Early language experience facilitates gender processing in Spanish heritage speakers

Justin Davidson, Israel de la Fuente, Silvina Montrul, Rebecca Foote

1. Introduction

In recent years, there has been a growing body of research comparing the linguistic abilities of adult second language learners and heritage speakers (e.g. Au et al., 2002; Montrul, 2010; Montrul et al., 2008). Heritage speakers are bilinguals whose first language is a minority or heritage language. Their heritage language is typically sociolinguistically and functionally less dominant than the majority language acquired simultaneously or as a second language. In the context of the United States, the majority language is English; the heritage language is an immigrant language. As adults, many heritage speakers come to the second language classroom to learn, relearn or improve their proficiency level in the family language as a way to reconnect with or preserve their cultural background. What second language learners and heritage speakers have in common is that they display comparable levels of nonnative-like ability in the secondary and less dominant target language, appearing to make very similar errors. From a pedagogical perspective, it is important to understand what type of linguistic knowledge heritage speakers already bring to the classroom and how their learning is different or not from that of second language learners.

At the same time, the comparison between second language learners and heritage speakers is also motivated by theoretical issues. Do heritage speakers have advantages over second language learners with early acquired aspects of language? Since heritage speakers acquired the heritage language early in life in the home environment and received aural input, whereas second language learners start acquisition of the second language well after the foundations of their first language are in place, typically in an instructed environment through aural and visual input, their acquisition and processing of the language may be different. If late language learning is the main reason behind nonnative-like linguistic knowledge in second language learners (Bley-Vroman, 1990; Clahsen & Muysken, 1986; DeKeyser, 2000;

[-] Justin Davidson, University of Illinois at Urbana-Champaign, jdavids2@illinois.edu; Israel de la Fuente, University of Illinois at Urbana-Champaign, delafue1@illinois.edu; Silvina Montrul, University of Illinois at Urbana-Champaign, montrul@illinois.edu; Rebecca Foote, University of Illinois at Urbana-Champaign, rfoote@illinois.edu. We are grateful to our undergraduate research assistants, Celeste Larkin, Gretchen Shaw, Kayla Pennoyer, Rachel Pirovano and Adam Bethune. Thank you to the audience of *BUCLD 35* for their comments and suggestions. All remaining errors are our own.

Long, 1990), heritage speakers should be more native-like than second language learners at least with early acquired aspects of the grammar. Research findings from morphosyntax have produced mixed results. Au et al. (2002) tested Spanish heritage speakers with very low productive ability in the language (what they call "overhearers") and incipient L2 learners on a 60-item grammaticality judgment task testing several features of the Spanish grammar (pronouns, gender, tense, agreement, subjunctive, etc.). Au et al. found no advantages for Spanish heritage speakers over second language learners in the domains of morphosyntax tested. However, Montrul, Foote & Perpiñán (2008) conducted a comprehensive study of Spanish gender agreement with over 140 Spanish heritage speakers and second language learners ranging from low to advanced proficiency in Spanish. They used two untimed written tasks and an untimed oral production task and found differences between the two groups depending on task type: second language learners were more target-like than the heritage speakers on the written tasks, whereas the heritage speakers were more target-like than the second language learners on the oral production task.

A potential limitation of Montrul et al.'s (2008) study is that because L2 learners, who learned the language primarily in the classroom, have more experience with written language than heritage speakers, the untimed written tasks employed may have overestimated their implicit knowledge of gender. The present follow-up study addresses this potential shortcoming by implementing a different set of tasks that might prove more efficient in tapping the participants' more automatic and implicit knowledge of grammatical gender. These tasks are an aural/oral word repetition task (RT), an aural gender-monitoring task (GMT) and an aural grammaticality judgment task (GJT). The tasks also vary on degree of explicitness: the GJT and the GMT require subjects to pay attention to form (by choosing grammatical/ungrammatical or masculine/feminine) and are therefore more explicit, while the RT taps more implicit and automatic knowledge and use. In general, we predict an advantage for heritage speakers, but if the degree of explicitness of the task, or metalinguistic awareness, matters for late bilinguals, the second language learners should be more accurate than heritage speakers on the GMT and the GJT than on the RT. Since heritage speakers were shown to perform better than second language learners on oral tasks, they will perform more native-like than the second language learners on the RT task in comparison to the other two more explicit tasks.

