

BUCLD 35 Proceedings
To be published in 2011 by Cascadilla Press
Rights forms signed by all authors

Grammatical Gender in L2: Where Is the Problem?

Theres Grüter, Casey Lew-Williams, and Anne Fernald

Grammatical gender is a difficult property to master for second language (L2) learners, an observation that raises the question of whether native-like mastery of grammatical gender in an L2 is attainable at all. Despite a wealth of research on the acquisition of grammatical gender by non-native speakers over the past decades, no consensus with regard to this question has been reached. Part of the difficulty in addressing the question lies in the definition of ‘mastery’ itself. From a generative perspective, mastery of grammatical gender crucially involves the acquisition of the functional gender feature which conditions gender agreement in the syntax. Thus recent work in this tradition has investigated learners’ performance on carefully constructed offline comprehension tasks, where good performance is argued to be possible only if the learner’s grammar includes the relevant functional feature (e.g., White et al. 2004, Montrul et al. 2008). These studies have shown that advanced-proficiency L2 learners perform at ceiling on such tasks, leading to the conclusion that the functional gender feature is acquirable in an L2, and that any remaining errors with grammatical gender that might occur in production must be due to a production-specific performance problem, consistent with the Missing Surface Inflection Hypothesis (MSIH, Prévost & White 2000).

This interpretation is consistent with data showing a dissociation between learners’ performance in spoken production versus offline comprehension. Yet it disregards a potentially important confound. The confound lies in the fact that offline tasks, such as the pen-and-paper comprehension tasks used in the studies cited above, are not time-constrained; however, language production by its very nature occurs in real time, and is thus always affected by the pressures of real-time processing. It therefore remains unclear whether persistent difficulty with grammatical gender in production is really a production-specific problem, or whether it might be a result of difficulty with the retrieval of gender information in real-time language use (both expressive and receptive) more generally.

From a psycholinguistic perspective, ‘mastery’ of grammatical gender has typically been operationalized in terms of sensitivity to gender cues and violations in real-time language processing. Findings from this line of research,

* Theres Grüter, Stanford University, tgruter@stanford.edu; Casey Lew-Williams, University of Wisconsin-Madison; Anne Fernald, Stanford University. This research was supported by grants from NIH to Theres Grüter (F32 DC010129-01) and to Anne Fernald (DC 008838). Many thanks also to Virginia Marchman and folks at the Center for Infant Studies at Stanford, as well as to all our participants.

however, have led to somewhat contradictory results in the case of L2 learners. Some report sensitivity to gender violations even in learners with relatively limited proficiency (Tokowicz & MacWhinney 2005, Sagarra & Herschensohn 2010), while others find differences between highly proficient L2 learners and native speakers (Guillelmon & Grosjean 2001, Gillon Dowens et al. 2010). Thus it remains an open question whether mastery of grammatical gender is possible in an L2, when ‘mastery’ is defined in terms of real-time processing.

To the best of our knowledge, no previous study has investigated grammatical gender in production, offline comprehension and online processing within the same learner group. Yet in order to tease apart the dimensions of production versus comprehension on the one hand, and offline versus real-time language use on the other, it is precisely such combined evidence – a crossing of the two factors experimentally – that is required (see Table 1).

Table 1. Summary of experiments in the present study

	OFFLINE	REAL-TIME
PRODUCTION	-- ¹	<u>Exp. 2</u> Elicited production
COMPREHENSION	<u>Exp. 1</u> Sentence-picture matching	<u>Exp. 3</u> Eye-tracking during listening

1. Goals and predictions

The goal of the present experiments was to investigate whether the persistent difficulty with grammatical gender that might be observed in the spoken language of highly proficient L2 learners is best characterized as a production-specific performance problem in the spirit of the MSIH (Hypothesis 1), or as difficulty with the retrieval of gender information in real-time language use – expressive and receptive – more generally (Hypothesis 2). Under Hypothesis 1, we may expect differences between L1 and the L2 groups in production (Exp. 2), but crucially, no differences are predicted for comprehension, assessed either by offline (Exp. 1) or online (Exp. 3) measures. Under Hypothesis 2, by contrast, between-group differences are expected in both production and online processing, but not in offline comprehension.

