

Article

The Role of Language Experience in the Acquisition of Spanish Gender Agreement: A Study with Nonce Nouns

Silvina Montrul ^{1,2,*} , Sara Ann Mason ^{1,2} and Andrew Armstrong ^{1,2}

¹ Department of Spanish and Portuguese, College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; agarmst2@illinois.edu (A.A.)

² Department of Linguistics, College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

* Correspondence: montrul@illinois.edu

Abstract: Why is learning the gender of nouns so difficult for some bilinguals? We test the hypothesis that different language learning backgrounds or life experience with Spanish determine how learners follow different morphosyntactic cues for gender assignment in Spanish by testing learners with early and late language experience in an experiment with invented nouns. A total of 44 monolingually raised native speakers, 44 heritage speakers, and 44 L2 learners of Spanish were trained to learn 24 nonce words in Spanish presented in four input conditions that manipulated the number and type of cues to gender marking (determiner, word marker, adjective). After the learning sessions, the participants completed a word naming task, an elicited production task, and a debriefing questionnaire. The L2 learners were different than native speakers and heritage speakers in learning nonce nouns. They used morphosyntactic cues differently, relying on adjectives as their most-used strategy to assign gender, unlike native speakers and heritage speakers who used all cues. Our findings confirm processing differences between L2 learners and heritage speakers and suggest language learning background determines how learners discover reliable morphosyntactic cues to the gender of nouns in the input.

Keywords: lexical gender assignment; gender agreement; Spanish; nonce words; native speakers; heritage speakers second language acquisition; oral production



Citation: Montrul, Silvina, Sara Ann Mason, and Andrew Armstrong. 2024. The Role of Language Experience in the Acquisition of Spanish Gender Agreement: A Study with Nonce Nouns. *Languages* 9: 45. <https://doi.org/10.3390/languages9020045>

Academic Editors: Alejandro Cuza and Jeanine Treffers-Daller

Received: 8 September 2023

Revised: 17 January 2024

Accepted: 18 January 2024

Published: 26 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many languages have a gender classification system for nouns (Corbett 1991). Once typically developing native speakers of gendered languages acquire gender agreement in their language, they tend to produce it very accurately, with the occasional exception, perhaps, of some irregular words or very infrequent words (Carroll 1989; Mariscal 2009). In stark contrast, gender agreement in noun phrases is very difficult for second language (L2) learners, and more so if their first language (L1) does not classify nouns on the basis of gender (Carroll 1989, 1999; Franceschina 2005; Sabourin and Stowe 2008). Particularly in oral production, gender errors tend to persist even in advanced learners and near-native speakers (Franceschina 2001; Grüter et al. 2012). However, in untimed written production, judgment, and comprehension tasks, intermediate and advanced proficiency L2 learners can perform very accurately on gender agreement with nouns that have transparent endings for gender, as in Spanish (White et al. 2004). Interestingly, heritage language speakers of Spanish show similar gender agreement patterns to L2 learners (Alarcón 2011; Foote 2014, 2015), but unlike L2 learners, these early bilinguals tend to show higher error rates in written tasks (and tasks that rely on metalinguistic knowledge) than in oral production tasks (Bowles 2011; Montrul et al. 2008). Since many L2 learners typically start learning the language in a classroom setting around or after puberty (late) whereas heritage speakers learn

their language naturalistically in the home beginning at birth (early), their different performance in written and oral tasks is likely related to their age of acquisition and different language learning experiences with the target language.

Why is the task of learning gender so difficult for some bilinguals? Perhaps because there are at least two dimensions of gender acquisition, one lexical and one syntactic. Assuming that language learning operates on rich levels of linguistic representations (phonology, morphology, syntax, semantics), learning gender requires one to first assign gender features to newly acquired nouns from parsing the input (Carroll 1989). Once stored in the mental lexicon with its assigned gender, nouns must then be retrieved to use in noun phrases (or DPs), and once inserted in the syntax, the speaker must compute gender agreement by marking modifiers such as determiners, adjectives, and pronouns in phrases and sentences. Native speakers and L2 learners of French (Carroll 1989, 1999; Hawkins and Franceschina 2004) and Spanish (Franceschina 2001, 2005) have been claimed to parse the input and build linguistic representations for gender differently, especially if the native language of the L2 learner does not represent gender as an abstract category in the grammar, like English. Much research on Spanish has investigated whether the differences between native speakers and L2 learners lie in the syntactic mechanisms for gender agreement (Alemán Bañón et al. 2014; Franceschina 2001; White et al. 2004). More recent studies have found that most errors made by L2 learners (Alarcón 2011; Hopp 2013, 2016; Grüter et al. 2012; Lemhöfer et al. 2014; Martoccio 2022) and heritage speakers (Alarcón 2011; Montrul et al. 2013) can be attributed to incorrect lexical assignment rather than to the maloperation of the syntactic mechanisms for agreement. This suggests that noun classification in the lexicon is an important locus of difficulty for gender agreement in some bilinguals (Hopp 2015). However, there is very little understanding of how this may happen at the initial stages of acquisition. The first aim of this study, therefore, is to examine how different learners go about classifying newly encountered nouns at the initial stages of learning by focusing specifically on what morphosyntactic cues from the input (i.e., word markers, determiners, adjectives) learners rely on.

Functionalist approaches to language learning generally emphasize the importance of structural cues in the input to guide lexical expectations through an association network during learning (MacWhinney 1987). Carroll's (1989) information processing proposal, which assumes generative grammar, demonstrated more specifically how gender assignment cues for French nouns are found in the syntax from specific words that agree with the noun; namely, determiners and adjectives. Spanish is very much like French, although it also offers another robust morphological cue when the nouns end in the vowels *-o* (masculine word marker) or *-a* (feminine word marker). We assume a syntactic account of gender agreement. Once nouns are assigned a gender feature and stored in the learner's mental lexicon, they can then be retrieved and inserted in the syntax of a noun phrase (Det-N-Adj), for example. The gender feature of nouns triggers agreement with other elements of the noun phrase (determiner, adjective) during production.

The second aim of this study is to investigate how different language learning experiences affect the initial learning of the gender of nonce words in Spanish. We test the hypothesis, based on previous studies (Bordag et al. 2006; Caffarra et al. 2017; Lemmerth and Hopp 2019), that the different language learning backgrounds or life experience with Spanish of monolingually raised native speakers, L2 learners, and heritage speakers may guide and determine how learners discover reliable cues to the gender of nouns in the input. Carroll (1989, p. 596) argued that determiners in French have a privileged status as the main cue for gender assignment by French-speaking children. Similarly, processing studies using eye-tracking and ERPs indicate that the determiner is the most robust cue to gender marking in Spanish for child and adult native speakers (Lew-Williams and Fernald 2007b). Learning the first language from auditory input in childhood requires the detection and segmentation of determiners and nouns from the acoustic signal (Lew-Williams and Fernald 2007b; Wicha et al. 2004), leading child learners and adult native speakers to use the grammatical information in determiners to predict the noun referent based on the gender of the determiner (Lew-Williams and Fernald 2007a). Using the same paradigm, Fuchs (2022)

also showed that Spanish heritage speakers use the gender of determiners predictively to anticipate nouns.

Adult L2 learners, who are exposed to auditory and visual input from the onset of L2 acquisition, may not engage in predictive processing of gender through determiners because they often learn words in isolation and tend to focus on morphophonology and the ending of the noun (Arnon and Ramscar 2012; Grüter et al. 2012; Gillon Dowens et al. 2010; Martoccio 2022), at least at intermediate stages of development (Hopp 2013, 2016; Morales et al. 2016). In a study of gender assignment with low- and high-frequency nouns in Spanish, Martoccio (2022) did find that L2 learners of Spanish paid attention to the determiner but used this strategy significantly less than native speakers. Montrul et al. (2013) and Montrul et al. (2014) found that heritage speakers, who as bilingual native speakers also learn their language from auditory input, were more native-like in processing and producing gender agreement in Spanish than proficiency-matched L2 learners. Foote (2014) found that L2 learners were able to use the determiner as a cue but only in conditions that presented gender-incongruent determiners and nouns. Following Gollan et al.'s (2008) weaker links hypothesis, Montrul et al. (2013) suggested that the lexical links between determiners and nouns seem to be stronger in heritage speakers than in L2 learners, probably because heritage speakers have been using Spanish longer than the L2 learners and the quality of the input differs (home context, spoken rather than written, etc.). Based on studies of monolinguals and bilinguals using picture naming with real words, the weaker links hypothesis attributes these effects to the frequency of words.

In this study, we assess the extent to which determiners are crucial cues for gender assignment in Spanish, compared to adjectives and word markers, especially for native speakers and heritage speakers, who may have developed different processing strategies from those used by L2 learners of Spanish due to an early age of acquisition of Spanish and more experience with gendered nouns over their lifetime. However, understanding how learners with different linguistic histories may use different mechanisms for discovering the gender of nouns requires controlling for experience with nouns. Therefore, we conducted an experiment with nonce nouns presented in different learning conditions to control for previous knowledge and frequency of exposure. We report the results of a naming task, an oral production task eliciting determiner + nonce noun + adjective phrases, and a debriefing questionnaire. Before presenting the details of the experiment, we describe how gender agreement is realized in Spanish and how it is acquired by monolinguals and bilinguals.

2. Spanish Gender Acquisition in Monolinguals and Bilinguals

Spanish nouns are classified as feminine or masculine. According to Teschner and Russell (1984), 99.9% of nouns that end in the vowel *-o*, such as *libro* “book”, *cuento* “story”, *banco* “bank, stool”, are masculine and 96.3% of nouns that end in the vowel *-a*, as in *cama* “bed”, *hoja* “leaf”, *taza* “cup” are feminine. Although the word markers *-a* and *-o* are not actual inflectional morphemes (Harris 1991), they are sometimes used productively to derive the gender of new nouns and to distinguish gender in animate nouns, such as *perro* “male dog” vs. *perra* “female dog”, *niño* “boy” vs. *niña* “girl”, and *tío* “uncle” vs. *tía* “aunt” (Falk 1978).

As in all languages, there are exceptions to these general tendencies, and Spanish also has many non-transparent nouns that end in other vowels and consonants, as well as exceptional masculine nouns that end in *-a* and exceptional feminine nouns that end in *-o*, as shown in Table 1.

Table 1. Morphological transparency of Spanish inanimate nouns based on noun-ending.

