. Yet, when we entered written proficiency as a covariate in the statistical analyses of the three main tasks and the oral task reported in Montrul et al. (2013), written proficiency was neither significant nor did it interact with any of the other within-subjects variables.

Participants completed a Picture Naming Task (PNT) that the two experimental groups performed in English and in Spanish separately to establish their degree of language dominance. The same pictures were used in both versions. The Spanish PNT was always administered first, after the background questionnaire and the English PNT was administered last, after all the other experimental tasks. Participants saw 48 black and white images (the same images in the two languages) on a computer screen and were prompted to say the name of the object as quickly as possible. In the Spanish naming task, participants were prompted by the instruction “*Diga*”. Both accuracy and reaction times were measured. Two independent one-way ANOVAs (one with accuracy as dependent variable and one with speed as dependent variable) compared the three groups (native speakers, heritage speakers, L2 learners) in the Spanish PNT. The heritage speakers and the L2 learners were also compared on their speed and accuracy of naming pictures in English, this time with independent samples t -tests since there were only two groups, one t -test for each dependent variable. To assess language dominance within each group, the L2 learners and the heritage speakers were also independently compared on their own performance in the two PNTs (English and Spanish) through paired samples t -tests on accuracy and speed in the two languages. The results are summarized in Table 3.

The Spanish native speakers were significantly faster ($F(2,85) = 12.13, p < .001$) and more accurate ($F(2,85) = 11.12, p < .001$) than both the heritage speakers and the L2 learners in the Spanish Picture Naming Task, according to Tukey's tests ($p < .001$). The heritage speakers and the L2 learners were not significantly different from each other in speed and accuracy on either of the Spanish and English Picture Naming Tasks

Table 3. Mean accuracy and reaction times in the Spanish and English Picture Naming Tasks.

Group	N	Spanish Picture Naming Task		English Picture Naming Task	
		Speed (ms)	Accuracy (%)	Speed (ms)	Accuracy (%)
		Mean SD	Mean SD	Mean SD	Mean SD
Native					
speakers	23	788 (185.3)	95 (2.9)	–	–
Heritage					
speakers	29	1110 (277.8)	88 (7.2)	855 (190.2)	98 (2.4)
L2 learners	37	1227 (450.9)	88 (7.7)	764 (204.6)	99 (1.5)

($p > .05$ for each independent samples t-test for accuracy and speed in English and in Spanish). The fact that the two experimental groups are faster ($F(1,62) = 63.41$, $p < .001$) and more accurate ($F(1,62) = 116.48$, $p < .001$) naming words in English than naming the same words in Spanish suggests that Spanish is their weaker language and English is their stronger language.

Experimental materials

The experimental materials for the Gender Monitoring Task (GMT) in Experiment 1, the Grammaticality Judgment Task (GJT) in Experiment 2, and the Word Repetition Task (WRT) in Experiment 3 consisted of 300 noun phrases with determiners, pronominal adjectives, and target nouns (Det-Adj-N), following a similar procedure for the selection of nouns as in Bates et al. (1996) for Italian and Guillelmon and Grosjean (2001) for French. The 300 noun phrases were created from a set of three determiners (the definite singular masculine determiner *el* “the”, the definite singular feminine determiner *la* “the”, and the gender neutral singular possessive determiner *su* “his/her”), seven adjectives (five transparent, ending in *-a* or *-o* depending on the gender of the noun as in *quinto/quinta* “fifth” and two invariable or opaque regardless of gender, as in *peor* “worst”), and 150 inanimate nouns. All the nouns were disyllabic or trisyllabic with stress on the second syllable, began with a stop consonant and had a word frequency of at least three per million. Word frequency counts were collected from the *Léxico informatizado del español* (LEXESP) database (Sebastián Gallés, Martí, Carreiras & Cuetos, 2000).

All nouns were divided into masculine canonical *-o* like *el cuerpo* “body” (40 nouns), feminine canonical *-a* like *la guerra* “war” (40 nouns), non-canonical masculine and feminine ending in *-e* as in *el puente* “bridge” and *la torre* “tower” (60 nouns), and the remaining 10 were non-canonical nouns ending in *-o* if they were feminine (*la mano* “hand”) and in *-a* if they were masculine (*el tema* “topic”). Thus, feminine nouns ending in *-a*

and masculine nouns ending in *-o* are canonical ending nouns. All other endings (*-e*, consonant, opposite vowel) were non-canonical. Across these four groups, nouns were matched for length in syllables (one-way ANOVA: $F(3,146) = .37$, $p = .778$) and the syllable at which each noun was distinguishable from all other words in Spanish – the uniqueness point ($F(3,146) = 1.56$, $p = .201$). The uniqueness point of a word is the point in the word when its spoken form becomes unique to that word in comparison to all other words in the language (cohort). For example, the uniqueness point of *camino* falls on the last phoneme (/o/), since only at that point can the hearer be certain that he or she is hearing *camino* as opposed to other possibilities such as *camina* or *caminas*. However, due to the difficulty in noun selection given the stimuli constraints, nouns were not matched for frequency ($F(3,146) = 10.19$, $p < .001$). There were significant differences in frequency between the group of masculine non-canonical nouns, on the one hand, and the groups of both feminine and masculine canonical nouns, on the other (all $ps < .001$). We will return to this point in the discussion.

In order to construct the 300 Spanish noun phrases, we followed the experimental set-up of Guillelmon and Grosjean (2001), except that we implemented other changes. A native-speaking female of Mexican Spanish was recorded in a sound-proof studio uttering a variety of noun phrases (at a normal rate), such as *el gran capital* “the big/great capital~money”, *la gran capital* “the big/great capital~city”, and *su gran capital* “his/her big/great capital~money/capital~city”. From these phrases, the five best exemplars of each determiner were chosen by the evaluation of two native-speaker judges, and spliced out. Noun phrases with adjectives and nouns from the target list of items selected for the experiments were recorded individually, and then combined with the best example determiners to create the target set of 300 phrases for each experiment (total 900 experimental items). The 300 items for each task were formed from 150 nouns, each repeated twice (once with a transparent adjective

Table 4. Example of stimuli for all three experiments.

Grammaticality	Noun gender	Adjective transparency	Noun ending	
			Canonical	Noncanonical
Grammatical	Feminine	Opaque	la gran guerra	la gran calle
		Transparent	la quinta guerra	la quinta calle
	Masculine	Opaque	el peor texto	el peor viaje
		Transparent	el quinto texto	el quinto viaje
Ungrammatical	Feminine	Opaque	*el gran guerra	*el gran calle
		Transparent	*el quinto guerra	*el quinto calle
	Masculine	Opaque	*la peor texto	*la peor viaje
		Transparent	*la quinta texto	*la quinta viaje
Neutral ^a	Feminine	Opaque	su gran guerra	su gran calle
		Transparent	su quinta guerra	su quinta calle
	Masculine	Opaque	su peor texto	su peor viaje
		Transparent	su quinto texto	su quinto viaje

^a The neutral condition was not included in the Grammaticality Judgment Task.

and once with an opaque adjective). One third of the noun phrases in the GMT and WRT experiments (100) were grammatical, 100 were ungrammatical, and 100 were neutral. Since there was no neutral condition in the GJT, 150 items were grammatical and 150 were ungrammatical. For each experiment, nouns were randomly assigned to grammatical, neutral, and ungrammatical conditions (GMT and WRT), or grammatical and ungrammatical conditions (GJT) (i.e., *casa* “house” was grammatical in the GJT, randomly, but *viento* “wind” was assigned to the neutral condition in the GJT). Within each experiment, nouns were matched across grammaticality conditions for length in syllables (GMT/WRT: $F(11,138) = .15, p = .999$; GJT: $F(7,142) = .20, p = .985$), and uniqueness point (GMT/WRT: $F(11,138) = 1.14, p = .337$; GJT: $F(7,142) = 1.49, p = .174$), but there were differences in word frequency (GMT/WRT: $F(11,138) = 2.90, p < .01$; GJT: $F(7,142) = 4.78, p < .001$). Because of these differences in word frequency across grammaticality conditions, results from all three experiments were initially analyzed via a model-comparison approach, using mixed logit (accuracy) and linear (reaction times) models. This was done in order to determine whether word frequency interacted with grammaticality condition. According to the model results, although frequency did affect overall accuracy on the GMT and overall reaction times on the GJT, it did not interact with the grammaticality condition on any of the tasks, in either accuracy or RT analyses. Sample stimuli and experimental conditions for the three experiments are displayed in Table 4.