2. Method

2.1. Oral/Aural Repetition Task (RT)

The stimuli for the Spanish word repetition task consisted of 300 Spanish determiner-adjective-noun (DAN) phrases (determiner and adjective primes and noun targets, in the order determiner, adjective and noun). Following Bates et al. (1996) and Guillelmon & Grosjean (2001), all nouns and adjectives began with a stop consonant. We made this decision because the results of prior research on gender and lexical access (Bates, Devescovi, Pizzamiglio, D'Amico, & Hernandez,

1995) found that items that began with a fricative consonant resulted in significantly slower reaction times. Likewise, following Bates et al. (1996), all nouns had inanimate referents in order to eliminate the possibility for interactions between semantic and grammatical gender. Slang terms, proper names, acronyms and technical, highly abstract or context-specific words were not included. Additionally, care was taken to select nouns and adjectives according to norms for word frequency in Spanish (Alameda & Cuetos, 1995). Only words with a frequency of 3 or higher (calculated out of the written corpus of two millions words) were used.

The 300 DAN phrases were constructed from a set of 3 Spanish determiners, 7 adjectives and 150 nouns. The three determiners were *el* ('the' masc.), *la* ('the' fem.) and *su* ('his/her' neuter). In Bates et al. (1996), Italian determiners were not used; instead, a phonologically gender-transparent prenominal adjective was used as the sole primer for the following target noun. We chose to include entire DAN Spanish phrases (following Guillelmon & Grosjean, 2001) for three reasons. First, a DAN phrase that includes a determiner such as *la quinta pelota* ('the fifth ball') is considerably more frequent in Spanish than one that does not, i.e. *quinta pelota* ('fifth ball'). Second, prior research (Lew-Williams & Fernald, 2007 and references therein) suggests that information regarding the grammatical gender of a noun referent is accessed primarily at the determiner rather than at a concordant adjective or even in the gender suffix of the noun. Third, by incorporating the determiner as the primer for the gender of the noun referent, we were able to vary the phonological gender-transparency of the adjective (transparent *quinto*/*a* - 'fifth' masc./fem. vs. non-transparent *gran* 'big' masc./fem.). Such a distinction is important because in Guillelmon & Grosjean (2001), only non-transparent adjectives were used in order to establish the determiner as the sole primer of the gender of the noun referent. Though they found no effects of gender (in)congruency with respect to the determiner and noun for late bilingual French learners, they proposed for future study the incorporation of transparent adjectives, as their addition as a second primer may increase late-L2 learners' sensitivity to effects of gender (in)congruency. The DAN phrase construction used in the present study, therefore, facilitated the investigation of this proposal as well as the replication of results from Guillelmon & Grosjean (2001).

A total of 7 adjectives were used in the present study: 5 were transparent, ending in either *o* or *a* depending on a subsequent masculine or feminine noun referent respectively, and 2 were non-transparent, ending in *e* or a consonant. Though efforts were made to standardize all adjectives with respect to syllabic length and position of prosodic stress, the need to select only adjectives that were pragmatically acceptable in prenominal position, started with a stop consonant and had word frequencies of at least 3 made this impossible. As such, adjectives vary in syllabic length, from 1 to 4 syllables, and position of prosodic stress; the 5 transparent adjectives have paroxytone stress whereas the 2 non-transparent adjectives have oxytone stress.