	Transparent	Non-Transparent		Exceptions
Masculine	–a/–o <i>libro</i> “book” <i>techo</i> “roof”	–e <i>puente</i> “bridge” <i>diente</i> “tooth”	–consonant <i>motor</i> “engine” <i>papel</i> “paper”	–a/–o <i>planeta</i> “planet” <i>cometa</i> “comet”
Feminine	<i>libra</i> “pound” <i>casa</i> “house”	<i>noche</i> “night” <i>nube</i> “cloud”	<i>piel</i> “skin” <i>nariz</i> “nose”	<i>mano</i> “hand” <i>foto</i> “photo”

Of the 1000 most common nouns in Spanish, 24 end in the vowel –e, and these words can be masculine or feminine (Teschner and Russell 1984, p. 124). There are also many animate nouns that end in –e and refer to both genders (*el/la estudiante* “the student”, *el/la agente* “the agent”, *el/la cantante* “the singer”). Many nouns ending in the consonants –r, –l, –z, –n and –d can also be feminine or masculine. Finally, the masculine nouns *problema* “problem”, *cometa* “comet”, *planeta* “planet” and the feminine nouns *mano* “hand”, *moto* “motorcycle”, *foto* “photo” are exceptions.

Although the classification of nouns into masculine or feminine gender is lexical, gender is manifested syntactically through agreement in the noun phrase and in the verb phrase. Nouns come lexically determined with a gender feature (Carroll 1999; Carstens 2000). The gender feature is an interpretable feature in nouns and an uninterpretable (formal) feature in determiners and adjectives, which must be checked through agreement (Chomsky 1995). Nouns check their gender features in specifier–head (for noun–adjective concord) and head–head (for determiner–noun concord) relations. Thus, gender agreement is an operation handled by the syntax. Examples (1a) and (2a) show transparent –o masculine and –a feminine-ending nouns, while examples (1b) and (2b) show non-transparent feminine and masculine nouns ending in a consonant.

- (1) a. La taza roj-a *feminine transparent*
 the-FEM cup-FEM red-FEM
 “The red cup”
- b. Una flor abiert-a *feminine non-transparent*
 a-FEM flor-FEM open-FEM
 “An open flower”
- (2) a. El techo roj-o *masculine transparent*
 the-MASC roof-MASC red-MASC
 “The red roof”
- b. Un árbol caíd-o *masculine non-transparent*
 a-MASC tree-MASC fallen-MASC
 “The fallen tree”

Definite (*el, la* “the”) and indefinite (*un, una* “a”) determiners in Spanish carry gender information, whereas possessive pronouns (*su, sus* “his/hers/their”) do not. If we compare the number of cues to gender in these noun phrases, the phrases with transparent nouns (1a, 2a) have three cues: one on the determiner, one on the word marker of the noun, and one on the adjective, whereas the noun phrases in (1b) and (2b) only have two cues: one on the determiner and one on the adjective; the word marker is ambiguous for gender. The examples in (3) include transparent and non-transparent nouns in noun phrases with possessive determiners and adjectives that are gender neutral.¹ Thus, in the phrase in (3a) the only cue to gender is the word marker on the noun. The phrase in (3b) has no cues to gender.

- (3) a. Sus ojos/miradas tristes *transparent noun*
 his/her eyes-MASC/looks-FEM sad
- b. Su piel/cutis suave *non-transparent noun*
 his/her skin-FEM/skin-MASC soft
- “His/her soft skin”

Non-transparent nouns present a particular challenge to language learners since the morphophonological form of the other items in the phrase that agree with the noun (i.e., determiners and adjectives) are key to discovering their gender assignment. Several studies of different languages that present these characteristics have shown that even native speakers take longer to learn and process gender assignment and agreement with non-transparent nouns (Bates et al. 1995; Taraban and Kempe 1999; Taraban and Roark 1996). When Spanish-speaking children make gender errors, these occur most often with non-transparent nouns (Hernández Pina 1984; Mariscal 2009), and the same pattern is attested in L1 attrition (Montrul 2011), heritage language acquisition, and L2 acquisition (Alemán Bañón et al. 2014; Alarcón 2011; Montrul et al. 2008, 2013, 2014). When non-transparent nouns contain a diminutive morpheme, however, the diminutive regularizes the word by making it transparent: e.g., *el elefante* “the elephant” becomes *el elefantito* “the little elephant”, and *la nariz* “the nose” *la naricita* “the little nose” (Montrul et al. 2013). The abundance of diminutives in child language helps children learn the gender of non-transparent nouns by revealing the transparent word ending cue to gender (Kempe and Brooks 2001; Savickienė and Dressler 2007).

When learners make gender errors in Spanish, masculine gender tends to be overgeneralized to feminine nouns more than the reverse (Faber 2017; McCarthy 2008; Montrul et al. 2008, 2013; White et al. 2004), and this pattern is consistent with the view that masculine gender is the default form in Spanish (Domínguez et al. 1999; Harris 1991; McCarthy 2008). There is also an asymmetry between gender agreement errors with determiners and adjectives in noun phrases. Some studies report more errors on adjectives than on determiners in L2 Spanish (Alarcón 2011; White et al. 2004).

Persistent difficulty acquiring gender in different syntactic contexts led to critical-period-based proposals that the problem with gender agreement in L2 acquisition results from maturational constraints associated with an inability to acquire the syntactic features necessary for syntactic computations during gender agreement (Carroll 1989; Hawkins and Franceschina 2004; Franceschina 2001). However, recent studies using elicited oral production and auditory processing have found that most gender errors in the L2 learners tested at the intermediate level and above were related to lexical assignment rather than to syntactic agreement (Grüter et al. 2012; Hopp 2013, 2015; Montrul et al. 2013). Hopp (2013, 2016) found that variability and instability in lexical gender assignment affects the processing of gender agreement. Consider the phrases in (4) with the non-transparent masculine noun *coche* “car”.

(4)	a.	el-MASC coche-MASC rojo-MASC	target
	b.	*la-FEM coche-MASC roja-FEM	assignment error
	c.	*el-MASC coche-MASC roja-FEM	agreement error
	d.	*la-FEM coche-MASC? / FEM? rojo-MASC	ambiguous error

If the target phrase is as in (4a), *el coche rojo*, and a learner says **la coche roja* (“the-FEM **car** red-FEM”) as in (4b), where the determiner and the adjective match with each other on gender but not with the noun, it is likely that the learner has misclassified the head noun *coche* as feminine in the mental lexicon and then performed agreement between determiner and adjective correctly in the syntax. If the learner says **el coche roja* (“the-MASC **car** red-FEM”) instead, as in (4c), where the determiner and the noun match but the adjective does not, one can assume that the error is syntactic rather than lexical misclassification; that is, in the gender agreement rule between noun, determiner, and adjective in the syntax. Although agreement with adjectives is generally more problematic than with determiners, especially with adjectives that are nonadjacent to the nouns that they modify (Foote 2015; Keating 2009), some errors are not easily classified as assignment or agreement errors and may arise from processing limitations as a function of syntactic distance (Keating 2009).

In the language acquisition literature, the gender of the determiner is often taken as evidence for lexical assignment of gender in French (Carroll 1989) and Spanish (Lew-Williams and Fernald 2010). Other potential errors include **la coche rojo* (la-FEM **car** red-MASC), as in (4d), where the noun and the adjective match, but the determiner and the adjective do not match. These are ambiguous between assignment or agreement errors because the

gender of the determiner may indicate incorrect lexical assignment, but since the determiner and the adjective do not match either, this would also be lack of concord. In this study, we consider Det-N concord an instance of assignment, as in (4b). Grüter et al. (2012) and Montrul et al. (2013) found that about 80% of gender errors were of the lexical assignment type and the rest were largely syntactic errors. Ambiguous errors like (4d) are quite infrequent. A reason why gender agreement errors with determiners are fewer than with adjectives may have to do with the prominent status of the determiner for gender assignment in Spanish. Studies of L1 acquisition (Lleó 1998; López Ornat 1997; Pérez-Pereira 1991) and of language processing (Afonso et al. 2014) have shown that the prenominal determiner seems to be the main morphosyntactic cue for predicting and assigning gender to nouns in Spanish. When children in the one-word stage produce nouns in Spanish, these are usually preceded by a 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 utterances suggest that there is a very tight association between determiners and nouns in the lexicon, at least for native speakers. Guided by prosodic cues or transitional probabilities, the child later segments the auditory stream for these utterances into determiner and noun. Lew-Williams and Fernald (2007a, 2007b, 2010) and Grüter et al. (2012) suggest that noun-gender associations are strong in the Spanish L1 lexicon as a consequence of early chunking and subsequent speech segmentation. In their studies of sentence processing using the Visual World paradigm, Lew-Williams and Fernald (2007a, 2007b, 2010) found that both adult native Spanish speakers and 3–4-year-old Spanish-speaking children used gender information in determiners to predict the referents of nouns during spoken word recognition on different-gender trials. In recent processing experiments, Fuchs (2021, 2022) found that heritage Spanish speakers, like the native speaker controls, used gender information on definite articles (masculine and feminine alike) to predict upcoming nouns during online auditory sentence comprehension.

The importance of exposure to noun phrases for gender learning was examined by Arnon and Ramscar (2012) with an artificial language experiment using auditory input, where they contrasted nouns with and without determiners. Although the study sought to determine if exposure to single nouns “blocked” learning of a variable (gender-marked) determiner, it did yield some information about the role of determiners in gender acquisition. One group of learners was first exposed to article + noun sequences in whole sentences, and then to nouns without determiners (noun-label only). The other group was exposed to nouns without determiners first and then to article + noun sequences. By the end of the experiment, both groups had received the same input but in different orders. This allowed the researchers to manipulate the size of the initial units learners were exposed to while keeping frequency of exposure constant. They found that the participants in the article + noun sequence condition were both more likely to choose the sentence with the correct article in a forced-choice task and more likely to produce the appropriate article for a given noun in a production task. This finding suggests that initial exposure to whole noun phrases, as in the naturalistic input that children are exposed to when acquiring the L1, enhances learning the correct gender-marked determiner when compared to initial exposure to single nouns.

Studies with nonce nouns or with artificial languages (e.g., Morgan-Short et al. 2010) are well suited to investigate the cues to gender assignment because they simulate a typical learning situation encountered by learners. They also control for experience with real words as a function of frequency and exposure. Although the L2 participants in Lew-Williams and Fernald’s (2010) study were unable to use gender in the determiner predictively with real nouns like the native speakers in Experiment 1, their Experiments 2 and 3 included the teaching and testing of four nonce nouns with transparent endings (two feminine and two masculine). On different-gender trials, when the article is potentially informative and the noun has a transparent ending, the L2 learners were able to use the determiner predictively like the native speakers to pair sentences to objects in Experiment 2 (definite determiners in training and test trials) but not in Experiment 3 (indefinite determiners in training and

definite determiners in trials). Thus, when exposed to the nonce word/object pairs the same number of times (Experiment 2), native and non-native speakers were successful at this task, compared to their performance on the same experiments with real words and with different determiners.