In addition to the variable grammaticality of agreement based on the determiner (grammatical neutral, grammatical, ungrammatical), we examined transparency

of adjective (opaque vs. transparent) and canonicity of noun ending (canonical vs. non-canonical), the latter being a central variable in our study. Recall that studies that have reported successful acquisition of gender by non-native speakers of Spanish typically include nouns with canonical endings only (White et al., 2004). Studies that have included nouns with non-canonical endings have found that gender is highly problematic for L2 learners of Spanish and Spanish heritage speakers (Alarcón 2011; Montrul et al., 2008).

Procedure

Participants completed the three tasks in the order WRT, GMT, and GJT (the more implicit before the two more explicit tasks), in addition to the Spanish and English Picture Naming Tasks, an eye-tracking task, and an elicited oral production task (the latter is reported in Montrul et al., 2013). We describe and present the results of the tasks in the order from more explicit to less explicit: GMT, GJT, and WRT.

Experiment 1: Gender Monitoring Task (GMT)

The participants met individually with a research assistant and completed the Gender Monitoring Task. They were given a set of headphones and asked to listen to a series of three-word phrases (300 total) and to push one of two buttons on the keyboard: one for feminine and one for masculine, depending on the gender of the target noun. Half of the subjects had the feminine button to the right of the masculine button, while the other half had their placement reversed. A solid blue screen was shown on the

computer and after a two-second pause, the first phrase was played. The program calculated the subjects' reaction time of pushing either button from the onset of the target noun. The experimental session was preceded by a 12-item practice session. Both accuracy and reaction times were measured.

Experiment 2: Grammaticality Judgment Task (GJT)

The stimuli for the Grammaticality Judgment Task came from the audio databank previously described. Since phrases with *su* "his/her" (neutral determiner) are always grammatical, these phrases were not included in this experiment. The procedures for Experiment 2 are identical to those in Experiment 1, with the exception of the renaming of the keyboard buttons from feminine and masculine to grammatical and ungrammatical. Participants were asked to listen to a series of three-word phrases (another set of 300) and to push one of the two buttons on the keyboard to indicate whether or not the phrase was grammatical or ungrammatical. The experimental session was preceded by a 12-item trial session. Reaction times and accuracy were both recorded.

Experiment 3: Word Repetition Task (WRT)

Participants were given a set of headphones with a recording microphone attached and were seated in front of a computer. They were asked to listen to a series of Det-Adj-N phrases and to repeat the last word of each phrase as quickly and accurately as possible after they heard it. A blank screen was shown on the computer (which remained blank throughout the entire experiment) and after a 2 second pause, the first noun phrase was played. The program recorded the time to initiate production of the target noun from the onset of the recorded noun. The experimental session was preceded by a 12-item trial session. Both accuracy and reaction times were measured.

The same type of stimuli – Det-Adj-N phrases – were used in the three experiments, and all three experiments relied on timed spoken language recognition. Yet, the three experimental tasks vary in their degree of explicitness. The Gender Monitoring Task (GMT) and the Grammaticality Judgment Task (GJT) favor a more controlled mode of gender processing. The GMT requires participants to focus on gender explicitly upon hearing the noun phrases and to make a metalinguistic judgment, by classifying nouns as masculine or feminine. In the GJT, participants are not asked to focus explicitly on noun gender, but rather are asked to decide whether a Det-Adj-N sequence is grammatical or ungrammatical. Since gender marking is the basis of the ungrammaticality, this task is an indirect way to induce conscious attentive processing to the gender dimension. By comparison, the Word Repetition Task (WRT) is implicit: all participants

have to do is to orally repeat the last word they hear, in this case the noun. The WRT requires no metalinguistic decision and no attention whatsoever to gender or its morphological markers. In all three experiments, a difference in accuracy and reaction times between the grammatical and ungrammatical conditions (faster and more accurate on grammatical than on ungrammatical conditions) would indicate that the participants are sensitive to gender congruency and incongruency or ungrammaticality, as has been demonstrated for native speakers of French (Guillelmon & Grosjean, 2001) and for native speakers of Italian (Bates et al., 1996) in similar tasks. If we are able to replicate these findings with Spanish native speakers, the question is whether L2 learners and heritage speakers will be equally sensitive to gender cues. Guillelmon and Grosjean (2001) found that early bilinguals were sensitive to gender cues in a repetition task while late bilinguals were not. If age of acquisition is a factor in gender representation and processing, the heritage speakers (early bilinguals) should be as sensitive to the gender congruency and incongruency conditions in all tasks as the native speakers, while L2 learners should be less so. But if, as we predict, the explicitness of the tasks matters more than age of acquisition in making comparisons between L2 learners and heritage speakers (Bowles, 2011; Montrul et al., 2008; Montrul, 2011a), then the heritage speakers should have an advantage over the L2 learners only in the WRT, the implicit task. These task-based predictions are summarized in Table 5.

Based on results of previous work showing that L2 learners and heritage speakers have difficulty with gender agreement with non-canonical ending nouns in off-line tasks and in oral production (Alarcón, 2011; Montrul et al., 2008, 2013), we also expect L2 learners and heritage speakers to be more affected by the canonicity of noun endings than the native speakers in these on-line tasks.

Results

Experiment 1: The Gender Monitoring Task (GMT)

Mean accuracy scores and reaction times for the GMT were each submitted to a mixed ANOVA with grammaticality (grammatical, ungrammatical, neutral) and canonicity (canonical, non-canonical) as within-subjects variables, and group (native speakers, heritage speakers, L2 learners) as a between-subjects variable in the by-subjects analysis, and with group as a within-items variable and grammaticality and canonicity as between-items variables in the by-items analysis. There was a main effect of grammaticality ($F_1(2,160) = 37.50, p < .001, \eta_p^2 = .31; F_2(2,145) = 18.70, p < .001, \eta_p^2 = .20$), a main effect for canonicity ($F_1(1,80) = 315.51, p < .001, \eta_p^2 = .79; F_2(1,145) = 54.33,$

Table 5. Predictions on task performance based on the degree of explicitness of each task.