2. Participants

Participants were 19 L2 learners of Spanish (L2 group) and 19 native Spanish speakers (L1 group). Participants in the L2 group were native speakers of English with no exposure to Spanish or other gender-marking languages before age 11. Criteria for inclusion in the L2 group were an overall score within

¹ Our design does not include an offline measure of production, as we could not conceive of an appropriate task that does not rely heavily on metalinguistic knowledge.

the top two tiers of the Versant Spanish Test (Pearson 2009), indicating advanced to near-native oral proficiency, as well as self-report of daily use of Spanish. Three measures of proficiency were collected: self-rating, performance on a written cloze test (White et al. 2004), and performance on the Versant Spanish Test as a measure of oral ability. Scores for both groups are summarized in Table 2. Although independent-samples *t*-tests indicate statistically significant differences between the L1 and the L2 group on all three measures (all $t(36) > 2.9$, $p < .01$), mean scores in the L2 group are close to ceiling, indicating advanced to near-native oral and written proficiency in Spanish.

Table 2. Results from proficiency measures (means and ranges).

	SELF-RATING (SPEAKING, 0-10)	WRITTEN CLOZE TEST (/50)	VERSANT SPANISH TEST (/80)
L1 (n=19)	9.9 (9-10)	48.6 (46-50)	79.9 (79-80)
L2 (n=19)	8.5 (6-10)	44.9 (37-49)	76.8 (69-80)

3. Experiment 1: Sentence-picture matching

3.1 Stimuli and procedure

Experiment 1 constituted a replication of the first experiment in Montrul et al. (2008), using a task originally developed by White et al. (2004).² Taking advantage of the fact that Spanish allows nouns to be omitted from a noun phrase when their referent is salient in the discourse (nominal ellipsis, or N-drop), this task was designed to investigate whether learners can identify the referent of a null noun in such a construction solely on the basis of gender (and number) marking on the determiner and/or the adjective. For example, the modifier *otra* ('other') in (1) is morphologically marked for feminine gender, thus constraining the choice of referent to those denoted by feminine nouns.

- (1) Tenemos que buscar otra ____.
 must-1-PL COMP find other-SG-FEM
 'We must find another one.'

As White et al. (2004: 116) argued, "this task provides a means of determining, via comprehension rather than production, whether abstract [gender and number] features are present in learner grammars."

The experiment consisted of 32 experimental and 13 filler trials. On experimental trials, participants were presented with a written sentence containing a noun-drop construction as in (1), with gender (and number) marked on the determiner and/or the adjective, together with a choice of three pictures each labeled with a bare noun, only one of which matched the gender marked on the modifiers. Half of the target nouns were masculine, half were feminine, and

² We would like to thank Silvina Montrul for sharing the stimuli used in the Montrul et al. (2008) study.

both gender conditions included singular as well as plural targets. As the number manipulation was not relevant here, singular and plural targets were collapsed for analysis. Filler trials consisted of sentences without noun-drop, also accompanied by a choice of three labeled pictures. The task was presented via a web interface. Participants were instructed that they would read sentences which were part of an overheard conversation between two people packing their suitcases for a trip, and their task was to guess what these people were talking about. Each trial was presented on a separate screen, and participants could advance to the next trial at their own pace after selecting one of the three choices. Experimental items were presented in the same single, semi-randomized order as in Montrul et al. (2008).

3.2 Results

Results showed that both groups performed similarly, and at ceiling, on this task, with a mean accuracy of 98% ($SD = 5.5$) in the L1 and 96% ($SD = 7.5$) in the L2 group. Thus, as predicted by both Hypotheses 1 and 2, the highly proficient L2 learners in this study performed like native speakers on an offline receptive measure targeting grammatical gender. Following White's rationale cited above, we may therefore conclude that these L2 learners have successfully acquired the abstract gender feature posited in linguistic theory.