Faber (2017) examined how L1 and L2 Spanish speakers assign, retain, and process gender with nonce nouns. The L2 learners were native speakers of English (a language with no gender) and native speakers of Brazilian Portuguese (a language with gender). One of the experiments presented short stories auditorily. The stories introduced two nonce items appearing once with an indefinite determiner and once with an adjective with gender agreement. Participants were required to respond to a question about each story, by producing the nonce noun and an agreeing adjective. Faber found that the three groups assigned gender differently: the L1 Spanish and L1 Brazilian Portuguese speakers were guided by syntactic cues, i.e., the determiner, to assign gender to nonce words in Spanish. The L1 English speakers used morphophonological cues on the noun and were more inaccurate on determiners than the other two groups. The Brazilian learners of Spanish, like the native speakers of Spanish, had experience assigning gender to nouns since childhood, and perhaps paying attention to the prenominal determiner as a syntactic cue, whereas the English-speaking learners of Spanish did not bring childhood experience processing gender. Contradictory findings with the L2 learners by Lew-Williams and Fernald (2010) and Faber (2017) could be related to the different materials and methodologies, including the words. Lew-Williams and Fernald (2010) used transparent *-o* and *-a* ending words only, while Faber included transparent endings and the non-transparent *-e* ending (as in *noche_F* “night”, *leche_F* “milk”, *coche_M* “car”). Therefore, noun ending transparency is important to manipulate in studies of nonce nouns.

To summarize, research suggests that the main locus of difficulty with gender production in Spanish may lie in lexical assignment more than in syntactic agreement, especially for L2 learners at the intermediate level. Difficulty with gender assignment could be related to the fact that L2 learners may not use gender in determiners as a cue for gender assignment as much as native speakers, and perhaps because of this different strategy they are more inaccurate with gender agreement than native speakers overall. Instructed L2 learners usually learn Spanish nouns in auditory and written form, either in isolation (word lists) or in a phrase or sentence in context. Since visual input is already segmented into words through blank spaces, L2 learners may not establish a close connection between the determiner and the noun like native speakers who learned to classify nouns from non-segmented auditory input. Since determiner–noun associations may not be strong for L2 learners (Grüter et al. 2012), they might pay more attention to the word marker (if the word ends in *-o* or *-a*) or to the postnominal adjective instead. Heritage speakers of Spanish have also been found to be inaccurate with gender (Montrul et al. 2008), but it is unlikely that the source of their difficulty could be same as in L2 learners because both groups have different language learning experiences and may process input differently. No studies to date have examined how heritage speakers learn nonce words, or which cues they may follow during the initial stages of learning to assign gender to nouns, so this study contributes to filling this important gap. The findings from recent studies (Fuchs 2022) lead us to predict that heritage speakers would pattern with monolingually-raised native speakers of Spanish and will use determiners, perhaps more than adjectives, as a cue to gender classification when learning nonce nouns in Spanish.

3. The Study

We designed an experiment with nonce nouns to address the following research questions:

- (1) Which morphosyntactic cues are used by native speakers, heritage speakers, and L2 learners for gender assignment in Spanish with transparent and non-transparent nouns? Does the syntactic position of the cue (det vs. adj) or the number of cues matter?

- (2) Do learners with different language learning experience use different morphosyntactic cues, or use cues differently, when assigning gender to nonce nouns in Spanish?

Noun frequency has been recently shown to play a role in Spanish heritage speakers' accuracy on gender agreement (Hur et al. 2020). Using nonce words allows us to control for the participants' experience with the new words during the experiment while focusing directly on lexical gender assignment and agreement mechanisms independent of word frequency. We created four training conditions manipulating the transparency (transparent, non-transparent) of noun endings, determiners and adjectives, the type of syntactic cue (determiner, adjective) and the number of gender cues in the input sentences (from zero to three cues).

Native speakers of Spanish and heritage speakers of Spanish were exposed to Spanish since birth and acquired words in Spanish by parsing auditory input. In their childhood, they segmented chunks into determiners and nouns. L2 learners of Spanish started acquisition later in life in an instructed setting and learned words auditorily and visually. Visual input is already segmented for determiners, nouns, and adjectives. Since we strictly manipulated the input experience during the experiment in the training sessions, the null hypothesis would be that there are no differences among the groups. Alternatively, if life experience with a particular language determines how speakers process input when learning new nouns in the language (Grüter et al. 2012; Hopp 2016), we expect native and heritage speakers of Spanish to produce gender agreement more accurately after input conditions with informative gender cues on determiners versus input conditions with informative gender cues on adjectives. Based on the patterns of gender accuracy reported in the literature (Franceschina 2001; Holmes and de la Bâtie 1999), L2 learners of Spanish are expected to use morphophonological cues on word endings and/or the adjective more than the determiner as more informative cues to gender in Spanish.

3.1. Method

Participants

There were three groups of adult learners: native speakers of Spanish, L2 learners of Spanish, and Spanish heritage speakers. Our goal was to recruit a minimum of 20 participants per group to complete the word-learning session in each of the four input conditions, which would be a minimum of 80 participants per group. But because we needed to also match heritage speakers and L2 learners on proficiency in each input condition, we made the decision to recruit a minimum of 40 participants per group instead and assign them to two input conditions each, testing them in two lab sessions at least a week apart.² A total of 145 individuals volunteered to participate. All participants completed a language background questionnaire with questions about age, sex, country of origin, place of current residence, childhood language(s), language of education, etc. Thirteen participants were eliminated for not meeting some of the participation criteria as indicated by their questionnaire responses or for issues with the tasks.³ The total number of participants retained is summarized in Table 2.

Table 2. Participants' mean age at testing, mean age of first exposure to Spanish (AoA), and mean Spanish proficiency scores (ranges in parentheses).

	Native Speakers (n = 44)	Heritage Speakers (n = 44)	L2 Learners (n = 44)
Age	29.89 (18–55)	21.27 (18–43)	25.89 (18–53)
AoA Spanish	birth	birth	10.67 (4.5–16) SD = 3.44
Spanish proficiency score	--	33.48 (13–50) SD = 10.38	39.61 (19–50) SD = 8.21

The native speakers (22 male, 22 female) were born,⁴ raised, and educated in Spanish-speaking countries representing all major regions of the Spanish-speaking world, and were

living in the United States at the time of testing. They were between the ages of 18 and 55, mean age 29.64 years. L2 speakers (10 male, 34 female) were between the ages of 18 and 53, mean age 25.89 years. All reported being born in the United States or Canada; some were heritage speakers of languages other than Spanish, but these were languages with no gender (Chinese, Korean). Heritage speakers (10 male, 34 female) were between the ages of 18 and 43, mean age 21.27 years. The heritage speakers were university students of Mexican descent. Six reported being born abroad and arriving in the U.S. between the ages of 6 months and 7 years; the rest were born in the United States or Canada. All reported early childhood exposure to Spanish and use of Spanish with their families. The L2 learners were all English native speakers enrolled in intermediate- and advanced-level Spanish classes at the university. The mean age of first exposure and acquisition of Spanish is in Table 2. The mean age at the start of formal study of Spanish was 13.4, $SD = 6.55$ (range 7–50). At least 15 participants started learning Spanish as a second language at school between the ages of 7–10, the rest after age 11.

The heritage speakers and the L2 learners completed a commonly used measure of proficiency used in L2 Spanish research (same tests used in Montrul et al. 2008 and discussed in Montrul et al. 2013). The L2 learners' mean proficiency was 33.48 ($SD = 10.38$, range 13–50) and the heritage speakers' mean was 39.61 ($SD = 8.21$, range 19–50). An ANOVA indicated that L2 and heritage speakers had significantly different proficiency test scores overall ($F(1,158) = 15.35, p < 0.001$).

3.2. Materials and Design

The current study consisted of a background questionnaire and proficiency test (described above), a gender proficiency test with real nouns, two separate learning sessions with nonce nouns, a nonce word picture naming task after each learning session, an oral production task with nonce words after each learning session, and a debriefing questionnaire at the end of the study.

3.3. Gender Proficiency Pre-Test with Real Nouns

Before assessing how different learners assign gender to newly encountered nonce nouns, it was important to obtain a baseline measure of their general knowledge of gender with real nouns, with which learners already have experience. This would allow us to establish two things: (1) whether general proficiency and gender proficiency are related, and (2) whether language experience with particular words affects gender accuracy in different types of speakers (e.g., the weaker links hypothesis), especially when we look at the results of nonce nouns.

In this task, a noun and an adjective (either in masculine or feminine form irrespective of the gender of the noun) were presented in all caps in the center of the screen, as in Figure 1.

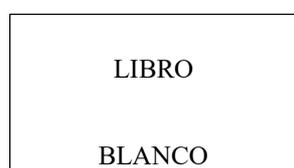


Figure 1. Sample screen presentation of items in the gender proficiency pre-test with real nouns.

The participants had five seconds to verbalize a complete NP minimally containing the correct definite determiner, the noun, and the correct adjective form (i.e., *el libro (es) blanco* “the book is white”). There were 24 nouns (*boca* “mouth”, *leche* “milk”, *mercado* “market”, etc.), 4 in each of six categories (2 genders and 3 endings, transparent (–a or –o), –e, –consonant). The instructions and the training examples made it clear that sometimes the form of the adjective had to be changed to agree in gender with the noun. A series of one-way ANOVAs confirmed that these nouns were normed by gender and transparency

on frequency, length in letters, and length in syllables.⁵ All adjectives were transparent (e.g., *raro/a* “rare”, *largo/a* “long”, *famoso/a* “famous”, etc.). (See https://osf.io/j6p5b/?view_only=9794107ee2884da99aaf8623d76537ca, accessed on 11 January 2024). An anonymous reviewer questioned the fact that in this task the adjectives appear in two different forms, sometimes matching and sometimes mismatching the noun in gender, and suggested that the task used by Montrul et al. (2008), in which participants themselves provided the adjective, would have been a better methodology. We chose this task format, following Bates et al. (1996), Guillelmon and Grosjean (2001), and Montrul et al. (2014), because it was important to maximize the probability of the participants producing full NPs with adjectives, since in the Montrul et al. (2008) study 22.7% of the native speakers, 36.24% of the heritage speakers, and 41.67% of the L2 learners did not produce adjectives. Presenting all adjectives in masculine form in the current task would lead to overextensions of masculine agreement to feminine nouns, whereas presenting adjectives in gender-matching form with the target noun on every trial would be giving away the response, especially to the L2 learners whose gender accuracy we were assessing. Therefore, to avoid biasing responses toward either gender, half of the items in each category appeared with a gender-matching adjective and half appeared with an adjective in the opposite gender (mismatching adjective). The consequence of this choice is that the learners would have to override the non-matching presentation in half of the items to produce correct agreement. We acknowledge that failure by the participants to override the gender of the mismatching adjectives would inflate the rate of gender assignment errors, and we consider this fact in our analyses.