Task	Type of response	Degree of explicitness	Advantages for heritage speakers over L2 learners?
Gender Monitoring Task (GMT)	Decide whether noun is feminine or masculine	Very explicit	No
Grammaticality Judgment Task (GJT)	Decide whether noun phrase is grammatical or ungrammatical	Explicit	No
Repetition Task (RT)	Repeat last word in phrase	Implicit	Yes

$p < .001$, $\eta_p^2 = .27$), and a main effect of group ($F_1(2,80) = 33.94$, $p < .001$, $\eta_p^2 = .45$; $F_2(2,290) = 171.38$, $p < .001$, $\eta_p^2 = .54$). Post-hoc tests with Bonferroni correction revealed that native speakers had significantly higher accuracy scores ($M = 97.2\%$) than the heritage speakers ($M = 82.5\%$) and the L2 learners ($M = 81.6\%$) ($p < .001$), whose scores did not significantly differ from one another. Additionally, there were three two-way interactions. One interaction was between grammaticality and group ($F_1(4,160) = 5.36$, $p < .001$, $\eta_p^2 = .11$; $F_2(4,290) = 14.15$, $p < .001$, $\eta_p^2 = .16$). The interaction revealed that all groups showed an effect of grammaticality favoring higher accuracy in the grammatical and neutral conditions over the ungrammatical condition; this was confirmed statistically with follow-up repeated measures ANOVAs within each group with grammaticality and canonicity as within-subject factors (all groups: $p < .001$). The effect was significantly stronger, however, for the heritage speakers and L2 learners, whose accuracy fell to well below ceiling in the ungrammatical condition ($M = 71.1\%$ and $M = 72.9\%$, respectively). The second two-way interaction, significant only in the by-subjects analysis, was between grammaticality and canonicity ($F_1(2,160) = 30.02$, $p < .001$, $\eta_p^2 = .27$; $F_2(2,145) = 0.59$, $p = .55$, $\eta_p^2 = .008$), and was due to a greater discrepancy in accuracy rates between canonical and non-canonical nouns in the ungrammatical condition (difference = 20.3%) in comparison to the other two grammaticality conditions (difference, grammatical = 14.4%; neutral = 11.9%). The third two-way interaction was between canonicity and group ($F_1(2,80) = 55.49$, $p < .001$, $\eta_p^2 = .58$; $F_2(2,290) = 76.63$, $p < .001$, $\eta_p^2 = .34$). While all three groups showed a canonicity effect favoring higher accuracy in the canonical condition (native speakers $M = 98.8\%$, heritage speakers $M = 91.1\%$, L2 learners $M = 94.8\%$) over the non-canonical condition (native speakers $M = 95.6\%$, heritage speakers $M = 74.0\%$, L2 learners $M = 68.5\%$), with the follow-up repeated measures ANOVA confirming significant differences in all groups (all $ps < .001$), the effect was larger in the heritage-speaker and the L2-learner groups.

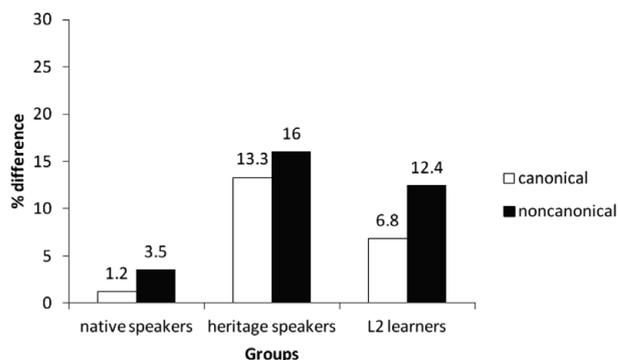


Figure 1. Gender Monitoring Task (GMT): Difference between mean accuracy scores of grammatical and ungrammatical conditions by canonicity.

Figure 1 displays these interactions. For each group, we graphed the difference between the mean percentage accuracy of grammatical phrases minus ungrammatical phrases by canonicity. (We omitted the neutral condition for simplicity and visual clarity, and also because it was not always significant in our results). For example, the accuracy score for native speakers for grammatical phrases with canonical ending nouns was 99.1 and the mean accuracy for ungrammatical phrases with canonical endings was 97.9, a difference of 1.2. The same difference for the non-canonical ending phrases was 3.5. The length of the bars represents the size of the difference between grammatical and ungrammatical phrases, which is a quantitative measure of the magnitude of the grammaticality or congruency effect. Because the bars are on the positive values of the Y-axis, this means that all subjects were more accurate with grammatical than with ungrammatical phrases, and the grammaticality effect is in the right direction for all groups, although stronger in the two experimental groups, and particularly with non-canonical ending nouns.

We now turn to the speed of responses. Only the reaction times for correct responses were analyzed. Before analyses were conducted, reaction times faster than 100 ms or slower than 3000 ms were trimmed to the

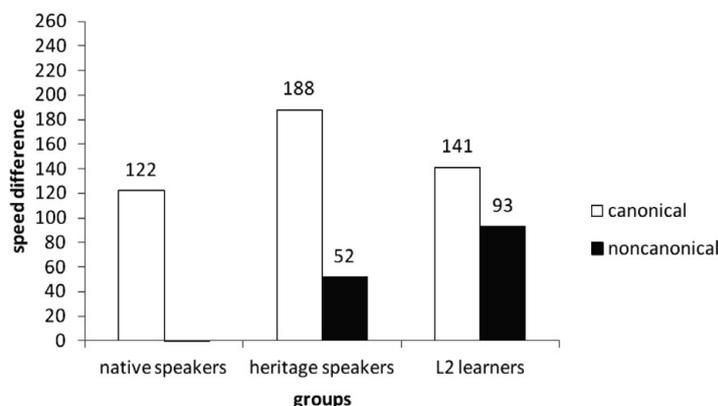


Figure 2. Gender Monitoring Task (GMT): Difference in mean reaction times between ungrammatical and grammatical conditions by canonicity.

corresponding cutoff point. This affected 3.4% of the data. The mixed ANOVA with grammaticality (grammatical, ungrammatical, neutral) and canonicity (canonical, non-canonical) by group (native speakers, heritage speakers, L2 learners) performed on the reaction times revealed main effects of grammaticality ($F_1(2,158) = 46.62, p < .001, \eta_p^2 = .37; F_2(2,144) = 22.69, p < .001, \eta_p^2 = .24$) and of canonicity ($F_1(1,79) = 166.85, p < .001, \eta_p^2 = .67; F_2(1,144) = 140.12, p < .001, \eta_p^2 = .49$), in addition to an effect of group that was marginal in the by-subjects analysis, but significant in the by-items analysis ($F_1(2,79) = 2.70, p = .074, \eta_p^2 = .06; F_2(2,288) = 206.23, p < .001, \eta_p^2 = .58$). All groups differed from each other significantly (all $ps < .001$), with the native speakers showing the fastest reaction times ($M = 1215$), the L2 learners the next fastest ($M = 1329$), and the heritage speakers the slowest ($M = 1412$). In addition to the main effects, there was a group by grammaticality interaction that was significant only in the by-items analysis ($F_1(4,158) = .75, p = .557, \eta_p^2 = .02; F_2(4,288) = 2.63, p < .05, \eta_p^2 = .03$). Follow-up repeated measures ANOVAs within each group with grammaticality and canonicity as within-subjects factors revealed a slightly different reaction time pattern in the heritage speakers than in the other groups: in the native-speaker and the L2-learner groups, reaction times in the grammatical and the neutral conditions did not differ from each other (all $ps = 1.000$), but they both differed from reaction times in the ungrammatical condition (all $ps < .001$). However, in the heritage-speaker group, reaction times in the neutral condition and in the grammatical condition differed marginally from each other in the by-items analysis ($p = .098$), and while reaction times differed in the grammatical and ungrammatical conditions to a similar extent in the heritage speakers and the other two participant groups, reaction times in the neutral condition did not differ as much from reaction times in the ungrammatical condition in the heritage speakers as in the native speakers and the L2 learners (heritage speakers:

difference = 88 ms; native speakers: difference = 112 ms; L2 learners: difference = 118 ms).