4. Experiment 2: Elicited production

4.1 Stimuli and procedure

This task was modeled after the third experiment in Montrul et al. (2008), with some methodological modifications. The goal of the task was to elicit determiner-noun-adjective sequences with a given set of Spanish nouns (25 fem., 25 masc.).³ Unlike in Montrul et al.'s experiment, the task was presented entirely in the auditory modality. On each trial, participants saw two images depicting the target noun, with the images contrasting along at least one salient dimension (e.g., color). For example, for the target noun *mariposa* ('butterfly'), images of a red butterfly and a yellow butterfly were presented. Participants heard a prerecorded question, and were asked to make a choice between the two pictures by naming one of them (*¿Cuál mariposa prefieres?* 'Which butterfly do you like better?'). Their oral responses were audiotaped and later transcribed.

Only responses minimally containing a determiner and a gender-marked adjective were retained for further analysis. Approximately 70% of all responses fulfilled this criterion (L1: 70.5%, L2: 69.9%). These responses were coded for accuracy of (i) determiner-noun agreement, and (ii) determiner-adjective agreement, yielding four logically possible response types, illustrated in Table 3.

³ With a few exceptions due to difficulty with imageability, the same nouns as in Montrul et al. (2008) were used.

Table 3. Possible response types in Experiment 2.

	Det-N match	Det-N mismatch
	TARGET	ASSIGNMENT ERROR
Det-Adj match	a. <i>la (mariposa) roja</i>	b. <i>el (mariposa) rojo</i>
	AGREEMENT ERROR	(BOTH/?)
Det-Adj mismatch	c. <i>la (mariposa) rojo</i>	d. <i>el (mariposa) roja</i>

Type (a) represents a correct response, while (b)-(d) constitute possible error types. Type (b), where determiner and adjective agree with each other, but do not instantiate the gender required by the noun, suggests that speakers have misclassified the noun with regard to its gender class, an error at the lexical level. Consistent with previous work, we call these ERRORS OF GENDER ASSIGNMENT (e.g., Dewaele & Véronique 2001, Montrul et al. 2008). Errors of type (c) consist of a correct determiner but incorrect adjective. On the assumption that determiner choice is the most immediate reflection of a noun's lexical gender (Carroll 1989), this error type suggests that the speaker has correctly classified the noun with regard to its gender class, but has failed to compute or express gender agreement on the adjective. Thus errors of this type, which we will call ERRORS OF GENDER AGREEMENT, are most likely of a (morpho)syntactic nature. Finally, errors of type (d), with an incorrect determiner but what appears to be correct agreement between the noun and the adjective, are somewhat difficult to interpret, but might be considered errors of both gender assignment and agreement.

4.2 Results

As summarized in Table 4, results showed errors were exceedingly rare in the L1 group, where a mean of 98.7% of all responses were of type (a). In the L2 group, by contrast, significantly fewer responses were target-like ($M = 80.0\%$, Kolmogorov-Smirnov $Z = 2.76$, $p < .001$), with most of the remaining responses ($M = 17.2\%$) constituting assignment errors. Agreement errors, however, were rare in both groups, with no significant between-group differences.

Table 4. Mean proportion of response types by group in elicited production (standard deviations in parentheses).

	Det-N match	Det-N mismatch
	TARGET	ASSIGNMENT ERROR
Det-Adj match	L1: 98.7% (1.7) L2: 80.0% (14.6)	L1: 1.0% (1.6) L2: 17.2% (12.8)
	AGREEMENT ERROR	(BOTH/?)
Det-Adj mismatch	L1: 0.3% (0.9) L2: 1.5% (2.3)	L1: 0% L2: 1.3% (2.6)

These findings indicate that difficulty with grammatical gender indeed persists, even in L2 learners at highest levels of proficiency. Interestingly, however, these

difficulties appear to be mostly confined to errors of gender assignment, indicating that lexical, rather than morpho-syntactic properties of gender are the source of the problem. We will return to this observation in the discussion.

5. Experiment 3: Looking while listening

5.1 Stimuli and procedure

Experiment 3 employed the looking-while-listening procedure (Fernald et al. 2008), where participants view pairs of pictures projected onto a screen while listening to sentences naming one of them. Participants' eye movements are videotaped, and this record stamped with a digital time-code time-locked to the auditory stimuli. Using custom software, eye movements are then coded offline, frame by frame, with 33-ms resolution, allowing for a precise time-course analysis of participants' looking patterns as the speech signal unfolds.