150 Spanish nouns were used, all of which had paroxytone prosodic stress, began with a stop consonant and had a word frequency of at least 3. They consisted

of 20 masculine disyllabic nouns ending in *o*, 20 masculine trisyllabic nouns ending in *o*, 20 feminine disyllabic nouns ending in *a*, 20 feminine trisyllabic nouns ending in *a*, 20 masculine disyllabic nouns ending in *e*, 20 masculine trisyllabic nouns ending in *e*, 12 feminine disyllabic nouns ending in *e*, 8 feminine trisyllabic nouns ending in *e*, 3 masculine disyllabic nouns ending in *a* and 7 masculine trisyllabic nouns ending in *a*. It was not possible to complete the final four sets of 20 nouns due to our condition of a minimal word frequency of 3. The group of 10 masculine nouns ending in the canonical feminine suffix *a* (such as *tema* 'theme' masc.) were incorporated in our study because they may help better assess the contribution of overt phonological cues (in this case, "mis-cues") to the recognition and processing of grammatical gender (Bates et al. 1996: 996). Additionally, care was to taken to match word frequency and uniqueness point across all noun groups.[1] Feminine nouns ending in "a" and masculine nouns ending in "o" are canonical ending nouns. All other endings ("e", "consonant", opposite vowel) are non-canonical. We included nouns with both canonical and non-canonical endings because Montrul, Foote and Perpiñán (2008) found that canonicity made a difference for both second language learners and heritage speakers: both groups were more accurate with gender agreement on canonical than with non-canonical ending nouns.

In order to construct the 300 total DAN Spanish phrases, we followed the experimental set-up of Guillelmon & Grosjean (2001). The DAN phrases were prepared in stages. In stage 1, exactly three determiners were chosen, namely *el*, *la* and *su*, which were eventually spliced in as the determiners for all DAN phrases. A native Spanish-speaking female was recorded in a sound-proof studio uttering a variety of DAN phrases (at a normal rate) with a group of nouns starting with a stop consonant and which can be masculine and feminine, for example, *el gran capital*, *la gran capital* and *su gran capital* ('the/his/her big capital~money'). None of these nouns was included in the experiment. From these DAN phrases, the five best exemplars of each determiner were chosen by the evaluation of 2 judges, and spliced out. The exemplar-spliced determiners were then measured and the best exemplar of each of the three determiners was manipulated using the acoustic editing software Praat so that they were of the same duration: 0.156571 seconds. At the end of stage 1, we obtained 3 .wav files, each containing one of *el*, *la* and *su* of good acoustic quality and of identical duration.

In stage 2, 300 adjective-noun pairs were created by randomly assigning an adjective of both transparency types to each of the 150 nouns, which avoided complications regarding potential semantic effects of an adjective with a noun. The

1. Means of word frequency were collected from the Léxico informatizado del español (LEXESP) database (Sebastián Gallés, N., Cuetos, F., Carreiras, M., & Martí, M.A., 2000). 1-way ANOVA tests were performed for all noun groups (except the group of masculine nouns with a non-canonical *a*-ending, as there were too few nouns on which to calculate any statistically significant difference in mean frequency or uniqueness point length) and yielded F ratios of <1.00, suggesting no significant differences across conditions in word frequency and uniqueness point.

same speaker from stage 1 was recorded saying all 900 possible DAN phrases with each of the 300 adjective-noun pairs (i.e., 6 DAN phrases per noun, such as *el peor texto*, **la peor texto*, *su peor texto*, *el quinto texto*, **la quinta texto* and *su quinto texto*). In all cases, transparent adjectives agreed with the determiner, or in the case of the neutral *su*, were concordant with the noun.

Next, all 150 nouns were randomly though evenly stratified to one of the determiners from stage 1 (*el*, *la* or *su*). The noun *texto*, for example, was randomly assigned to *la*, its ungrammatical determiner, meaning that the *la* from stage 1 would eventually be spliced onto 2 stage 2 DAN phrases with *texto*, one with a transparent adjective (*quinto*/*a*) and one with a non-transparent adjective (*peor*). The only remaining complication involved deciding which 2 of the 6 total DAN variations of *texto* from stage 2 would be used as the base onto which stage 1 *la* would be spliced, for which another process of evenly stratified randomization was performed.