In addition to the group by grammaticality interaction, there was also a significant canonicity by grammaticality interaction ($F_1(2,158) = 15.72, p < .001, \eta_p^2 = .16; F_2(2,144) = 3.94, p < .05, \eta_p^2 = .05$). The effect of canonicity decreased in magnitude from the grammatical condition (a difference of 238 ms) to the ungrammatical condition (174 ms) to the neutral condition (124 ms), such that for canonical nouns reaction times were fastest in the grammatical condition ($M = 1154$), followed by the neutral condition ($M = 1227$) and the ungrammatical condition ($M = 1308$), whereas for non-canonical nouns reaction times were fastest in the neutral condition ($M = 1351$), followed by the grammatical condition ($M = 1392$) and the ungrammatical condition ($M = 1482$). The canonicity by group interaction ($F_1(2,79) = 11.35, p < .001, \eta_p^2 = .22; F_2(2,288) = 30.25, p < .001, \eta_p^2 = .17$) indicated that the native speakers and the heritage speakers were equally affected by canonicity in their RTs (difference of about 135 ms between canonical and non-canonical nouns), but the L2 learners were more affected, showing a difference of about 265 ms between canonical and non-canonical ending nouns.

Figure 2 displays the reaction times (RTs) for the GMT. For each group, we graphed the difference in speed (in ms) between ungrammatical phrases (which should be slower) and grammatical phrases (which should be faster) by canonicity. (Here as well we omitted the neutral condition.) For example, the mean RT for native speakers for ungrammatical phrases with canonical ending nouns was 1211 ms and the mean RT for grammatical phrases with canonical endings was 1089, a difference of 122 ms. The same difference for the non-canonical ending phrases was -4 (1324–1328).

Since there was a significant grammaticality effect for all groups in both accuracy and reaction times, the results of the GMT suggest that even when the experimental groups are overall slower and less accurate than the native

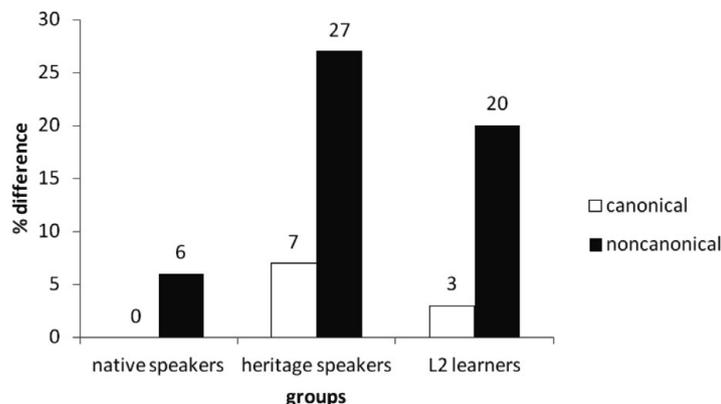


Figure 3. Grammaticality Judgment Task (GJT): Difference between mean accuracy scores of grammatical and ungrammatical conditions by canonicity.

speakers, they are sensitive to the gender congruency effect, like the Spanish native speakers. All groups appear to use gender cues on determiners in noun recognition. Although all groups showed an effect of canonicity in the accuracy results, this effect was larger in the heritage speakers and the L2 learners; they were the only groups that showed an effect in the reaction time data. As predicted, there were no advantages for the heritage speakers over the L2 learners in the GMT.

Experiment 2: The Grammaticality Judgment Task (GJT)

Mean accuracy scores and mean reaction times for each group were submitted to mixed ANOVAs similar to those conducted for the GMT: one for accuracy, one for reaction times. Grammaticality (grammatical, ungrammatical) and canonicity (canonical, non-canonical) were the within-subjects variables, and the between-subjects variable was group (native speakers, heritage speakers, L2 learners). The ANOVA for accuracy revealed main effects for grammaticality ($F_1(1,81) = 62.57, p < .001, \eta_p^2 = .436$; $F_2(1,146) = 58.11, p < .001, \eta_p^2 = .28$) and for canonicity ($F_1(1,81) = 343.31, p < .001, \eta_p^2 = .80$; $F_2(1,146) = 113.01, p < .001, \eta_p^2 = .43$) by which responses on canonical ending nouns ($M = 94.8\%$) were more accurate than on non-canonical ending nouns ($M = 79.7\%$). A main effect of group was also found ($F_1(2,81) = 36.39, p < .001, \eta_p^2 = .47$; $F_2(2,292) = 199.31, p < .001, \eta_p^2 = .57$). Post-hoc tests with Bonferroni correction revealed that native speakers had significantly higher accuracy scores ($M = 96.8\%$) than the heritage speakers ($M = 84.4\%$) ($p < .001$) and the L2 learners ($M = 80.5\%$) ($p < .001$), whose scores did not significantly differ from one another. In addition, there was a three-way interaction between grammaticality, canonicity, and group ($F_1(2,81) = 13.54, p < .001, \eta_p^2 = .25$; $F_2(2,292) = 12.95, p < .001, \eta_p^2 = .08$). When we followed up on the interactions, the Spanish native speakers showed no effect of canonicity, although

they were less accurate in the ungrammatical condition on non-canonical nouns than on canonical nouns (97.9% vs. 92%). By contrast, for the heritage-speaker and L2 groups, the effect of canonicity held for both grammaticality conditions, with an increased magnitude of effect in the ungrammatical condition over the grammatical condition. (Recall that the neutral condition was not part of the design of the GJT.) We present these results in Figure 3, which graphs the difference between accuracy on grammatical and ungrammatical phrases by canonicity, as described for the GMT.

As for latencies in the GJT, only the reaction times for correct responses were analyzed. The same reaction time cutoffs used in the GMT were applied to the GJT results; this affected 5.9% of the data. The ANOVA on reaction times with grammaticality (grammatical, ungrammatical) and canonicity (canonical, non-canonical) as within-subjects factors and group (native speakers, heritage speakers, L2 learners) as a between-subjects factor also revealed a main effect of grammaticality ($F_1(1,80) = 98.01, p < .001, \eta_p^2 = .55$; $F_2(1,146) = 65.62, p < .001, \eta_p^2 = .31$), canonicity ($F_1(1,80) = 195.69, p < .001, \eta_p^2 = .71$; $F_2(1,146) = 85.08, p < .001, \eta_p^2 = .36$), and group ($F_1(2,80) = 27.36, p < .001, \eta_p^2 = .406$; $F_2(2,292) = 1348.29, p < .001, \eta_p^2 = .90$). Post-hoc tests with Bonferroni corrections revealed that reaction times were fastest for the native speakers ($M = 1076$), followed by the heritage speakers ($M = 1414$), and lastly the L2 learners ($M = 1694$) ($p < .01$ between each group).

Parallel to the interaction effect for the accuracy scores, there was a three-way interaction for the reaction times of the GJT between grammaticality, canonicity, and group ($F_1(2,80) = 5.74, p < .01, \eta_p^2 = .12$; $F_2(2,292) = 3.19, p < .05, \eta_p^2 = .02$). The interaction was caused by the fact that the Spanish native speakers showed no effect of canonicity overall, but a difference in the ungrammatical condition (122 ms difference), whereas for the heritage-speaker and L2 groups, the effect of canonicity held for both grammaticality conditions, with an increased

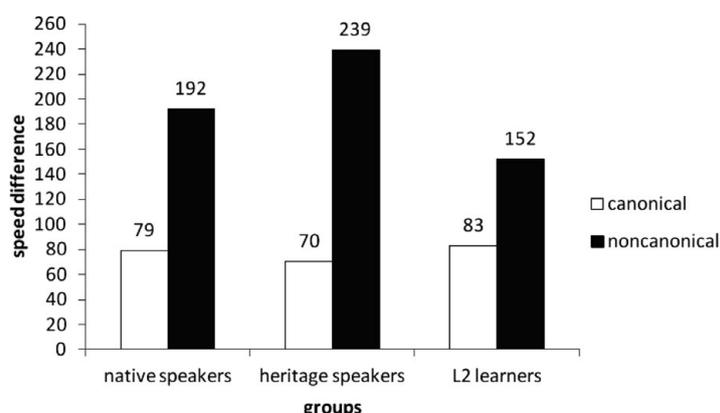


Figure 4. Grammaticality Judgment Task (GJT): Difference in mean reaction times between ungrammatical and grammatical conditions by canonicity.

magnitude of effect in the ungrammatical condition (239 ms difference for the L2 learners and 203 ms for the heritage speakers) over the grammatical condition (170 ms difference, 34 ms for the heritage speakers).