3.4. Nonce Words

We created 24 nonce words in Spanish by randomly combining phonotactically acceptable CV-CV(-C) sequences (Table 3). We consulted with several native speakers during the creation of the task to make sure that the nonce words were not very similar to real words; however, complete control of all possibilities was not possible. As the results of our debriefing questionnaire show, some native speakers did relate some of the nonce words to real words in their variety of Spanish. Nonce nouns varied both by transparency and gender. Transparent nouns ended in *-a* or *-o*, and non-transparent nouns ended in *-e* or a consonant (*-l*, *-n* or *-r*). There were 8 transparent nouns (4 masculine, 4 feminine), 8 nouns ending in *-e* (4 masculine, 4 feminine), and 8 nouns ending in a consonant (4 masculine, 4 feminine).

Table 3. Nonce nouns invented for the experiment.

Transparent Ending		Non-Transparent Ending			
-o	-a	-e		Consonant	
<i>masculine</i>	<i>feminine</i>	<i>masculine</i>	<i>feminine</i>	<i>masculine</i>	<i>feminine</i>
reco	tida	naje	mibe	nejil	bejur
lafo	nala	repe	nete	disán	rechal
nocho	jisa	jabe	nune	tufán	serul
sodo	foda	bure	sire	chelel	jerán

The nonce nouns were presented with invented objects, which were pictures created in Paint and Pixelmator, as shown in Figure 2. They were randomly assigned to an invented object using a list randomizer.

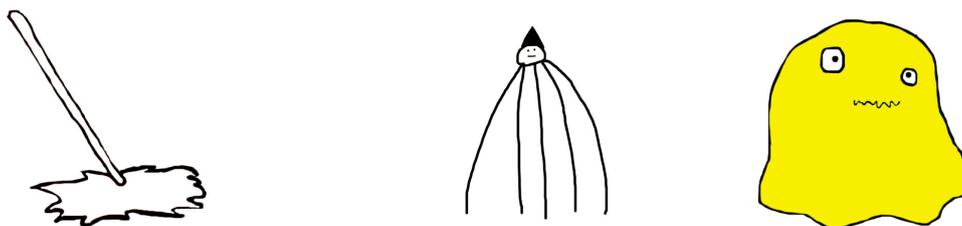


Figure 2. Sample objects used as referents of nonce nouns.

3.5. Learning Sessions

The nonce nouns were randomly divided into two lists of 12 words each, with two words from each gender and ending category presented in Table 1. During the learning sessions, the words were presented audio-visually in one of four different input conditions. The input condition refers to the sentence context (syntax) in which the nonce nouns were presented during the word-learning sessions. Input conditions varied in the type and number of gender cues, including the transparency of the nonce word ending, the transparency of the article, and the transparency of the adjective. Twelve words (list 1) appeared in input conditions 1 and 3 and the other 12 (list 2) appeared in input conditions 2 and 4 (Table 4). All participants learned the 24 words, but divided in two different learning conditions, as illustrated in Figure 3.⁶

Table 4. Input conditions during learning sessions.

Gender Cue	Transparent N	# of Cues	Non-Transparent N	# of Cues
1. no cue on determiner or adjective	su sodo gris	1	su jabe gris	0
2. cue on adjective	su reco rojo	2	su serul roja	1
3. cue on determiner	la tida azul	2	el disán azul	1
4. cue on determiner and adjective	el lafo negro	3	el chelel negro	2

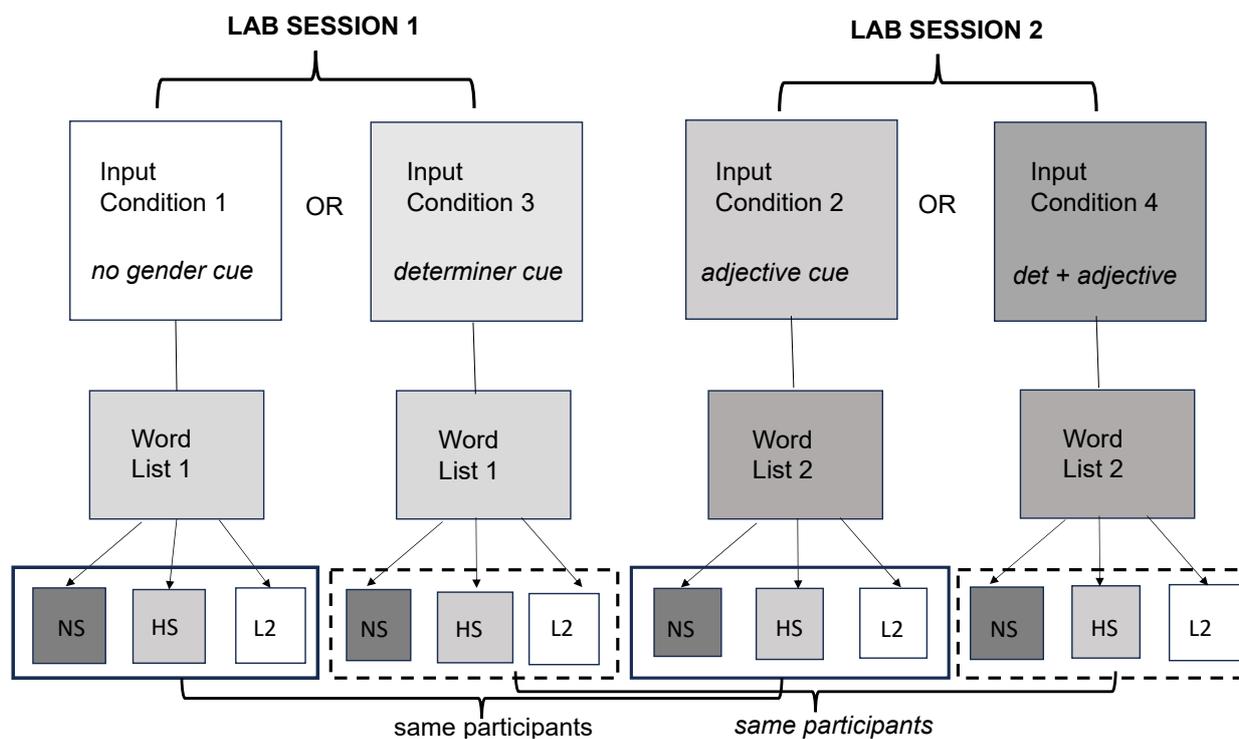


Figure 3. Overview of research design.

During each word-learning session, the input condition participants were assigned to consisted of 96 block-randomized question–answer pairs that presented each of the 12 nonce words 8 times. Preceding each question–answer pair, the nonce word was introduced embedded in a sentence with an audio-visual prompt while its object was displayed for up to 10 s (e.g., *Juan trajo su/el reco rojo* “Juan brought his/the red reco”). The nonce words were then elicited from the participant in four question-answer pairs (Table 5) used twice with each nonce word to yield the 8 presentations of each nonce word. The answers produced by the participants during the learning session finished with an NP consisting of a determiner (*su* or *el/la*, depending on the input condition), the nonce noun, and the adjective.

Table 5. Phrases illustrating/eliciting nonce nouns following each audio-visual sentential prompt in the different input conditions used during learning sessions.

Input Condition	Question	Answer
1 (no cue)	<i>¿Qué trajo Juan?</i> “What did Juan bring?”	<i>Juan trajo</i> (su + nonce noun + opaque adj.) “Juan brought (his/her + nonce noun + opaque adj.)”
2 (adjective)	<i>¿Qué ve Juan?</i> “What does Juan see?”	<i>Juan ve</i> (su + nonce noun + transparent adj.) “Juan sees (his/her + nonce noun + transparent adj.)”
3 (determiner)	<i>¿Qué trajo María?</i> “What did Maria bring?”	<i>María trajo</i> (det + nonce noun + opaque adj.) “Maria brought (the + nonce noun + opaque adj.)”
4 (determiner + adjective)	<i>¿Qué ve María?</i> “What does Maria see?”	<i>María ve</i> (det + nonce noun + transparent adj.) “Maria sees (the + nonce noun + transparent adjective)”

Each word-learning session took 20 min to complete. The number of gender cues a participant was exposed to and required to produce during the word-learning session depended on which input condition they were assigned to. The lists used in the test can be found in the OSF repository (https://osf.io/j6p5b/?view_only=9794107ee2884da99aaf8623d76537ca).

Although previous studies with nonce nouns have used auditory stimulus presentation (Arnon and Ramscar 2012), we made the decision to present the stimuli in written and auditory modalities (bimodally) to level the playing field, so to speak, especially because L2 learners perform less native-like than heritage speakers in auditory tasks (Sánchez-Walker and Montrul 2020). For both heritage and L2 speakers, this makes the input presentation more similar to typical instructed settings and makes our results potentially more applicable to a classroom setting while still manipulating the number of gender cues to reflect input differences based on varying language experiences.

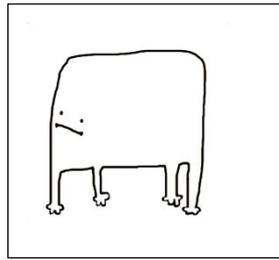
3.6. Experimental Tasks

After the word-learning session, participants completed two tasks to assess their knowledge of the nonce nouns introduced in the session: (1) a vocabulary (word naming) task that assessed their verbal memory of the nouns, and (2) an oral production task that elicited the nonce nouns in noun phrases and tested what gender participants assigned to the newly learned nouns. Finally, at the end there was a short debriefing questionnaire with five questions about what they learned.