This randomization process was performed in order to account for any effects of determiner type on the pronunciation of the 900 DAN phrases of stage 2. Crucially, the 450 stage 2 DAN phrases with non-transparent adjectives were treated differently than the 450 with transparent adjectives, as is explained below. The 450 with non-transparent adjectives were randomly divided into 3 categories according to determiner type (grammatical, ungrammatical or neutral). For example, the DAN phrases with *texto* were randomly assigned neutral, and thus of the 3 aforementioned DAN phrase variations of *texto* with the non-transparent adjective *peor*, the one with *su* was chosen as the base onto which to splice the stage 1 determiner *la*, yielding **la peor texto*. Therefore, a third of the final 150 DAN phrases with non-transparent adjectives were derived from stage 2 bases whose determiners were grammatical, ungrammatical and neutral.

The same process could not be performed with the 450 remaining DAN phrases with transparent adjectives, since any randomly assigned to a stage 1 ungrammatical determiner would require an agreeing adjective only recorded with an ungrammatical determiner in stage 2. In other words, since *texto* was randomly assigned the stage 1 determiner *la*, the agreeing explicit adjective in the DAN phrase chosen from stage 2 must be marked with the feminine suffix *a*: *quinta*. The only DAN phrase recorded in stage 2 with *quinta* is **la quinta texto*, and thus only the stage 2 DAN phrase **la quinta texto* could be used as the base onto which to splice stage 1 *la*.

For nouns not assigned a stage 1 ungrammatical determiner, another evenly stratified randomization process was needed, i.e., since *cuerpo* (body - masc.) was randomly assigned a grammatical stage 1 determiner (*el*), either of stage 2 DAN phrases *el cuarto cuerpo* and *su cuarto cuerpo* could serve as a valid base onto which to splice stage 1 *el*. Half of these kinds of final DAN phrases were derived from stage 2 DAN phrases with a grammatical determiner and the other half were derived from stage 2 DAN phrases with neutral *su*. We summarize the synthesis process below in Table 1.

Table 1 - Overview of DAN Phrase Synthesis

Noun	Assigned Adjective (Random)	Stage 2 DAN Phrases	Assigned Stage 1 Deter-miner (Random)	Stage 2 Candidates for Splicing (**Bolded** Random Candidate Choice)	Final DAN Phrase
texto 'text'	*peor* 'worst'	*el peor texto* **la peor texto* *su peor texto*	**la*	*el peor texto* **la peor texto* ***su peor texto***	**la peor texto*
	quinto/a 'fifth'	*el quinto texto* **la quinta texto* *su peor texto*		**la quinta texto*	**la quinta texto*
cuerpo 'body'	*peor* 'worst'	*el peor cuerpo* **la peor cuerpo* *su peor cuerpo*	*el*	***el peor cuerpo*** **la peor cuerpo* *su peor cuerpo*	*el peor cuerpo*
	cuarto/a 'fourth'	*el cuarto cuerpo* **la cuarta cuerpo* *su cuarto cuerpo*		*el cuarto cuerpo* ***su cuarto cuerpo***	*el cuarto cuerpo*
guerra 'war'	*peor* 'worst'	**el peor guerra* *la peor guerra* *su peor guerra*	*su*	****el peor guerra*** *la peor guerra* *su peor guerra*	*su peor guerra*
	cuarto/a 'fourth'	**el cuarto guerra* *la cuarta guerra* *su cuarta guerra*		*la cuarta guerra* ***su cuarta guerra***	*su cuarta guerra*

Adjustments were made to eliminate discrepancies in sound volume between the stage 1 determiners and the spliced stage 2 adjective-noun pairs. In some instances, volume adjustments were sufficient to mask changes in pitch between the stage 1 determiners and spliced stage 2 adjective-noun pairs. When the pitch was too drastically different to pass for a sample of natural speech, which occurred in 29 cases out of 300, we rerecorded the appropriate stage 2 DAN phrases with the same native speaker. We provide Table 2 below as a brief synthesis of the types of stimuli created using the aforementioned synthesis techniques.