Figure 4 plots the difference in RTs between ungrammatical and grammatical phrases by canonicity.

To sum up, the results of the GJT are very similar to the results of the GMT. All groups are sensitive to the gender grammaticality effect. Yet, while heritage speakers and L2 learners are affected by the canonicity of noun endings, the native speakers are not affected to the same extent. As predicted, in terms of processing gender in a native-like way, there were no advantages for the heritage speakers over the L2 learners in the GJT.

Experiment 3: The Word Repetition Task (WRT)

Accuracy scores on the WRT were at 100% across the three groups and therefore were not subjected to further statistical analyses. After trimming the reaction time data in the same way as was done with the GMT and GJT data and removing button box errors (1.6% of the data were affected), reaction times for the WRT were entered into a mixed ANOVA of the same design as the ANOVAs conducted on the GMT results. The two within-subjects factors were grammaticality (grammatical, ungrammatical, neutral) and canonicity (canonical, non-canonical). The between-subjects factor was group (native speakers, heritage speakers, L2 learners). The analysis revealed a main effect of grammaticality that was significant only in the by-subjects analysis ($F_1(2,164) = 17.97, p < .001, \eta_p^2 = .18$; $F_2(2,143) = 1.70, p = .186, \eta_p^2 = .02$) and an effect of canonicity, again only significant in the by-subjects analysis ($F_1(1,82) = 10.47, p = .001, \eta_p^2 = .11$; $F_2(1,143) = 1.25, p = .266, \eta_p^2 = .01$). Additionally, there was a three-way interaction between grammaticality, canonicity and group that was significant only in the by-subjects analysis ($F_1(4,164) = 2.56, p < .05, \eta_p^2 =$

.06; $F_2(4,286) = 1.74, p = .141, \eta_p^2 = .02$). The same effect revealed by this three-way interaction in the by-subjects analysis was shown in the by-items analysis via two, two-way interactions: one between group and grammaticality ($F_2(4,286) = 3.61, p < .01, \eta_p^2 = .04$) and one between group and canonicity ($F_2(2,286) = 8.44, p < .001, \eta_p^2 = .05$). The interaction revealed that for the native speakers and the heritage speakers, reaction times in the ungrammatical condition were significantly slower than reaction times in the neutral and the grammatical conditions, whereas this grammaticality effect did not occur in the L2-learner group; the L2 learners had the same mean reaction time to both the ungrammatical and the grammatical conditions ($M = 800$). Furthermore, the grammaticality effect found in the native-speaker and heritage-speaker groups was limited to canonical nouns only; non-canonical nouns did not show a grammaticality effect in any group.

To follow up on this result, we conducted three independent ANOVAs, one for each group, with grammaticality and canonicity as the within-subjects factors. The analysis for the native speakers revealed a main effect of grammaticality ($F(1,22) = 28.11, p < .0001, \eta_p^2 = .56$) and a grammaticality by canonicity interaction ($F(1,22) = 25.81, p < .0001, \eta_p^2 = .54$). The native speakers repeated canonical nouns faster in grammatical ($M = 726$) than in ungrammatical phrases ($M = 792$) ($t(23) = 7.20, p < .0001$), while the difference in speed of repetition for non-canonical nouns was not significant in grammatical ($M = 749$) and ungrammatical phrases ($M = 757$). The ANOVA conducted on the heritage speakers showed the same profile: a main effect of grammaticality ($F(1,28) = 15.95, p < .0001, \eta_p^2 = .36$) and a grammaticality by canonicity interaction ($F(1,28) = 11.30, p = .002, \eta_p^2 = .23$). The heritage speakers repeated canonical nouns faster in grammatical ($M = 838$) than in ungrammatical phrases ($M = 876$) ($t(23) = 3.5, p < .001$), while the speed of word repetition for non-canonical nouns was not significantly different between grammatical

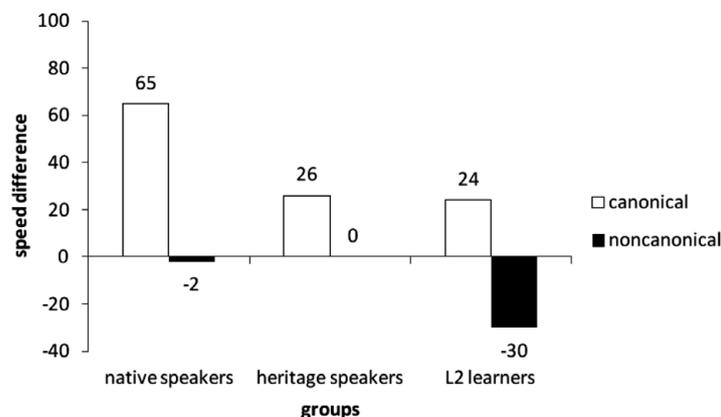


Figure 5. Word Repetition Task (WRT): Difference in mean reaction times between ungrammatical and grammatical conditions by canonicity.

($M = 832$) and ungrammatical phrases ($M = 838$). The ANOVA for the L2 learners revealed no effect for grammaticality ($F(1,33) = .06, p = .94, \eta_p^2 = .23$). There was a significant canonicity by grammaticality interaction for this group as well ($F(1,33) = 25.81, p < .0001, \eta_p^2 = .26$). The canonical nouns showed the expected pattern, but non-canonical nouns show a pattern that is the opposite of what was expected: the L2 learners repeated non-canonical nouns in ungrammatical phrases faster ($M = 795$) than non-canonical nouns in grammatical phrases ($M = 825$) ($t(33) = 9.44, p = .004$), while canonical nouns in grammatical phrases were repeated faster ($M = 775$) than in ungrammatical phrases ($M = 804$) ($t(33) = 4.16, p < .0001$).

Figure 5 plots the difference in RTs between ungrammatical phrases and grammatical phrases by canonicity.

In summary, although the three groups showed a canonicity by grammaticality interaction, the patterns and directions of the interactions were different. The native speakers and the heritage speakers showed a grammaticality effect overall and with canonical nouns. Grammaticality did not have an effect for non-canonical nouns in these two groups. The L2 learners patterned in the right direction with canonical nouns, but showed a grammaticality effect in the opposite direction with non-canonical nouns, showing no overall effect of grammaticality unlike the other two groups. This result is consistent with our hypothesis that, when it comes to grammatical knowledge and native-like processing, heritage speakers appear to exhibit more native-like patterns than L2 learners in less explicit tasks.

Discussion

Unlike native speakers of Spanish who do not usually make errors, heritage speakers and L2 learners make errors with gender agreement in noun phrases. Even though heritage speakers and L2 learners differ from

Spanish native speakers in their level of proficiency and ultimate attainment in the language, one of the goals of the present study was to investigate whether age of acquisition and early language experience provides advantages (i.e., more target-like performance) to Spanish heritage speakers over L2 learners with gender processing, since they were exposed to the language in early childhood, when gender agreement is acquired and mastered by Spanish-speaking children. Previous studies comparing these two groups on off-line tasks found task effects by which L2 learners outperformed heritage speakers in written tasks while heritage speakers outperformed L2 learners in less metalinguistic tasks and, particularly, in oral production (Alarcón, 2011; Bowles, 2011; Montrul et al., 2008, 2013). To minimize the effects of written language and potential reliance on metalinguistic ability, the present study implemented a different methodology. We used three on-line spoken word recognition tasks with systematic variation on the type of response required of each task to tap knowledge and processing of gender more or less implicitly. Similar tasks have detected sensitivity to gender violations in native speakers of French (Guillelmon & Grosjean, 2001) and of Italian (Bates et al., 1995, 1996). Guillelmon and Grosjean found that late bilinguals (L2 learners of French whose L1 was English) were not sensitive to gender violations in noun phrases in a word repetition task, while early bilinguals and native speakers were. Their study, however, did not address whether L2 learners would be equally insensitive to gender agreement violations in more metalinguistic tasks. In the present study we pursued this question by including a gender monitoring task and a grammaticality judgment task.