3.7. Vocabulary (Word Naming) Task with Nonce Nouns

To make the learning sessions more meaningful, the purpose of the experiment was presented as a memory game. Retrieval of the nonce words was tested via a picture naming task. The nonce words were sorted into two blocks and randomized within blocks such that both items from the same category did not appear in consecutive trials (i.e., masculine *-e* items back-to-back). There were five practice trials using familiar Spanish words. A white image of a nonce object was presented on the computer screen (Figure 4) and participants had 5 s to produce the nonce word (i.e., sees an *apple*, says “apple”, in this case sees a *sodo*,

says “sodo”). The microphone registered the vocal responses and continued to the next item automatically.



Expected response: SODO

Figure 4. Sample stimulus in the Vocabulary Task (word naming).

3.8. Oral Production Task with Nonce Nouns

The main task to test the hypotheses of our study was an oral production task that elicited the nonce nouns in noun phrases. The purpose of this task was to see what gender participants would assign to the newly learned words in production, and whether they would make errors in gender production. This task followed the same format as the gender proficiency pre-test described above (Figure 5).

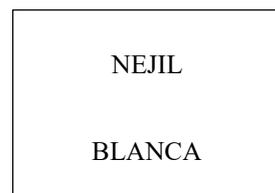


Figure 5. Sample screen presentation of items in the oral production task with real nouns.

Participants saw a nonce noun and an adjective (masculine or feminine) in all caps in the center of the screen and had five seconds to verbally produce a complete NP minimally containing a definite determiner, the noun, and the adjective (i.e., *el reco (es) rojo*). All adjectives were the gender-transparent color adjectives (*rojo* “red”, *negro* “black”, *blanco* “white”, *amarillo* “yellow”) and were assigned randomly to the different items. Each nonce noun was tested twice, once with a color adjective that matched in gender (*reco-rojo*) and once with an adjective that mismatched in gender (*reco-roja*). (In the learning session with gender cues on the adjective, the noun was presented with only one version of a gender-transparent adjective, the one assigned to the noun in the experiment). The items were sorted into two blocks and randomized within blocks. There were five practice trials using familiar Spanish words. The practice items made it clear that sometimes the form of the adjective needed to be changed.

3.9. Procedure

The background questionnaire and Spanish proficiency test were completed in the lab online before the beginning of the first treatment session. Based on their Spanish proficiency score, participants were invited to complete two lab sessions at least one week apart and were told they would be participating in a memory game during each session. They were assigned to two input conditions (input condition 1: no cue and input condition 2: adjective cue OR input condition 3: determiner cue and input condition 4: adjective and determiner cues) (Table 3). All participants learned all 24 words but in different learning conditions and sessions (Figure 2). The order of the conditions that the participants were exposed to was counterbalanced between the two sessions. The first lab visit was 90 min long; the second lasted an hour.

In the first lab session (Day 1 in Figure 6), participants completed (1) the gender proficiency test with real Spanish words, (2) the first 20-min nonce-word-learning session for one list of 12 nonce words in an assigned input condition (1 or 3), (3) the nonce-word vocabulary task, and (4) the oral production task with the 12 nonce words presented during the learning session. In the second lab session (Day 2 in Figure 6), participants completed the second nonce-word-learning session for the other list of 12 nonce words in the other input condition they were assigned to (2 or 4). For example, participants completed either a learning session in input condition 1 in the first lab visit and then a learning session in input condition 2 in the second lab visit, or a learning session in input condition 3 in the first visit and then a learning session in input condition 4 in the second visit (Figure 2). In the second lab visit, participants completed (1) the other nonce-word-learning session for a new set of 12 nonce words, (2) the nonce-word vocabulary task, and (3) the oral production task with the new nonce words learned in that session. At the end of the second lab session, participants completed the debriefing questionnaire.

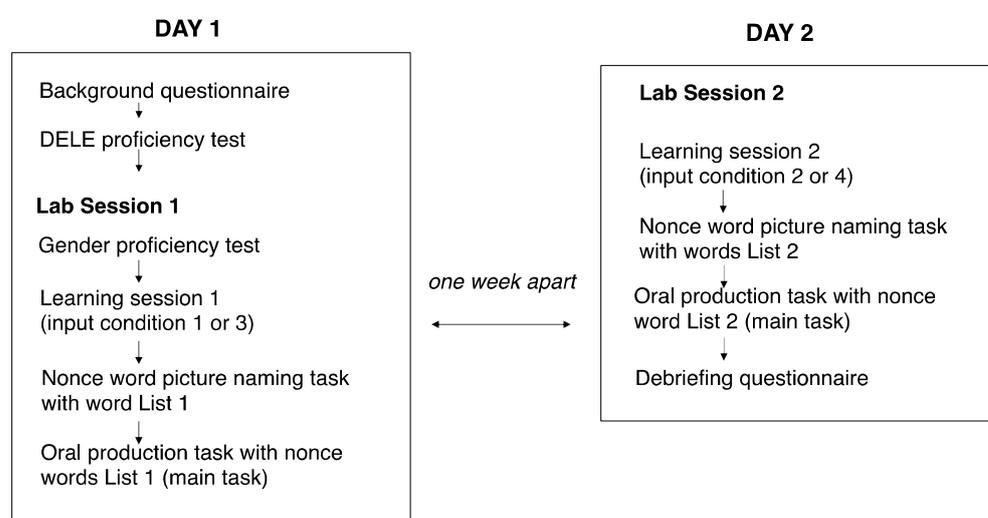


Figure 6. Summary of the study timeline and procedure.

4. Results

Oral responses of all three experimental tasks were transcribed, coded, and double-checked by two research assistants. Target responses were coded as correct; those with one or more discrepancies from the target response were coded as incorrect. (Data files and codes for all tests reported can be found at https://osf.io/j6p5b/?view_only=9794107ee2884da99aaf8623d76537ca). All statistical models were run using the *glm* and *glmer* functions in the *lme4* package (Bates et al. 2015) for R (v4.1.3, R Core Development Team). We started with the maximum random effect structure (Barr et al. 2013) and chose the best fit models that converged with our data. Alpha was set at 0.05 for all analyses. However, to lower the possibility of Type II errors, interactions with *p* values less than 0.10 were explored (Larson-Hall and Herrington 2010). Confidence intervals (Cis) for the fixed effects estimates were calculated using the *confint* function.

4.1. Gender Proficiency Pre-Test with Real Nouns

The overall percentage of correct responses for each group was 99.7% for the native speakers, 92.2% for the heritage speakers, and 87.5% for the L2 learners. As most studies of gender agreement show, participants were more accurate on gender agreement with transparent nouns (Bordag et al. 2006; Caffarra et al. 2017; Foote 2014) and with masculine gender, as shown in Figure 7.

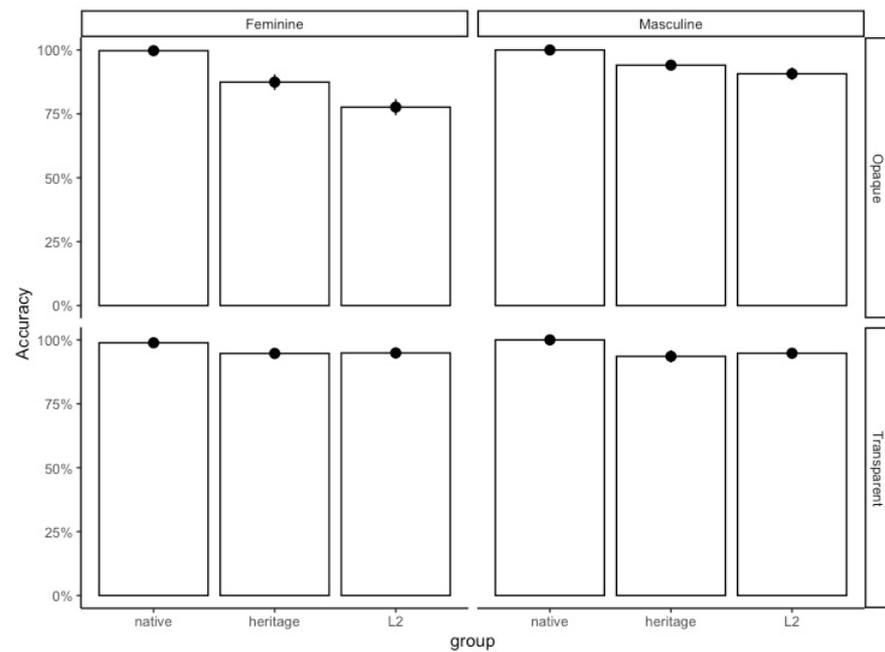


Figure 7. Gender Proficiency Pre-test with Real Nouns. Mean percentage gender agreement accuracy by transparency, gender, and group.

Table 6 shows that the L2 learners and the heritage speakers produced assignment and agreement errors with all nouns in all conditions.

Table 6. Gender Proficiency Pre-test with Real Nouns. Percentage of response type by participant group; 95% confidence intervals appear in parentheses.

Group	Response Type			
	% Ambiguous Errors	% Assignment Errors	% Agreement Errors	% Correct Responses
Heritage	0.70 (0–2.23)	3.19 (1.70–4.72)	3.89 (2.40–5.42)	92.22 (90.72–93.74)
L2	0.79 (0–2.70)	6.13 (4.25–8.04)	5.53 (3.66–7.45)	87.55 (85.67–89.46)
Native	0.10 (0–0.43)	0	0.19 (0–0.52)	99.71 (99.51–100)

Recall that half of the nouns were presented on the screen with gender-matching adjectives (*viento*-MASC /*nuevo*-MASC; *calle*-FEM/*cerrada*-FEM) and half with non-gender-matching adjectives (*hotel*-MASC/*famosa*-FEM; *pared*-FEM/*alto*-MASC). We compared the overall results of gender accuracy by group by noun–adjective gender-matching condition (yes, no). The native speakers were equally accurate with gender-matching (99.8%) and non-matching (99.6%) items; while the L2 learners and the heritage speakers were more accurate with the gender-matching (HS = 98.6%, L2 = 93.9%) than the non-matching (HS = 85.9%, L2 = 81.3%) items. The native speakers were different from the heritage speakers ($\beta = -4.03$, $SE = 0.86$, $z = -4.659$, $p < 0.001$) and from the L2 learners ($\beta = -4.88$, $SE = 0.86$, $z = -5.622$, $p < 0.001$). The group by gender-matching interaction was not significant. Thus, noun–adjective gender non-matching trials did not affect the native speakers’ accuracy at all. For the L2 learners, the percentage of assignment errors (6.13%) was slightly higher than agreement errors (5.53%), but not by much. We checked the non-matching trials in each group by gender. We found very few errors on the masculine items. Yet, on feminine items, the L2 learners made twice the number of errors with assignment in mismatch

trials than the heritage speakers. With masculine items, the number of errors was equal in both groups.