Table 2 - Example Stimuli for All 3 Experimental Tasks

DAN Phrase Condition	Noun Gender	Noun Ending	
		Canonical	**Non-Canonical**
Grammatical	Feminine	*la gran guerra* *la quinta guerra*	*la gran calle* *la cuarta calle*
	Masculine	*el peor texto* *el quinto texto*	*el peor viaje* *el quinto viaje*
Ungrammatical	Feminine	** el gran guerra* ** el quinto guerra*	** el gran calle* ** el cuarto calle*
	Masculine	** la peor texto* ** la quinta texto*	** la peor viaje* ** la quinta viaje*
Neutral (Control)	Feminine	*su gran guerra* *su quinta guerra*	*su gran calle* *su cuarta calle*
	Masculine	*su peor texto* *su quinto texto*	*su peor viaje* *su quinto viaje*

Once all 300 DAN phrases were synthesized, they were amplified to 60 decibels and placed as .wav files in SR Research Experiment Builder v.1.4.624, an EyeTracker software program that allows for a computer presentation of audio stimuli, can record subject speech and measure the reaction times of subject oral production after a given stimulus. The presentation order of each DAN phrase was randomized for every subject.

Participants met individually with a research assistant and completed the RT in the UIUC Second Language Acquisition and Bilingualism lab. Participants were given a set of headphones with a recording microphone attached and sat in front of a computer. They were asked to listen to a series of 3-word phrases and to repeat the last word of each phrase as quickly and accurately as possible after they heard the entire 3-word phrase. A blank screen was shown on the computer (which remained blank throughout the entire experiment) and after a 2 second pause, the first DAN phrase was played. The program recorded the subjects' speech throughout the experiment without pause, though upon the subject's production onset of the target noun, the program calculated the reaction time from the onset of the noun in the DAN phrase audio file. The next blank slide (a seamless white screen to the subject) was displayed 2 seconds after the subjects' production onset of a target noun, or in the case of silence from the subject, 4 seconds after the end of the audio file. The next DAN phrase would then play 2 seconds after the transition to a new blank slide, repeating until the end of the experiment. A 12 item trial session was performed, consisting of 12 un-manipulated natural speech DAN phrases with animate nouns and the same adjectives and determiners as in the RT. Both accuracy and reaction times were measured.

2.2. Aural Gender Monitoring Task (GMT)

The stimuli for the gender-monitoring task came from the audio databank previously described for the RT in section 2.1. The same 150 noun-adjectives pairs and stage 1 determiners were used, though their distribution was unique. For this task, the third of the 150 nouns that in the RT was assigned a stage 1 ungrammatical determiner was here assigned a stage 1 grammatically neutral determiner. The 50 that were assigned a stage 1 grammatical determiner in the RT were here assigned a stage 1 ungrammatical determiner, and the remaining 50 assigned a grammatically neutral determiner in the RT were here assigned a stage 1 grammatical determiner. The same synthesis process (as outlined in Table 1 above) was performed once more, creating 300 new DAN phrases.

The 300 DAN phrases (each amplified to 60 decibels) were placed into E-Prime v.1.2, a psychology software program that presents audio and visual stimuli and records the reaction times of subjects pushing a button on the computer keyboard.