Although the three word recognition tasks employed in this study did not rely on written language, they still varied in their degree of explicitness due to the type of response required by the participants. The GMT and the GJT required participants to decide whether the nouns were feminine or masculine or whether the phrases

were grammatical or ungrammatical, thus prompting participants to focus on the form of the phrases more explicitly. The WRT only required participants to repeat the last word they heard. In principle, if heritage speakers are better in oral production and comprehension than L2 learners because they were predominantly exposed to spoken language since early childhood, then they should perform better than the L2 learners in all these tasks. But if the degree of explicitness of the task or reliance on some sort of metalinguistic knowledge helps the L2 learners, we predicted that the advantage for heritage speakers over L2 learners would be measurable in the WRT, the most implicit task, but it would not surface necessarily in the GMT and GJT, the more explicit tasks. This is because instructed L2 learners in general have more experience than heritage speakers with metalinguistic tasks typically used in classroom instruction.

The second objective of the study was to investigate whether the canonicity of noun endings would affect gender violation effects in the participants. Studies of gender with native speakers have shown that irregular ending nouns take longer to process (Bates et al., 1995; Taraban & Kempe, 1999; Taraban & Roark, 1996), and several studies comparing L2 learners and heritage speakers have shown that both types of bilinguals are more inaccurate at producing gender agreement with non-canonical ending nouns than with canonical nouns (Alarcón, 2011; Montrul et al., 2008, 2013).

Even though we used a different research design including different adjectives and words with canonical and non-canonical endings, we found that, like previous studies with Italian (Bates et al., 1996) and French native speakers (Guillelmon & Grosjean, 2001), our Spanish native speakers performed largely at ceiling and were sensitive to gender violation effects in the three tasks: that is, they responded faster and more accurately on grammatical than on ungrammatical phrases. The canonicity of the noun ending or the type of adjective did not affect their performance to the extent that it affected the performance of the other two groups.

The heritage speakers and the L2 learners were slower and less accurate than the native speakers on all tasks, and since Spanish is their weaker language, this is not a surprising result. Consistent with our first two hypotheses outlined in Table 5, there were no differences between the heritage speakers and the L2 learners in the GMT and the GJT: that is, there were no advantages for heritage speakers. Like the native speakers, the two groups showed sensitivity to gender violations. However, the two groups were significantly more affected by the type of noun ending than the native speakers: they were more accurate and faster in the GMT and GJT when the nouns had canonical masculine *-o* and feminine *-a* endings than when the nouns ended in non-canonical word markers, confirming the results of previous studies

with off-line tasks (Alarcón, 2011; Montrul et al., 2008, 2013). Since non-canonical ending nouns are more infrequent than canonical ending nouns (recall that it was not possible to match nouns in frequency due to their endings), this result can also be due to frequency. Gollan, Slattery, Goldenberg, van Assche, Duyck and Rayner (2011) proposed the frequency-lag hypothesis to explain that the bilingual disadvantage in language processing and lexical retrieval centers primarily on low frequency words in reading and in speaking. The results of the L2 learners and the heritage speakers in the GMT and the GJT are consistent with this hypothesis, although our tasks involved auditory comprehension. At the same time, another experiment would have to be carried out to tease apart the independent effects of frequency and canonicity with these groups. Crucially, however, the heritage speakers had advantages over the L2 learners in the WRT, in line with our predictions. In this experiment, all groups exhibited a gender incongruity or grammaticality effect with canonical ending nouns, yet differed on sensitivity to gender incongruity with non-canonical ending nouns. The native speakers and the heritage speakers repeated non-canonical ending words in grammatical and ungrammatical phrases equally fast (no effect), but the L2 learners repeated non-canonical ending nouns in ungrammatical phrases faster than words in grammatical phrases, an effect in the opposite direction. One reason for the opposite-than-expected pattern could be that some of these nouns may have been classified incorrectly for these speakers, so that the ones that are feminine for the L2 learners are really masculine. If they had performed native-like on the items to which they assigned gender correctly this would suggest that L2 learners are sensitive to the gender feature and that the issue with gender is largely lexical. This explanation is actually supported by the findings of Grüter et al. (2012) and Montrul et al. (2013), where the same individuals performed a picture description task. The L2 learners in the Montrul et al. (2013) study also produced significantly more gender errors with non-canonical ending nouns than the heritage speakers, and the vast majority of the errors were lexical misclassifications of gender.

Our overall findings are similar to what Guillelmon and Grosjean (2001) found in French because the early bilinguals in their study showed overall sensitivity to grammaticality in the WRT like the native speakers while the late bilinguals did not, although Guillelmon and Grosjean did not test non-canonical ending nouns as we did in our study. If implicit knowledge of gender applies to both canonical and non-canonical ending nouns, the lack of effect of overall grammaticality for L2 learners in the WRT suggests that L2 learners may lack the same implicit knowledge of gender that both native monolinguals and heritage speakers may access when completing oral, less metalinguistic tasks, as Guillelmon and Grosjean (2001)

suggested. At the same time, the fact that the L2 learners showed sensitivity to grammaticality in the WRT with canonical ending nouns like the other two groups weakens the claim that L2 learners may not have the same type of implicit representation of gender as native speakers and heritage speakers, unless it is assumed that canonical and non-canonical ending nouns are handled by entirely different mechanisms. What is significant, however, is that the L2 learners' pattern of response with non-canonical nouns was very different from that of the native speakers and the heritage speakers in this task. The native speakers and the heritage speakers repeated non-canonical nouns faster than canonical nouns, while the L2 learners were slower with non-canonical nouns than with canonical nouns. According to Gollan et al.'s (2011) frequency-lag hypothesis, word frequency affects bilinguals' production in their non-dominant language. If word repetition is a form of production, our results suggest that the relative infrequency of non-canonical ending might have affected the L2 learners, but it did not affect the native speakers and the heritage speakers to the same extent, suggesting that when it comes to activating implicit knowledge of gender in production, native speakers and heritage speakers are very similar to each other with both canonical and non-canonical ending nouns. Canonicity of nouns seems to affect L2 speakers and heritage speakers differently in word repetition.

Although L2 learners may have acquired that Spanish nouns have a gender feature, it does not seem to be integrated and processed in the same way as in native speakers and heritage speakers during oral repetition and production. In other words, L2 learners with a late onset of acquisition do not seem to use the gender feature as efficiently during word recognition as the heritage speakers and the native speakers, replicating the finding of Guillelmon and Grosjean (2011) with L2 learners of French. The results of the WRT reported in the present study are also consistent with the results of the oral production task with the same participants reported in Montrul et al. (2013). The combined results of these two studies show that producing gender agreement in Spanish is more difficult for the L2 learners than for the heritage speakers.