Combining experimental stimuli previously developed by Lew-Williams and Fernald (2007, 2010), Experiment 3 contained two different experimental conditions. In the FAMILIAR NOUN condition, speech stimuli consisted of simple Spanish sentences using one of two sentence frames, *Encuentra...* ('Find') or *¿Dónde esta...* ('Where is'), followed by one of eight determiner-noun sequences, four masculine and four feminine: *la pelota* ('ball'), *la galleta* ('cookie'), *la vaca* ('cow'), *la rana* ('frog'), *el zapato* ('shoe'), *el carro* ('car'), *el pájaro* ('bird'), and *el caballo* ('horse'). Visual stimuli were colorful digital images of animals and objects. Participants were exposed to two trial types: SAME-GENDER trials, where the images depicted objects referred to by nouns of the same grammatical gender (e.g., *la pelota, la galleta*), and DIFFERENT-GENDER trials, where the images depicted objects referred to by nouns of different gender (e.g., *la pelota, el zapato*). In this latter case, the gender-marked determiner provided a potentially informative cue regarding the object being named. On SAME-GENDER trials, by contrast, no disambiguating information was provided before the onset of the noun itself. Thus, if listeners process gender-marking on determiners online, we expect their looks to converge on the target image faster on DIFFERENT- compared to SAME-GENDER trials. This is indeed what Lew-Williams and Fernald (2007) found in an experiment with adult as well as 3-year-old native speakers of Spanish. Here we ask whether the highly proficient L2 learners in this study demonstrate the same processing advantage.

Interspersed with these FAMILIAR-NOUN trials were trials in the NOVEL NOUN condition. During the first part of the experiment, participants were exposed to four novel nouns, each paired with a novel object. Two novel nouns were feminine and two were masculine. On these teaching trials, speech stimuli consisted of a simple sentence frame (*¡Mira, es*, 'Look, it's') followed by one of four novel Spanish nouns, each preceded by an indefinite article (*una catela, una pifa, un durino, un tebo*). Visual stimuli consisted of a single image of the respective novel object. During the second part of the experiment, participants were exposed to NOVEL-NOUN test trials, structured analogously to the trials in the FAMILIAR-NOUN condition. Speech stimuli consisted of the frame *¿Dónde*

esta... ('Where is'), followed by the novel noun preceded by the definite article (*la/el*). On SAME-GENDER trials, visual stimuli consisted of two images depicting objects referred to by novel nouns of the same gender (e.g., *la catela, la pifa*), while on DIFFERENT-GENDER trials, images depicted objects referred to by novel nouns of different gender (e.g., *la catela, el durino*). The experiment took approximately 8 minutes to complete, and consisted of a total of 79 trials: 32 in the familiar-noun condition, 20 teaching and 24 test trials in the novel-noun condition, as well as 3 fillers designed to maintain participants' attention.

Since participants could not know in advance which picture would be named, they were by chance equally likely to be looking at the target or distractor picture at the onset of the determiner. If they were already looking at the correct picture (target-initial trials), they should maintain fixation; but if they were looking at the distractor picture (distractor-initial trials), they should shift quickly to the named picture. Distractor-initial trials were used to calculate reaction time (RT), the latency to initiate an eye movement toward the target picture. RT was calculated from determiner onset, that is, the first moment in the unfolding sentence where participants received relevant acoustic information. Shifts initiated in the first 200 ms were not included in analyses, because they were likely to represent random shifting that occurred prior to the possible influence of the determiner. RTs were included in analyses if they occurred between 200 and 800 ms from determiner onset.