We compared the results of the written proficiency test and the oral gender proficiency test. Since the native speakers did not take the written proficiency test and were categorically accurate in the oral gender proficiency test, they were not included in this model. We fit a linear mixed effects model with accuracy on the gender proficiency task as the dependent variable, group and written proficiency score as fixed factors, and participants as random effects. The model found a significant effect for group ($\beta = 2.27$, $SE = 0.89$, $z = 2.54$, $OR = 9.71$, $95\% CI = [0.52, 4.03]$, $p = 0.011$), for written proficiency score ($\beta = 0.15$, $SE = 0.02$, $z = 7.07$, $OR = 1.16$, $95\% CI = [0.11, 0.19]$, $p < 0.001$), and a group by written proficiency score interaction ($\beta = -0.07$, $SE = 0.03$, $z = -2.66$, $OR = 0.93$, $95\% CI = [-0.12, -0.02]$, $p = 0.008$). The heritage speakers were more accurate than the L2 speakers on the written proficiency task and the oral gender proficiency task, but the difference was bigger on the written proficiency task (HS mean = 79.6%, L2 mean = 67.6%) than in the oral gender proficiency task (HS mean = 92.2%, L2 mean = 87.5%).

To further investigate the differences between these two groups, several linear mixed effects models were run. The best fit for the data based on AIC scores was a simple linear model with accuracy as the dependent measure. Group (heritage, L2 learners), noun gender (masculine, feminine), and transparency (transparent, non-transparent) were included as fixed effects with an interaction between gender and noun transparency. We used treatment coding and set the reference levels to the heritage group, transparent nouns, and masculine gender. The model indicated that L2 learners are significantly less accurate with gender production on real nouns than the heritage speakers ($\beta = -0.53$, $SE = 0.15$, $z = -3.46$, $OR = 0.59$, $95\% CI = [-0.84, -0.23]$, $p < 0.001$), and that the probability of producing gender agreement correctly depends on noun transparency and the gender of nouns ($\beta = -1.27$, $SE = 0.38$, $z = -3.33$, $OR = 0.28$, $95\% CI = [-2.03, -0.53]$, $p < 0.001$): for both L2 learners and heritage speakers, accuracy on gender was lower with feminine non-transparent nouns than with all other nouns (Figure 7). Therefore, L2 learners, who were also of lower proficiency than the heritage speakers, had more difficulty than the heritage speakers with gender production with real nouns.

4.2. Vocabulary Task with Nonce Nouns

The percentage of nouns recalled correctly by each group in all learning sessions combined was 54% for native speakers, 57% for heritage speakers, and 66% for L2 speakers. All groups were more accurate at naming transparent ($M = 64.4\%$, $SD = 47.9$) than opaque-ending nouns ($M = 55.8\%$, $SD = 49.7$), as shown in Figure 8.

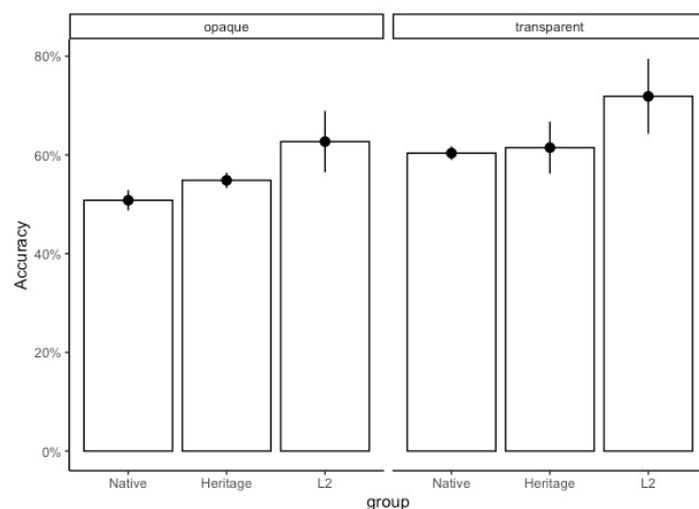


Figure 8. Vocabulary Task with Nonce Nouns. Mean percentage accuracy on word naming by group (all sessions).

We fit a logistic model with accuracy as the dependent variable, group and input condition as fixed factors, and participants as random intercepts (with native speakers and input condition 1 as the reference levels). Group and input condition had a significant effect on the probability of remembering the correct labels for the nonce nouns. The odds of the L2 learners being more accurate at naming nonce nouns was 1.86 times greater than the native speakers ($\beta = 0.64$, $SE = 0.29$, $z = 2.20$, $OR = 1.86$, $95\% CI = [0.06-1.22]$, $p = 0.027$), and the L2 learners were significantly more accurate in input condition 1 (no cues) ($\beta = -0.59$, $SE = 0.12$, $z = -4.697$, $OR = 1.58$, $95\% CI = [-0.85, -0.34]$, $p < 0.001$) and input condition 2 (adjective) ($\beta = -0.73$, $SE = 0.25$, $z = -2.906$, $OR = 1.58$, $95\% CI = [-1.24-0.24]$, $p < 0.001$).

The L2 learners had the highest accuracy scores of the three groups matching the nonce nouns to their pictures. The L2 learners also showed very different patterns of accuracy recall depending on the input condition in which the nonce words were presented during the learning session, as shown in Figure 9. The L2 learners who learned the nouns in the learning session with no cues and in the learning session with the cue on the adjective only were significantly more accurate retrieving labels than the L2 learners who learned the words with transparent determiners or with both determiners and adjectives. By contrast, the native speakers and the heritage speakers showed similar recall accuracy for nouns regardless of the input condition they were presented in during the learning session.

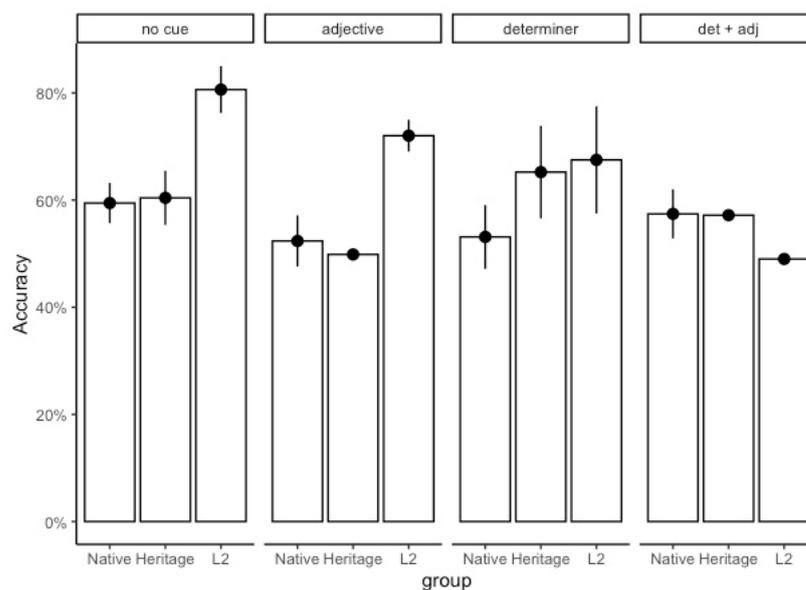


Figure 9. Vocabulary Task with Nonce Nouns. Naming accuracy for input condition by group.

We ran a logistic model with accuracy as the dependent variable, group and transparency as fixed factors, and participants as random intercepts (with native speakers and opaque nouns as the reference levels). The output indicated that group and transparency had a significant effect on the probability of remembering the correct labels for the nonce nouns. The odds of the L2 learners accurately naming nonce nouns was 1.86 times greater than the odds of the native speakers ($\beta = 0.62$, $SE = 0.30$, $z = 2.07$, $OR = 1.86$, $95\% CI = [0.03, 1.21]$, $p = 0.038$), and the odds of all groups being more accurate with transparent nouns was 1.67 times higher than with opaque nouns ($\beta = 0.51$, $SE = 0.15$, $z = 3.32$, $OR = 1.67$, $95\% CI = [0.21, 0.82]$, $p < 0.001$). We compared the performance of the participants in Lab Session 1 and Lab Session 2. A linear mixed effects model assessing accuracy by group and by session number confirmed that the probability of remembering words was higher in the second session than in the first session for all groups ($\beta = 0.30$, $SE = 0.09$, $z = 3.48$, $OR = 1.35$, $95\% CI = [0.13, 0.47]$, $p < 0.001$).

4.3. Oral Production Task with Nonce Nouns

Oral responses were recorded, transcribed, and coded for accuracy. Based on the accuracy of the determiner, noun, and adjective, error type was also coded for inaccurate responses (assignment, agreement, ambiguous; see examples in (4)). Responses were discarded if any part of the noun phrase was omitted, if a non-transparent determiner or adjective was substituted in the noun phrase, if the wrong noun was produced, if the response could not be understood by any member of the research team, or if a response was coded as an ambiguous item (e.g., **la reco rojo*) (there were only 41 errors of this type (19 for L2 learners, 18 for heritage speakers, 4 for native speakers) accounting for 0.7% of all errors). All transcripts were checked by two Spanish-speaking research assistants who also performed the interrater reliability (97.97%). This resulted in the removal of 122 responses for which agreement was not reached, or 2.03% of the dataset. The analyses are based on all items on which there was agreement.

A Welch two-sample *t*-test found that responses for Lab Session 1 (76.2% correct) were significantly less accurate than Lab Session 2 (78.8% correct) ($t = -2.41$, $df = 5892.7$, 95% $CI = [-0.047, -0.005]$, $p = 0.016$). However, as in the picture-naming vocabulary task, the increased performance on Lab Session 2 was true of all groups. This means that all participants got better at performing the learning task the second time around. Although we do not think that this difference in accuracy between the two lab sessions affected the overall patterns reported next, we acknowledge this is a limitation of our design.