Despite being slower and less accurate, the heritage speakers were sensitive to the grammaticality effect and generally detected gender violations or showed a null effect in all three tasks and with the two types of nouns (canonical and non-canonical) like the native speakers. The L2 learners only displayed sensitivity to gender marking errors in the two more explicit tasks and showed the opposite pattern of responses with non-canonical ending nouns in the more implicit task. Our findings are consistent with other recent studies showing that metalinguistic tasks improve linguistic performance in L2 learners (Bowles, 2011; Rebuschat & Williams, 2011),

and that L2 learners tend to display native-like knowledge of gender with canonical ending nouns (Montrul et al., 2008; White et al., 2004). At the same time, the native speakers were not affected by the ending of the noun to the same extent as the two experimental groups, for whom canonicity of the noun was highly significant in both accuracy and reaction times.

We believe that the fact that the heritage speakers patterned with the native speakers in the WRT (the more implicit task) and that the L2 learners did not differ from the heritage speakers in the more explicit tasks (GMT and GJT) may be related to differences in learning experience rather than merely age of acquisition. Because heritage speakers are born in a home environment where the heritage language is spoken, they are exposed to the language since birth and in early childhood in a naturalistic setting. The input they receive in the heritage language at that age is primarily through the auditory medium, and they use spoken language in social interactions with their caregivers. However, most heritage speakers receive limited to no schooling in their heritage language. By contrast, L2 learners start acquisition of the second language around or after puberty in a formal setting (classroom) or in a naturalistic environment. Although they have access to spoken language and receive auditory input, a great deal of input is actually written. Unlike heritage speakers who can be illiterate in the heritage language, L2 learners are fairly literate in their second language, exposed to both visual and aural input in the classroom. Thus, it is possible that in addition to age of acquisition (timing of input), modality of input, and experience with more or less spoken or written language may play a role in linguistic knowledge and input processing experience and strategies.

For example, we know that gender is in the lexicon. When children learn Spanish or any other language with gender, they hear sequences of determiners and nouns in the acoustic input and must identify nouns in the speech stream (through computations or transitional probabilities). In fact, very young monolingual and bilingual children produce their first nouns with a preverbal vowel (*e pie* 'the foot', *a queca* 'a doll', *u fo* 'a flower'; López Ornat, 1997), a protodeterminer according to Lleó (1998), which coincides with the vowels of gender-marked definite and indefinite determiners (*el, la, un, una*). These early productions may constitute unanalyzed chunks and suggest that there is a very tight association between determiners and nouns in the lexicon. With more input and experience, the child later segments the chunk into determiner and noun. Most recent experimental evidence by Lew-Williams and Fernald (2007a, b, 2010) suggests that noun-gender associations are strong in the L1 lexicon as a consequence of early speech segmentation. In their studies of the visual world paradigm they found that adult native Spanish speakers and three-four-year-old

Spanish-speaking children use gender information in determiners to predict nouns during spoken word recognition.

But second language acquisition around puberty is different. L2 learners at this age are exposed to visual input through reading and writing, in addition to aural input. They already know through their L1 that determiners and nouns are separate words. Visual input reinforces this idea because there are spaces between words. Because visual input gives information about word boundaries, L2 learners may not need to rely as much on distributional properties and transitional probabilities to segment the acoustic stream. As a result, the association between noun and determiners and noun and gender in the lexicon may not be very strong in the L2 (see also Grüter et al., 2012, for a similar explanation). It appears, then, that input modality affects language representation and processing, and may explain why L2 learners are typically less sensitive to gender marking than native speakers in implicit tasks.

Although heritage speakers are child learners, they make errors like L2 learners; their noun–gender lexical associations may be stronger than in L2 learners but weaker than in mature native speakers. The heritage speakers in our study showed the same pattern of responses and sensitivity to gender as our native speakers, except that they are quantitatively different (less accurate and slower). It is likely that reduced input and use of the minority language throughout the school-age period leads to reduced frequency of use of nouns and their associated genders by heritage speakers as they grow older. Gollan, Montoya, Cera and Sandoval (2008) proposed the WEAKER LINKS HYPOTHESIS to explain potential speed and accuracy differences between monolinguals and bilinguals in lexical access. Extending this hypothesis to the specific case of gender processing in heritage speakers, we can assume that noun–gender links may have been stronger in their childhood but they may have also progressively weakened as the first language became the secondary language. Weaker links due to reduced frequency of use lead to gender assignment errors, slower retrieval of nouns in the lexicon, slower insertion of nouns in the syntax, and slower speed at computing syntactic dependencies (concord with determiners, nouns, and adjectives), and all of these lead to gender agreement errors in Spanish heritage speakers.

Of course, problems with gender can be lexical, at the level of lexical assignment, or syntactic, by failing to perform concord among all the elements of the noun phrase. Given the nature of the tasks used in this study, where the most important cue to gender came from the determiner, we can only support an explanation related to problems of lexical assignment. In order to tease apart whether L2 learners and heritage speakers make lexical and/or syntactic errors, a different methodology involving production of determiners and

adjectives transparently marked for gender is more suitable to confirm these alternative possibilities, as used by Montrul et al. (2013). What remains to be explained is why canonicity of noun ending affects L2 learners and heritage speakers to such an extent. We have seen that L2 learners and heritage speakers are more accurate in tasks that use canonical ending nouns, and can even display at-ceiling performance, than when the tasks also use non-canonical ending nouns. Although gender is assigned in the lexicon, it does have an overt morphological expression in Spanish nouns, through the word markers *-a*, *-o*, *-e*, consonant (Harris, 1991). Feminine *-a* and masculine *-o* are regular, the rest are irregular, and L2 learners and heritage speakers use these phonological and morphological cues when assigning gender to nouns. According to the dual mechanism model of inflection (Pinker, 1999; Pinker & Prince, 1994; Pinker & Ullman, 2002), regular morphological processes are stored in procedural memory and irregularities are stored in declarative memory. Extending this approach to gender marking, once canonical ending nouns are learned, the gender assignment is automatized, stored in procedural memory and handled by rule (implicitly acquired in childhood by heritage speakers and learned later but automatized through practice in L2 learners). Non-canonical ending nouns, by contrast, need to be memorized. They are also more infrequent and, in our study, we were unable to completely match canonical and non-canonical ending nouns on frequency. We suggest that reduced input and use of Spanish by L2 learners and heritage speakers may affect storage in declarative memory. When this happens, L2 learners and heritage speakers resort to the regular rule, but regularity is less predictable for nouns ending in *-e* and in a consonant. Mature native speakers whose primary language is Spanish do not exhibit gaps with declarative memory because they use the language more frequently on a daily basis and the lexical association links remain strong for both canonical and non-canonical ending nouns (Gollan et al., 2008). This idea predicts that non-canonical ending nouns will be highly affected under L1 attrition. In fact, Montrul's (2011b) study of an adult Guatemalan adoptee showed that the vast majority of gender errors produced by the subject of the case study were precisely with non-canonical ending nouns.

To conclude, our study suggests that although both L2 learners and heritage speakers make gender agreement and assignment errors when compared to native speakers who have full command of the language and use it frequently, heritage speakers display more native-like patterns than L2 learners in implicit tasks that require aural comprehension, like the WRT used in this study. Due to differences in language learning experience, L2 learners are able to develop sensitivity to gender marking, but this knowledge is better manifested in visual and

auditory tasks that tap some metalinguistic component of their knowledge. The implication from these findings is that the research methodology and the types of tasks matter when it comes to reaching conclusions about the linguistic behavior and possible mental representations of second language learners and early bilinguals who differ in their language learning experience, and in theory construction more generally. Types of task and modality (aural, visual) need to be taken into account when comparing different bilinguals, especially in order to understand their linguistic knowledge and processing and to draw implications for the classroom.