5.2 Results

5.2.1 Familiar-noun condition

Figure 1 shows the time course of orienting to the correct picture on same- and different-gender trials in the FAMILIAR-NOUN condition, for both native speakers and L2 learners. Visual inspection suggests that native speakers were faster to converge on the target picture on different- compared to same-gender trials, consistent with the findings of Lew-Williams and Fernald (2007). In the L2 group, by contrast, the difference in performance between the two trial types appears to be smaller. When mean RTs were entered into a 2 (group) x 2 (trial type) ANOVA, a significant main effect for trial type emerged ($F(1,35) = 7.2, p = .01$), showing that participants overall responded faster on different- than same-gender trials. However, neither the main effect for group ($F(1,35) = .7, p > .3$), nor the group by trial-type interaction ($F(1,35) = .7, p > .3$), reached significance. While the absence of these latter effects indicates that the two groups did not perform differentially, the effect of trial type may not be as robust in the L2 group. As shown in Figure 2, the difference in RT between same- and different-gender trials was smaller in the L2 (32 ms) compared to the L1 group (63 ms). We next performed planned comparisons within each group, using Bonferroni-corrected alpha levels ($p < .025$). These comparisons indicated a significant difference between same- and different-gender trials for the L1 ($t(18) = 2.8, p = .01, d = .93$), but not the L2 group ($t(17) = 1.2, p = .26, d = .39$).

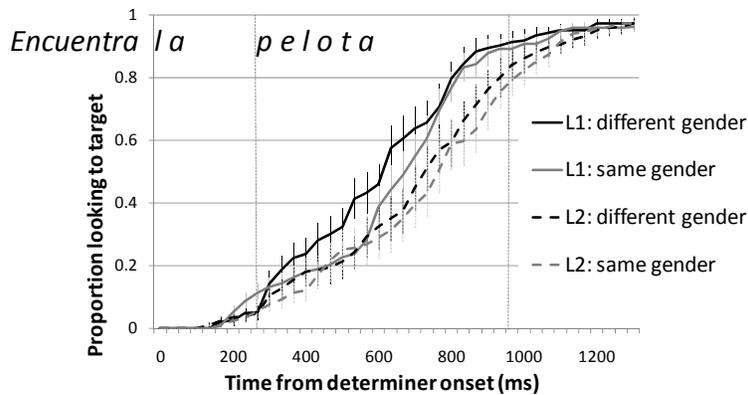


Figure 1. Time course of participants' looking to the target image on same- and different-gender trials in the FAMILIAR NOUN condition. Vertical dashed lines indicate onset and offset of the noun.

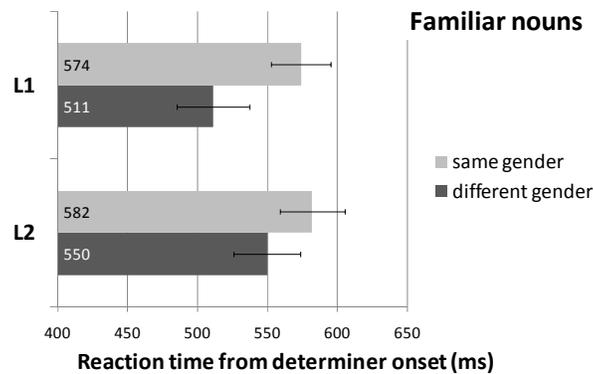


Figure 2. Participants' mean RTs in shifting from the distractor to the target picture on same- and different gender trials in the FAMILIAR NOUN condition. Error bars indicate standard errors of the means.

As the difference in the L2 group appeared to be in the same direction as that observed in the L1 group, the question arose as to whether performance in the L2 group as a whole was negatively impacted by those participants with somewhat lower proficiency. To explore this possibility, RTs for the subgroup of L2 learners with a perfect score (80) on the oral proficiency measure ($n=11$) were analysed separately. Mean RTs in this subgroup were not different for same- (548 ms) compared to different-gender (556 ms) trials, indicating the absence of a significant effect in the L2 group as a whole is unlikely to be due to differences in proficiency.

In sum, while the absence of a main effect for group and a group by trial-type interaction would indicate that participants in the L1 and L2 groups

performed alike, the failure to obtain a significant effect in the pair-wise comparison in the L2 group suggests that any processing advantage L2 learners may gain as a result of gender-marking on the determiner is somewhat weaker or less consistent than what can be observed in native speakers.