Participants met individually with a research assistant and completed the GMT in the UIUC Second Language Acquisition and Bilingualism lab. Participants were given a set of headphones and sat in front of a computer. They were asked to listen to a series of 3-word phrases 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. Subjects were asked to rest their index fingers on the masculine and feminine buttons throughout the duration of the experiment. A solid blue screen was shown on the computer and after a 1 second delay, the first DAN phrase was played. The program calculated the subjects' reaction time of pushing either button from the onset of the target noun. A solid white screen was displayed for 1 second once a subject pushed a button, or in the case of no subject response, 4 seconds after the end of the audio file. This screen indicated a pause between audio files, and after its 1-second duration was replaced with another solid blue screen that signaled to the subject that a new DAN phrase was about to play (after 1 second). This alternation of blue and white screens repeated until the end of the experiment. A 12 item trial session was performed using the same trial DAN phrases in the RT. Both accuracy and reaction times were measured.

2.3. Aural Grammaticality Judgment Task (GJT)

The stimuli for the grammatically judgment task came from the audio databank previously described for the RT and GMT. The same 150 noun-adjectives pairs and stage 1 determiners were used, though their distribution was unique. In this task, following Bates et al. (1996), the stage 1 neutral determiner *su* was not used. Since *su* is not marked for gender and the adjective in DAN phrases with *su* was made to always agree with the noun, every DAN phrase with *su* would be grammatical.

Native Italian speakers when performing this task in the study of Bates et al. (1996) hit the grammatical button immediately upon hearing *su*, thus not even waiting to hear the target noun. As such, we eliminated this stage 1 determiner from this task.

The third of the 150 nouns that in the RT was assigned a stage 1 ungrammatical determiner was here assigned a grammatical determiner. The 50 nouns that in the RT were assigned a neutral determiner were here assigned an ungrammatical determiner. The remaining 50 (which already received a grammatical determiner in the RT and an ungrammatical determiner in the GMT) were randomly split in half and added to the other two GJT groups, yielding 75 nouns to be spliced with a stage 1 grammatical determiner and 75 nouns to be spliced with a stage 1 ungrammatical determiner. The synthesis process for the RT (above in Table 1) was performed once more, resulting in 300 DAN phrases, over two-thirds of which were distinct from DAN phrases presented in the RT and GMT. These 300 DAN phrases (each amplified to 60 decibels) were placed into E-Prime v.1.2 once more.

The procedures for the GJT were identical to those of the GMT, 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 3-word phrases and to push one of the two buttons on the keyboard to signify whether or not the phrase was grammatical or ungrammatical. Both accuracy and reaction times were measured.

2.4. Subject Pool

A total of 23 Spanish native speakers, 29 heritage speakers and 33 intermediate to advanced L2 learners were recruited for testing. Participants were recruited from the Champaign-Urbana community and the UIUC student population. UIUC students enrolled in a Spanish bilingualism course received course credit for participating, whereas the rest were monetarily compensated. Subjects completed two proficiency measures, one a written metalinguistic proficiency exam, and the other a picture-naming task (completed once in Spanish and once in English, though Spanish native speakers only completed the task in Spanish). The picture-naming task was included as a non-metalinguistic oral production proficiency measure, following O'Grady (2009), since the RT, GJT, and GMT spanned a similar continuum from lower (and oral) to higher (and non-oral) metalinguistic tasks. Recall that Montrul et al. (2008) found significant effects of task type on performance, with heritage speakers performing better than L2 learners on oral, less metalinguistic tasks.

The picture-naming task consisted of 48 inanimate Spanish nouns (with a frequency of 3 or higher, cf. Alameda & Cuetos, 1995): 12 masculine nouns ending in canonical -o, 12 feminine nouns ending in canonical -a, 6 masculine nouns ending in a non-canonical consonant, 6 feminine nouns ending in a non-canonical consonant, 6 masculine nouns ending in non-canonical -e and 6 feminine nouns ending in non-canonical -e. Subjects viewed black and white images on a computer screen (randomly generated using SR Research Experiment Builder v.1.4.624) and

were prompted to say the name of the object in English or Spanish as quickly as possible after hearing "say" or "*diga*," respectively. Both accuracy and reaction times were measured.