References

- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*, 249–306.
- Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data. *Hispania, 92*, 814–828.
- 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*, 332–350.
- Albirini, A., Benmamoun, E., & Saddah, E. (2011). Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. *Studies in Second Language Acquisition, 33*, 273–304.
- Au, T., Knightly, L., Jun, S., & Oh, J. (2002). Overhearing a language during childhood. *Psychological Science, 13*, 238–243.
- Bates, E., Devescovi, A., Hernández, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics, 58*, 992–1004.
- Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernández, A. (1995). Gender priming in Italian. *Perception & Psychophysics, 57*, 847–862.
- Bialystok, E., & Ryan, E.B. (1985). A metacognitive framework for the development of first and second language skills. In D. L. Forrest-Pressley, G. E. Mackinnon & T. G. Waller (eds.), *Meta-cognition, cognition, and human performance*, pp. 207–252. New York: Academic Press.
- Bley-Vroman, R. (2009). The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition, 31*, 175–198.
- Bolonyai, A. (2007). (In)vulnerable agreement in incomplete bilingual L1 learners. *The International Journal of Bilingualism, 11*, 3–21.
- Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition, 33*, 247–272.
- Bruhn de Garavito, J. (2002). Verb raising in Spanish: A comparison of early and late bilinguals. In B. Skarabela, S. Fish & A. H.-J. Do (eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development*, pp. 84–94. Somerville, MA: Cascadilla Press.
- Carroll, S. (1989). Second language acquisition and the computational paradigm. *Language Learning, 39*, 535–594.
- Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry, 31*, 319–355.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499–534.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.
- Falk, J. (1978). *Linguistics and language: A survey of basic concepts and implications* (2nd edn.). New York: Wiley.
- Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics, 32*, 187–220.
- Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research, 17*, 213–247.
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: John Benjamins.
- Frigo, L., & McDonald, J. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218–245.
- Gollan, T. H., Montoya, R. I., Cera, C. M., & Sandoval, T. C. (2008). More use almost always means smaller a frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140*, 186–209.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in the L2: A production or a real time processing problem? *Second Language Research, 28*, 191–215.
- Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition, 29*, 503–511.
- Håkansson, G. (1995). Syntax and morphology in language attrition: A study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics, 5*, 153–171.
- Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry, 22*, 27–62.
- Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: The “failed functional features hypothesis”. *Second Language Research, 13*, 187–226.
- Hernández Pina, F. (1984). *Teorías psicosociolingüísticas y su aplicación a la adquisición del español como lengua materna* [Psychosociolinguistic theories and their application to the acquisition of Spanish as a mother tongue]. Madrid: Siglo XXI.

- Hulsen, M. (2000). Language loss and language processing: Three generations of Dutch migrants in New Zealand. Ph.D. dissertation, University of Nijmegen.
- Keating, G. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59, 503–535.
- Lew-Williams, C., & Fernald, A. (2007a). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 33, 193–198.
- Lew-Williams, C., & Fernald, A. (2007b). How first and second language learners use predictive cues in on-line sentence interpretation in Spanish and English. In H. Caunt-Nulton, S. Kulatilake & I. Woo (eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development*, pp. 382–393. Somerville, MA: Cascadilla Press.
- 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.
- Lleó, C. (1998). Proto-articles in the acquisition of Spanish: Interface between phonology and syntax. In R. Fabri, A. Ortman & T. Parodi (eds.), *Models of inflection*, pp. 175–195. Tübingen: Max Niemeyer Verlag.
- Long, M. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.
- López Ornat, S. (1997). What lies in between a pre-grammatical and a grammatical representation? Evidence on nominal and verbal form–function mapping in Spanish from 1;7 to 2;1. In A. T. Pérez-Leroux & W. Glass (eds.), *Contemporary perspectives on the acquisition of Spanish*, pp. 3–20. Somerville, MA: Cascadilla Press.
- Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, 5, 39–68.
- Montrul, S. (2004a). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition*, 7, 125–142.
- Montrul, S. (2004b). *The acquisition of Spanish*. Amsterdam: John Benjamins.
- Montrul, S. (2007). Interpreting Mood distinctions in Spanish as a heritage language. In K. Potowski & R. Cameron (eds.), *Spanish contact: Policy, social and linguistic inquiries*, pp. 23–40. Amsterdam: John Benjamins.
- Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.
- Montrul, S. (2009). Incomplete acquisition of Tense–Aspect and Mood in Spanish heritage speakers. *The International Journal of Bilingualism*, 13, 239–269.
- Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30, 3–23.
- Montrul, S. (2011a). Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition*, 33, 155–161.
- Montrul, S. (2011b). First language retention and attrition in an adult Guatemalan adoptee. *Language, Interaction and Acquisition*, 2, 276–311.
- Montrul, S. (in press). How native are heritage speakers? A look at gender agreement in Spanish. *The Heritage Language Journal*.
- Montrul, S., & Bowles, M. (2009). Back to basics: Differential Object Marking under incomplete acquisition in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 12, 363–383.
- Montrul, S., de la Fuente, I., Davidson, J., & 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, 87–118.
- Montrul, S., & Foote, R. Age of acquisition interactions in bilingual lexical access: A study of the weaker language in L2 learners and heritage speakers. *The International Journal of Bilingualism*, 16, doi:10.1177/1367006912443431. Published online by Sage, May 8, 2012.
- Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58, 3–53.
- Montrul, S., & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition*, 13, 449–473.
- Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish bilingual children. *International Journal of Bilingualism*, 11, 301–328.
- Mueller Gathercole, V. C. (2002). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In D. K. Oller & R. E. Eilers (eds.), *Language and literacy in bilingual children*, pp. 207–219. Clevedon: Multilingual Matters.
- O’Grady, W. (2009). Assessing heritage language competence. Presented at the Third Heritage Language Institute, University of Illinois at Urbana–Champaign, June 2009.
- O’Grady, W., Kwak, H-K., Lee, M., & Lee, O-S. (2011). An Emergentist perspective on partial language acquisition. *Studies in Second Language Acquisition*, 33, 223–246.
- O’Grady, W., Lee, M., & Choo, M. (2001). The acquisition of relative clauses by heritage and non-heritage learners of Korean as a second language: A comparative study. *Journal of Korean Language Education*, 12, 283–294.
- Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 83, 770–802.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: HarperCollins.
- Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima, R. L. Corrigan & G. K. Iverson (eds.), *The reality of linguistic rules*, pp. 321–352. Philadelphia, PA: John Benjamins.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456–463.
- Polinsky, M. (1997). American Russian: Language loss meets language acquisition. In W. Browne (ed.), *Formal approaches to Slavic linguistics*, pp. 370–407. Ann Arbor, MI: Michigan Slavic Publications.

- Polinsky, M. (2008). Russian gender under incomplete acquisition. *The Heritage Language Journal*, 6, 40–71.
- Polinsky, M. (2011). Reanalysis in adult heritage language: A case for attrition. *Studies in Second Language Acquisition*, 33, 305–328.
- Rebuschat, P., & Williams, J. (2011). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33, 1–28.
- Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in heritage Brazilian Portuguese. *The International Journal of Bilingualism*, 11, 359–389.
- Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processing during comprehension. *Language Learning*, 61, 98–116.
- Sebastián Gallés, N., Martí, M. A., Carreiras, M. F., & Cuetos, F. (2000). LEXESP: Léxico informatizado del español. CD-ROM. Barcelona: Edicions de la Universitat de Barcelona.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Taraban, R., & Roark, B. (1996). Competition in language-based categories. *Applied Psycholinguistics*, 17, 125–148.
- Taraban, R., & Kempe, V. (1999). Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics*, 20, 119–148.
- Teschner, R., & Russell, W. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, 9, 157–173.
- Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Language Research*, 23, 215–242.
- White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y.-K. I. (2004). Gender agreement in nonnative Spanish: Evidence against failed features. *Applied Psycholinguistics*, 25, 105–133.