5.2.2 Novel-noun condition

Figure 3 shows participants' time course of orienting to the correct picture on same- and different-gender trials in the NOVEL-NOUN condition. Visual inspection indicates that native speakers were faster to converge on the target picture on different- compared to same-gender trials, consistent with Lew-Williams and Fernald (2010). In the L2 group, we also observe a difference between different- and same gender trials, although this difference appears to occur somewhat later than in the L1 group. RT measures, illustrated in Figure 4, confirm these observations, showing substantial differences in RT between same- and different-gender trials in both groups (L1: 62 ms, L2: 102 ms).

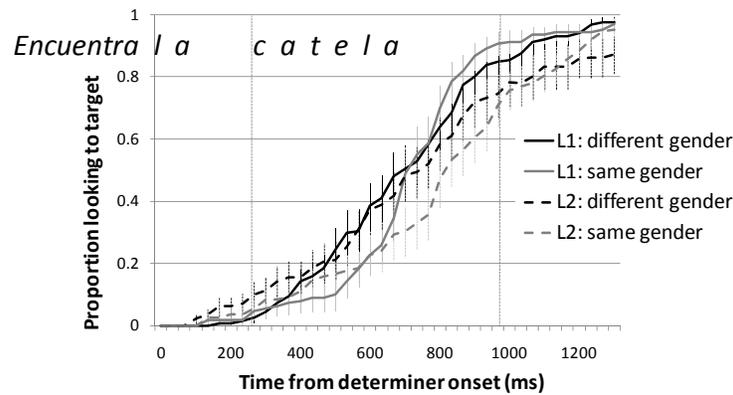


Figure 3. Time course of participants' looking to the target image on same- and different-gender trials in the NOVEL NOUN condition.

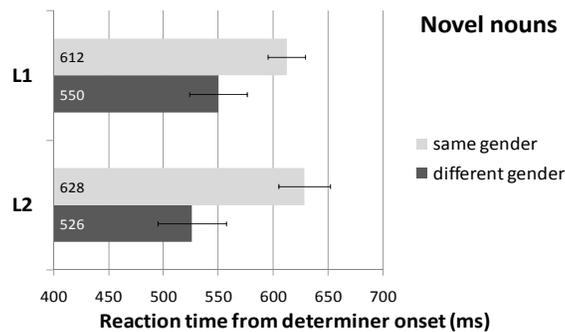


Figure 4. Mean RTs in shifting from the distractor to the target picture on same- and different gender trials in the NOVEL NOUN condition.

When mean RTs were entered into a 2 (group) x 2 (trial type) ANOVA, a significant main effect for trial type emerged ($F(1,33) = 10.3, p = .003$). Neither the main effect for group ($F(1,33) = .03, p > .8$) nor the interaction ($F(1,33) = .6, p > .4$) were significant. In Bonferroni-corrected planned comparisons within each group ($p < .025$), analogous to those performed in the familiar noun condition, the difference between same- and different-gender trials was marginally significant in both the L1 ($t(18) = 2.1, p = .05, d = .68$) and the L2 group ($t(15) = 2.4, p = .03, d = .86$).

5.3 Discussion

Experiment 3 confirmed earlier findings by Lew-Williams and Fernald (2007, 2010), demonstrating that native speakers of Spanish use gender-marking on the determiner as a predictive cue in online language processing. However, unlike Lew-Williams and Fernald (2010), who included a group of L2 learners of Spanish with intermediate levels of proficiency, we tested a group of highly proficient L2 learners, and obtained no main effect for group in the familiar- nor the novel-noun condition. This null finding may be taken to indicate that online processing of gender-marking on determiners by highly proficient L2 learners' does not differ from that observed in native speakers. Such a conclusion is qualified, however, by the observation that pairwise comparisons between RTs on same- versus different-gender trials did not reach significance in the L2 group in the familiar-noun condition, suggesting that processing advantages related to gender-marking may be weaker or more variable in non-native compared to native speakers. Interestingly, this qualification does not appear to apply to the novel-noun condition, an observation we will return to in the general discussion.