With respect to the written, highly metalinguistic proficiency exam, heritage speakers showed a similar range of scores to that of the L2 learners, though the L2 learner scores' distribution was wider than that for heritage speakers. Both groups scored within the range of 30 and 48 (out of a maximum of 50 points).

With respect to the oral picture-naming task, native speakers were significantly more accurate and faster than both heritage speakers and L2 learners ($p < .001$). Crucially, across both the Spanish and English versions of the task, heritage speakers and L2 learners did not significantly differ from one another with respect to accuracy and reaction time. These results are displayed below in Table 3.

Table 3 - Results of Picture-naming Tasks

	Reaction Times (ms)		Accuracy Scores (%)	
	Spanish	English	Spanish	English
NS	788.17	--	95.45	--
HS	1110.32	854.58	88.2	97.62
L2	1227.13	764.09	87.96	98.53

3. Results

3.1. Accuracy and Reaction Times on the GJT and GMT

Mean reaction times and mean accuracy scores were submitted to a mixed ANOVA analysis with grammaticality and canonicity as within-subject variables and group as a between-subjects variable. All three subject groups showed a significant ($p < .001$) effect of grammaticality on accuracy across the GJT and GMT, the two more metalinguistic tasks. We present these results below in Table 4.

Table 4 - Accuracy Results for the GJT & GMT (in %)

	GJT				GMT			
	Grammatical		Ungrammatical		Grammatical		Ungrammatical	
	Can	NonCa	Can	NonCa	Can	NonCa	Can	NonCa
NS	99	98	98	92	99	98	99	93
HS	96	90	90	62	96	81	82	60
L2	95	79	91	58	98	72	89	57

Accuracy was consistently lower for stimuli in the ungrammatical condition than in the grammatical condition, suggesting that all three groups used gender cues on determiners and adjectives in noun recognition. Moreover, whereas the native

monolinguals performed at ceiling (showing no significant effect of canonicity on accuracy), both experimental groups showed a significant effect of canonicity (Can vs. NonCa in Table 4) ($p < .001$), performing consistently more accurately with nouns with canonical endings. These results were similarly echoed with respect to reaction times across the GJT and GMT, in that all groups showed an effect of grammaticality. Additionally, heritage speakers and L2 learners were consistently slower with nouns of non-canonical endings ($p < .001$), whereas native monolinguals did not show this effect. The reaction time results for the GMT and GJT are illustrated below in Table 5.

Table 5 - Reaction Time Results for the GJT & GMT (in ms)

	GJT				GMT			
	Grammatical		Ungrammatical		Grammatical		Ungrammatical	
	Can	NonCa	Can	NonCa	Can	NonCa	Can	NonCa
NS	1010	1037	1103	1252	1085	1290	1238	1418
HS	1394	1463	1439	1778	1286	1538	1479	1631
L2	1532	1822	1667	1971	1123	1441	1265	1561

Across the GMT and GJT, heritage speakers and L2 learners patterned similarly, in that both showed effects of both canonicity and grammaticality on accuracy and reaction time, whereas native monolinguals only showed an effect of grammaticality. That heritage learners did not show any clear advantages over L2 learners across these two tasks was expected, as Montrul et al. (2008) found that heritage learners perform best at oral, less metalinguistic tasks. Accordingly, we now turn to these results of the RT in order to shed light on whether or not the lack of performance differences between the two experimental groups is consistent across even oral, less metalinguistic tasks.

3.2. Reaction Times on the RT

Accuracy scores on the RT were at ceiling across the three groups and, thus, were not subjected to further analyses. The reaction time data, on the other hand, showed that the heritage speakers patterned with the native monolinguals as opposed to with the L2 learners. The former groups both showed a significant effect of grammaticality ($p < .001$), whereas no such effect was present for the L2 learners. This result is illustrated below in Table 6.