The native speakers and the L2 speakers had a similar overall accuracy rate of 78.5% and 78.4% correct responses, respectively, whereas heritage speakers were overall slightly less accurate at 75.6% correct responses. The remaining responses consisted of assignment errors (**la-FEM reco-MASC roja-FEM*) or agreement errors (**el-MASC reco-MASC roja-FEM*). For all three groups, most error responses were assignment errors (native speakers: 19.6%, L2 speakers: 18.9%, heritage speakers: 22.6% of all responses), and very few were agreement errors (native speakers: 2.1%, L2 speakers: 2.6%, heritage speakers: 1.8% of all responses). As in the gender pre-test with real nouns, which used the same format, we compared responses on gender-matching versus non-matching noun–adjective trials. Participants were significantly more accurate on gender-matching trials ($\beta = 0.59$, $SE = 0.13$, $z = 4.287$, $p < 0.001$), and there was no main effect for group or gender-mismatch by group interaction. All groups were more accurate on gender-matching than non-matching trials (native = 81.4% matching vs. 75.3% non-matching, HS = 79.9% matching vs. 71.3% non-matching, L2 = 82.2% matching vs. 74.9% non-matching). All groups made more assignment and agreement errors on the non-match (native speakers: assignment 21.7%, agreement 3%; heritage speakers: assignment 24.2%, agreement 3.4%; L2 learners: assignment 21%, agreement 4%) than in the match trials (native speakers: assignment 17.4%, agreement 1.1%; heritage speakers: assignment 19.9%, agreement 0.2%; L2 learners: assignment 16.6%, agreement 1.2%). However, as shown in Figure 10, accuracy differed significantly depending on the gender and transparency of the nonce noun. Within all three participant groups, accuracy on non-transparent feminine nouns was markedly low, ranging from 40.8% to 57.1%. In contrast, accuracy with non-transparent masculine nouns ranged from 80.3% to 91.2%, and accuracy with transparent nouns was at ceiling in all groups.

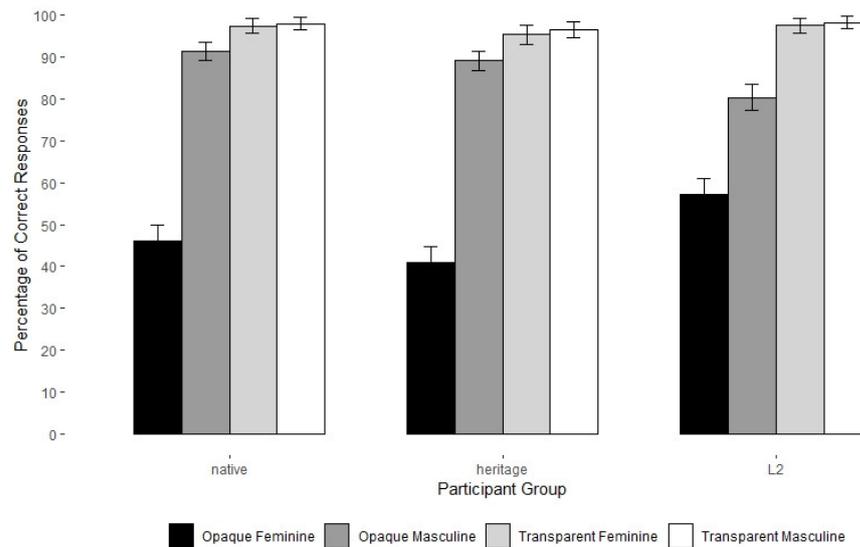


Figure 10. Oral Production Task with Nonce Nouns. Overall accuracy by nonce word gender and transparency for the four input conditions combined. Error bars represent 95% confidence intervals.

We ran a logistic model with accuracy as the dependent variable, group, gender, and transparency as fixed factors, an interaction between the three factors, and participants and items (nonce nouns) as random intercepts. We used treatment coding and set the reference levels to native speakers, masculine gender, and non-transparent nouns. The output in Table 7 shows a significant effect for group (native speakers are different from L2 speakers but not from heritage speakers), a significant effect for gender (feminine is less accurate than masculine), a significant effect for transparency (non-transparent less accurate than transparent), a gender by group interaction (the L2 learners are significantly different from the native speakers but the heritage speakers are not), a group by gender by transparency interaction (the L2 learners were different from the native speakers; the heritage speakers were not). Thus, all groups were more accurate producing gender on transparent than non-transparent nonce nouns. Within non-transparent nouns, all groups were more accurate on masculine nonce nouns presented than feminine nonce nouns. The L2 learners were more accurate than the native speakers and the heritage speakers on feminine non-transparent nouns. There were no differences between the native speakers and the heritage speakers (pairwise comparisons: native-heritage: $z = 1.092, p = 0.5189$; native-L2: $z = -2.084, p = 0.0931$; heritage-L2: $z = -3.160, p = 0.0045$).

Table 7. Oral Production Task with Nonce Nouns. Mixed effects logistic regression on accuracy by group, gender, and transparency.

Fixed Effects	Estimate	Std. Error	z Value	Odds Ratio	95% CI		p
					LL	UL	
(Intercept: reference level native)	2.67	0.28	9.50	14.42	2.12	3.22	0.001
groupHeritage	-0.21	0.28	-0.73	0.82	-0.76	0.35	0.466
groupL2	-1.02	0.28	-3.70	0.36	-1.56	-0.48	0.001
genderF	-2.87	0.32	-9.00	0.06	-3.49	-2.24	0.001
TransparentY	1.53	0.52	2.96	4.63	0.52	2.55	0.003
groupHeritage:genderF	-0.06	0.23	-0.25	0.94	-0.51	0.40	0.801
groupL2:genderF	1.53	0.22	6.93	4.60	1.09	1.96	0.001
groupHeritage:TransparentY	-0.25	0.51	-0.49	0.78	-1.24	0.74	0.618
groupL2:TransparentY	1.19	0.58	2.07	3.28	0.06	2.32	0.039
genderF:TransparentY	2.66	0.71	3.77	14.36	1.28	4.05	0.001
groupHeritage:genderF:TransparentY	-0.01	0.67	-0.02	0.99	-1.32	1.30	0.988
groupL2:genderF:TransparentY	-1.60	0.76	-2.09	0.20	-3.09	-0.10	0.036

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Since there was little variation with transparent nouns and non-transparent masculine nouns (Figure 10), we fit a logistic model to understand the combined effect on grammatical accuracy of participant group and cues to gender in the input using only non-transparent feminine nouns (Figure 11), which showed the greatest variation in accuracy in all groups.⁷

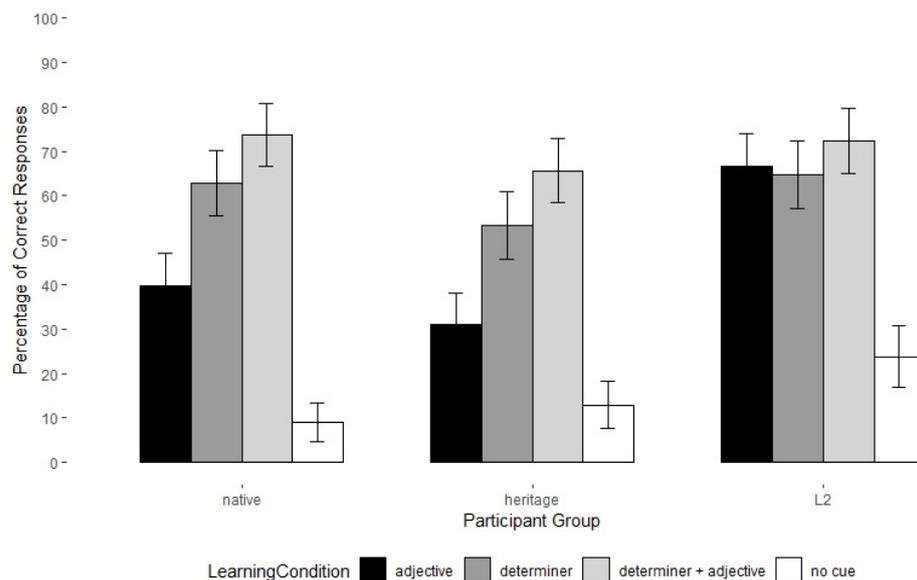


Figure 11. Oral Production Task with Nonce Nouns: Mean percentage accuracy on gender with non-transparent, feminine nonce nouns presented in the four input conditions. Error bars represent 95% confidence intervals.

The logistic model included group (native, HS, L2) and input condition (Input Condition 1: no cues; Input Condition 2: adjective cue; Input Condition 3: determiner cue; and Input Condition 4: determiner cue + adjective cue) as fixed effects, with an interaction between the two. There was a random effect for participant and for item (nonce noun). We used treatment coding and set the reference levels to native speakers and Input Condition 1, respectively. Table 8 shows the output of the model and Table 9 shows pairwise comparisons.

Table 8. Oral Production Task with Nonce Nouns. Mixed effects logistic regression on accuracy with non-transparent feminine nonce nouns.

Fixed Effects	Estimate	Std. Error	Z Value	Odds Ratio	95% CI		p
					LL	UL	
(Intercept: reference level native)	−3.05	0.54	−5.63	0.05	−4.11	−1.99	0.001
groupHeritage	0.32	0.57	0.57	1.38	−0.80	1.45	0.571
groupL2	1.36	0.54	2.50	3.88	0.29	2.42	0.013
LearningCondition2	2.45	0.61	4.03	11.57	1.26	3.64	0.001
LearningCondition3	3.73	0.53	7.00	41.55	2.68	4.77	0.001
LearningCondition4	4.17	0.73	5.72	64.75	2.74	5.60	0.001
groupHeritage:LearningCondition2	−0.82	0.49	−1.66	0.44	−1.79	0.15	0.097
groupL2:LearningCondition2	0.17	0.48	0.37	1.19	−0.76	1.11	0.714
groupHeritage:LearningCondition3	−0.87	0.74	−1.17	0.42	−2.31	0.58	0.241
groupL2:LearningCondition3	−1.26	0.73	−1.74	0.28	−2.68	0.16	0.083
groupHeritage:LearningCondition4	−0.52	0.75	−0.70	0.59	−1.99	0.94	0.485
groupL2:LearningCondition4	−1.30	0.74	−1.76	0.27	−2.74	0.15	0.078

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Table 9. Oral Production Task with Nonce Nouns. Pairwise comparisons for group by input conditions for the model involving non-transparent, feminine nouns.