6. General discussion

In this study, we investigated grammatical gender in production, offline comprehension and online processing within the same, highly proficient L2 learner group, with the goal of teasing apart whether advanced learners' difficulty with this property is confined to expressive language use (Hypothesis 1), or whether it affects the retrieval of gender information in real-time language use more generally (Hypothesis 2). Consistent with both hypotheses, the L2 group performed at ceiling in offline comprehension (Exp. 1). Also consistent with both hypotheses, we observed significant differences between the L1 and the L2 group in elicited production (Exp. 2). These differences, however, were confined to errors of gender assignment, which were significantly more frequent in the L2 group. Errors of gender agreement, on the other hand, were rare in both groups, suggesting that persistent difficulty with grammatical gender in L2 primarily affects lexical, rather than syntactic aspects of gender. Finally, evidence from the online processing of gender-marked determiners was obtained (Exp. 3) in order to adjudicate between the two hypotheses. Results, however,

were not entirely unequivocal. On the one hand, the absence of significant between-group differences and interactions would suggest that L2 learners and native speakers did not differ in their online processing of gender-marking on determiners, as predicted by Hypothesis 1. On the other hand, planned comparisons within each group, as well as visual inspection of the time course of looking patterns suggested that the effect in the familiar noun condition was not robust in the L2 group, indicating a weakness in the use of grammatical gender cues in L2 online processing, consistent with Hypothesis 2. This outcome would be unexpected under Hypothesis 1, where difficulties are predicted to be confined to expressive language use. Interestingly, a more robust effect was obtained in the novel word condition in the L2 group, suggesting that use of gender cues in online processing is not beyond the abilities of L2 learners in principle. We will return to this intriguing finding after considering in more detail a potential source of the L1-L2 difference observed in Experiment 2.

Findings from Experiment 2 suggested that lexical, rather than syntactic properties of gender may be the primary source of difficulty for advanced L2 learners. Let us thus take a closer look at the lexical representation of grammatical gender, and consider how such representations may be acquired. In gender-marking languages, noun class membership must be stored as an inherent property of each noun as part of its entry in the mental lexicon. One way of formalizing this is to assume that “all nouns of a given grammatical gender are linked to one gender node specifying that grammatical gender” (Schriefers & Jescheniak 1999: 577). Of relevance here is how such links or associations between nouns and gender nodes are established in L1 acquisition. Given that phonological and semantic cues alone are insufficient for establishing membership in gender classes in Spanish (Harris 1991), learners must rely extensively on co-occurrence relations between nouns and gender-marked modifiers, most importantly determiners, to detect a noun’s gender. The computation of co-occurrence relations and transitional probabilities more generally has been shown to be a key mechanism in infants’ early language learning (Saffran et al. 1996). As (backward) transitional probabilities between determiners and nouns are typically high (see Pelucchi, Hay & Saffran 2009), it is thus not surprising that young children occasionally fail to segment det-N sequences and treat them as unanalyzed chunks (e.g., Carroll 1989). This kind of error illustrates the tight associations that are formed between frequently co-occurring elements such as determiners and nouns in early L1 lexicons. While at early stages these tight associations between gender-marked determiners and nouns are likely to be lexically specific, it seems reasonable to assume that with increased vocabulary size they will give rise to associations between nouns and more abstract gender classes, as instantiated for example through the gender nodes posited by Schriefers and Jescheniak (1999). Importantly, given the strong initial associations between nouns and their gender-marked modifiers as a result of early language processing, associations between nouns and gender nodes in the more mature L1 lexicon can be expected to remain strong, making gender-marking a powerful cue for processing in a native language.

How do L2 learners establish a noun's membership in a gender class? While adults are able to use transitional probabilities to learn properties of an unfamiliar artificial language in laboratory settings, computing distributional information is not the only, and almost certainly not the most efficient way to learn nouns in a real-world L2 context. Importantly, learners beyond early childhood have access to many more cues than infants to identify and learn words from the input. These include translation equivalents from the L1, metalinguistic information, and perhaps most importantly, gaps between words in writing. Given the richness of these information sources, L2 learners are unlikely to rely on the computation of co-occurring elements – such as determiners and nouns – as much as infant L1 learners. As a consequence, the tight associations between determiners and nouns that emerge in the early L1 lexicon are unlikely to arise in developing L2 lexicons. There is good reason to assume, given advanced L2 learners' good performance on offline tasks manipulating gender, that representations of nouns including information on gender-class membership can be established in the L2 lexicon. Crucially, however, the associations between nouns and gender class information are unlikely to attain the same strength in L2 as in L1 lexicons, due to L2 learners' reliance on cues other than co-occurrence relations during word learning.