Table 6 - Reaction Time Results for the RT (in ms)

	Grammatical		Ungrammatical	
	Canonical	Non-canonical	Canonical	Non-canonical
NS	726	749	792	757
HS	838	832	876	838
L2	775	825	804	795

The lack of effect of grammaticality for L2 learners in the RT suggests that L2 learners may lack the implicit knowledge of gender that both native monolinguals and heritage speakers access when completing oral, less metalinguistic tasks. Our results therefore corroborate those of Montrul et al. (2008), in that heritage speakers were found to have an advantage over L2 learners only in performance in the RT, the oral and least metalinguistically demanding task of the three.

4. Conclusions

The present study re-examined whether or not heritage speakers have advantages in morphosyntax over L2 learners. Following up on Montrul, Foote & Perpiñán (2008), we removed off-line written tasks to more directly assess the potential involvement of implicit linguistic knowledge rather than metalinguistic knowledge. Using proficiency-matched heritage speakers and L2 learners of Spanish (according to proficiency measures varying in metalinguistic focus), we have shown that heritage speakers do have advantages in morphosyntax over L2 learners, specifically on oral, less metalinguistic tasks such as the RT in comparison with the GMT and GJT. Only on the RT did heritage speakers pattern with the native monolinguals in showing an effect of grammaticality, indicative of a possible lack of implicit knowledge on the part of L2 learners with gender processing. Prior research offers insight as to the potential cause for this morphosyntactic advantage for heritage speakers - heritage speakers have an earlier onset age for bilingualism than L2 learners (Guillelmon & Grosjean, 2001), and heritage speakers differ from L2 learners both regarding the context for language acquisition (naturalistic vs. formal instruction) and their overall experience with oral production (Montrul, in press).

While we have offered evidence in support of a morphosyntactic advantage for intermediate to advanced proficiency heritage speakers over L2 learners on oral, less metalinguistic tasks (cf. Au et al., 2001), future investigation regarding how gender is processed and accessed by heritage speakers and L2 learners is still warranted. For example, recall that the stimuli used in the present study consisted of DAN phrases, precisely half of which contained a transparent adjective canonically marked for gender. Data analysis is currently in progress in order to compare performance differences between experimental groups and across task types as affected by the presence of an adjective transparently marked for gender. A lack of effect of adjective transparency, for example, could suggest that information about gender

used in noun recognition is gathered primarily from the determiner in noun phrases (corroborating Lew-Williams & Fernald, 2007 and references therein). Moreover, data analysis is also currently ongoing as to any effects of the type of non-canonical noun ending on performance across the three tasks; a more detailed look, for example, at performance with noun stimuli such as *el tema* (theme - masculine), which feature a noun ending canonical to the opposite gender of the noun, can offer insight as to how well different groups of learners access gender information offered in the determiner and adjective, particularly since for these stimuli, the vowel ending of the noun is deceivingly non-canonical.

References

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

Au, T., Knightly, L, Jun, S., & Oh, J. (2002). Overhearing a language during childhood. *Psychological Science* 13, 238–243.

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

Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender priming in Italian. *Perception & Psychophysics* 57, 847-862.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis,* 20, 3-49.

Clahsen, H., & Muysken, P. (1986). The UG paradox in L2 acquisition. *Second Language Research,* 5, 1-29.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition,* 22, 499-534.

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

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

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition,* 12, 251-285.

Montrul, S. (in press). Morphological variability in L2 learners and heritage speakers: The role of experience. *Studies in Second Language Acquisition.* Special issue on the linguistic competence of heritage speakers, 2011.

Montrul, S. (2010). How similar are adult second language learners and Spanish heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics,* 31, 167-207.

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.

O'Grady, W. (2009). Assessing heritage language competence. Paper presented at the Third Heritage Language Institute, University of Illinois at Urbana-Champaign, June 2009.

Sebastián Gallés, N., Cuetos, F., Carreiras, M., & Martí, M.A. (2000). LEXESP: Léxico informatizado del español [LEXESP: A computerized database of Spanish]. Barcelona: Universitat de Barcelona.