In sum, we propose a difference between native and non-native speakers with regard to the lexical representation of grammatical gender. Specifically, we suggest that the source of this difference lies in the strength of associations between nouns and what can be instantiated as gender nodes in the mental lexicon. This quantitative, rather than qualitative, difference between L1 and L2 manifests itself in slower retrieval of gender information in real-time L2 use, leading to occasional errors of gender assignment in production (as in Exp. 2), and less effective use of gender cues in online comprehension (Exp. 3).

The proposed account also yields a possible explanation for the somewhat surprising finding from Experiment 3, where L2 learners appeared to use gender-marking on the determiner as a predictive cue with newly learned novel nouns, but not with familiar nouns. Note that the novel nouns in Experiment 3 had to be learned solely on the basis of exposure to these nouns, preceded by gender-marked (indefinite) determiners, in spoken language during the teaching trials. No metalinguistic or written cues regarding gender class were available, forcing participants to rely on distributional information, i.e., gender-marking on the determiner, alone. This learning scenario simulates the word learning context of the infant L1 learner more so than the learning context typically encountered by the literate adult L2 learner. We thus speculate that as a consequence of participants in this experiment being forced to rely on distributional cues, their newly created lexical representations for these novel nouns contained stronger noun-gender associations than the familiar nouns, presumably learned with the aid of other cues, thus allowing for faster access to gender information. To further investigate this proposal, future work is needed to systematically manipulate the cues available during word learning (see Arnon and Ramscar (2009) for recent experimental evidence consistent with such an account).

References

- Arnon, Inbal & Michael Ramscar. 2009. Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Amsterdam.
- Carroll, Susanne. 1989. Second-language acquisition and the computational paradigm. *Language Learning*, 39, 535-594.
- Dewaele, Jean-Marc & Daniel Véronique. 2001. Gender assignment and gender agreement in advanced French interlanguage: a crosssectional study. *Bilingualism: Language and Cognition*, 4, 275-297.
- Fernald, Anne, Renate Zangl, Ana Luz Portillo & Virginia A. Marchman. 2008. Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental Psycholinguistics: On-line Methods in Children's Language Processing* (pp. 97-135). Amsterdam: John Benjamins.
- Gillon Dowens, Margaret, Marta Vergara, Horacio A. Barber & Manuel Carreiras. 2010. Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22, 1870-1887.
- Guillelmon, Delphine & François Grosjean. 2001. The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition*, 29, 503-511.
- Harris, James. 1991. The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 27-62.
- Lew-Williams, Casey & Anne Fernald. 2007. Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18, 193-198.
- Lew-Williams, Casey & Anne Fernald. 2010. Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63, 447-464.
- Montrul, Silvina, Rebecca Foote & Silvia Perpiñán. 2008. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58, 503-553.
- Pearson. 2009. Versant Spanish Test: Test description and validation summary. Pearson Knowledge Technologies, Palo Alto, California. Available online at <https://www.ordinate.com/technology/VersantSpanishTestValidation.pdf>.
- Pelucchi, Bruna, Jessica F. Hay & Jenny R. Saffran. 2009. Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113, 244-247.
- Prévost, Philippe & Lydia White. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103-133.
- Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Sagarra, Nuria & Julia Herschensohn. 2010. The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120, 2022-2039.
- Schriefers, Herbert & Jörg D. Jescheniak. 1999. Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, 28, 575-600.
- Tokowicz, Natasha & Brian MacWhinney. 2005. Implicit and explicit measures of sensitivity to violations in second language grammar: An Event-Related Potential investigation. *Studies in Second Language Acquisition*, 27, 173-204.
- White, Lydia, Elena Valenzuela, Martyna Kozłowska-MacGregor & Yan-Kit I. Leung. 2004. Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25, 105-133.