Input Cond.	Contrast	Estimate	Std. Error	z Ratio	Odds Ratio	95% CI		p
						LL	UL	
1 (no cue)	native–heritage	−0.32	0.57	−0.57	0.72	−1.67	1.02	0.838
	native–L2	−1.36	0.54	−2.50	0.26	−2.63	−0.08	0.034
	heritage–L2	−1.03	0.53	−1.94	0.36	−2.28	0.22	0.129
2 (adj.)	native–heritage	0.49	0.48	1.03	1.64	−0.63	1.62	0.559
	native–L2	−1.53	0.48	−3.22	0.22	−2.64	−0.42	0.004
	heritage–L2	−2.02	0.48	−4.20	0.13	−3.15	−0.90	0.001
3 (det.)	native–heritage	0.54	0.47	1.16	1.72	−0.55	1.63	0.477
	native–L2	−0.10	0.48	−0.21	0.91	−1.22	1.03	0.977
	heritage–L2	−0.64	0.48	−1.33	0.53	−1.76	0.49	0.379
4 (det. + adj.)	native–heritage	0.20	0.48	0.41	1.22	−0.93	1.32	0.911
	native–L2	−0.06	0.50	−0.12	0.94	−1.22	1.11	0.992
	heritage–L2	−0.26	0.49	−0.52	0.77	−1.41	0.89	0.861

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

The model returned a significant effect for group: across the four conditions for opaque feminine nonce nouns, L2 speakers were significantly more accurate on gender production than the native speakers, who were no different from the heritage speakers (Figures 10 and 11). Furthermore, the results show a significant effect of input condition; responses for Input Condition 1 (no cues) were statistically lower than on all other conditions. This is expected since the gender with these words is hard to tell and the participant had to make a prediction, which was predominantly masculine default.

In input condition 1 (no cue), the L2 learners were significantly more accurate than the native speakers and in input condition 2 (adjective cue), the L2 learners were significantly more accurate than the native speakers and the heritage speakers. Conversely, all three groups performed alike in input conditions 3 (determiner cue) and 4 (determiner cue + adjective cue). These differences are visualized in Figure 11, which shows a gradual increase in the proportion of accurate responses in conditions that include a determiner cue or determiner + adjective cues for heritage and native speakers, but not for L2 speakers, and confirmed by the pairwise comparisons in Table 9. These results indicate that the L2 learners are not affected by the inconsistent input on the adjective–noun pairs and use these cues more than the other two groups.

4.4. Debriefing Questionnaire

At the end of the second experimental session, participants answered four main questions to understand how they perceived the experiment and their strategies for learning the new words:

1. What was it like to learn the new words? Was it easy or hard? What strategies did you use?
2. How did you know if a word was masculine or feminine?
3. Did the new words make you think of any real Spanish words?
4. Do you have any other comments about the experiment?

With respect to question 1, 38.4% of participants (43 of 112) found learning the nonce words hard, but 6.25% (7 of 112) indicated it was a lot of fun. Fifty percent of participants (56 of 112) reported that strategies for learning and remembering the words were to create an association between part of the nonce noun and its corresponding object, or between some combination of the nonce noun, the nonce object, and real-world nouns and objects. For example, “I associated *nala* with the cow-like creation because the arch made an ‘n’”, or “I would try to think of something that rhymes and compare it. E.g., *la nete > la neta.*”

Others described extra-linguistic cognitive strategies like self-quizzing, selective attention, and creating a mental support structure.

We were most interested in the responses to questions 2 and 3 since these would reveal learners' awareness of the cues in the input when assigning gender to nonce nouns in Spanish. For question 2, which asked about learning the gender of nouns, participants sometimes gave more than one response. All responses were coded and classified into four main learning strategies related to morphosyntactic gender cues (the noun ending, the determiner, or the adjective, sometimes deduced from responses like "if [the noun] ended in -o or -a"), and chunking behaviors. Chunking behaviors, consisting of apparent attempts to learn the noun along with the determiner, the adjective, or both, were described by many participants. This included responses with explicit references to learning the article and/or adjective along with the nonce noun ("I tried to learn the article and the word together as one", "Remembering the phrases said in the recording"). Table 10 shows the distribution of responses by individuals in the groups who reported use of a gender cue or chunking behavior.

Table 10. *Debriefing questionnaire.* Percentage and number of participants per group (in parentheses) who utilized different strategies for assigning gender to nonce nouns.

Strategy	Groups		
	Native Speakers	Heritage Speakers	L2 Learners
noun ending	56.8% (<i>n</i> = 21/37)	41.0% (<i>n</i> = 16/39)	55.6% (<i>n</i> = 20/36)
determiner	27% (<i>n</i> = 10/37)	30.8% (<i>n</i> = 12/39)	16.7% (<i>n</i> = 6/36)
adjective	--	5.1% (<i>n</i> = 2/39)	25.0% (<i>n</i> = 9/36)
chunking	2.7% (<i>n</i> = 1/37)	20.5% (<i>n</i> = 8/39)	11.1% (<i>n</i> = 4/36)

Participants' responses indicated that for transparent nouns, the ending was by far the most robust cue to gender for all groups. Paying attention to the noun ending was mentioned by 56.8% of the native speakers, 41% of the heritage speakers, and 55.6% of the L2 speakers. All groups reported paying attention to the determiner, but many more native (27%) and heritage speakers (30.8%) than L2 learners (16.7%), while no native speakers and very few heritage speakers (5.1%) reported paying attention to the adjective: by contrast, 25% of the L2 learners reported paying attention to adjectives. This means that aside from the use of the transparent noun ending, which was the most common strategy by all three groups, native and heritage speakers most often reported use of the determiner, compared to L2 speakers who most often reported use of the adjective. Finally, 20.5% of heritage speakers relied on chunking, the strategy that is assumed to be used by very young children learning Spanish, more than the native speakers (2.7%) and the L2 learners (11.1%). Using the *party* R package (Hothorn et al. 2006), we fit the raw data with a classification inference tree (group ~ N-ending + determiner + adjective + chunking), a non-parametric class of regression trees, which confirmed that the main split in the data was the adjective strategy ($p < 0.001$), which set apart the L2 learners from the heritage speakers and the native speakers, as shown in the decision tree in Figure 12.

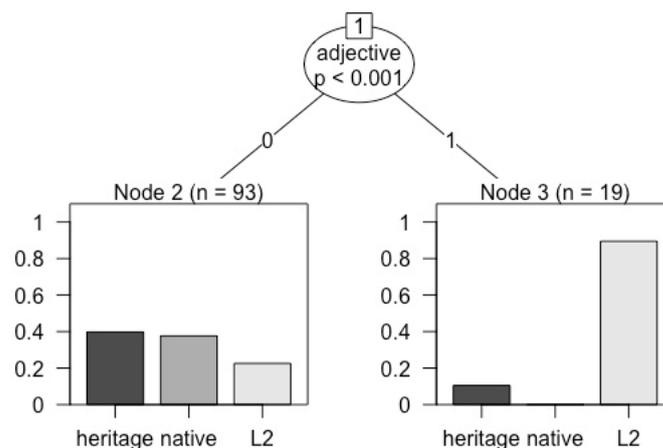


Figure 12. *Debriefing questionnaire.* Partition classification tree showing that the adjective strategy was different among the groups.

The analysis of the responses to the debriefing questionnaire confirms the quantitative and statistical tendencies found in the oral production task.

5. Discussion

Studies of monolingual acquisition show that the pre-nominal determiner is an informative cue to assign gender to Spanish nonce nouns, stronger than adjectives or word ending (Arias-Trejo et al. 2013; Anderson and Lockowitz 2009). If language learning experience determines how gender is assigned to new nouns in Spanish, we predicted differences among groups with different language learning histories. Heritage speakers and monolingually raised native speakers of Spanish would also use the determiner as the strongest cue for gender assignment to nonce nouns because they started learning Spanish in childhood in a naturalistic setting and had to extract nouns and split them from determiners in the auditory input, as supported by recent research (Lew-Williams and Fernald 2007b; Fuchs 2021, 2022). L2 learners typically learn the L2 through written form. Since the visual input they receive is already segmented, we predicted that they would be less likely than native speakers to closely link the determiner with the noun, as found by Martoccio (2022) with real known nouns.

Our study with nonce nouns largely confirmed the trends found in recent research. Our findings revealed that all groups relied on the determiner cue to similar extents, but the L2 learners prioritized adjectives as the most informative syntactic cue to gender classification in Spanish. Although gender in nonce nouns was assessed in phrases with gender-matching and non-matching adjectives (relative to the consistent gender in which the nouns and adjectives were presented in the learning sessions), all groups were more accurate on gender-matching than gender-mismatching trials. Still, the results suggest that the native speakers and the heritage speakers were better at paying attention to more than one cue, whereas the L2 learners relied more on adjectives when they contained the only cue despite adjectives not being the most reliable cue due to the gender-match and mismatch manipulations in the oral production task with nonce nouns. So, it appears that conflicting input is not very distracting to L2 learners (Hopp 2016).

We demonstrated that having morphosyntactic cues to gender in the syntactic environment is critical for gender classification at the lexical level. All groups were most “inaccurate” with gender production with the (transparent and non-transparent) nonce words presented with possessive determiners and opaque color adjectives, which are neutral for gender (*su tida azul/reco azul*) (learning condition 1, no cue). Since it was impossible to tell the gender of these determiners and adjectives, the probability of assigning masculine gender to most of the nonce words by the three groups was higher for this condition than for the other three conditions, which supports the hypothesis that masculine gender is the default gender in Spanish (Harris 1991). The high accuracy on gender agreement with transparent nouns by all groups in this task suggests that having more morphosyntactic

cues to gender in the syntax (Carroll 1989; Grüter et al. 2012), as in the exposure scenarios of input conditions 2, 3, and 4 (see Table 3), helps to classify nonce nouns in the right gender category.

If the determiner is the most commonly used cue for lexical gender assignment in Spanish, then higher accuracy in gender production was expected for participants who completed training sessions in learning conditions with transparent determiners + noun phrases (input conditions 3 and 4) as opposed to non-transparent determiners + noun phrases (input conditions 1 and 2). And if early language experience plays a role in how speakers process input when learning new nouns in the language, we expected native and heritage speakers to be more accurate with words presented with informative gender cues on determiners than on adjectives. Conversely, L2 learners of Spanish were expected to perform differently if they use word endings or the adjective more than the determiner as more informative cues to gender in Spanish.

Despite having lower proficiency in Spanish overall and in gender agreement in the gender proficiency pre-test than the heritage speakers and the native speakers, the L2 learners were more accurate than the other two groups at remembering the nonce nouns' labels. On the oral production task, accuracy on gender production was similar for all groups who learned the words in the input condition that presented the words with gender on transparent determiners (*la tida/naje azul*, condition 3), including for the L2 learners, for whom we predicted lower accuracy in this condition. While noun endings on both transparent and non-transparent nouns and determiners appear to be a stronger cue for gender assignment and agreement than adjectives for all learner groups, the L2 learners showed a different pattern because they were even more accurate with gender production after learning the nonce words in the input condition that presented the words with gender on adjectives only (*su naje negra*, condition 2) (Figure 11). While the use of determiners was consistent in the experiment, the presentation of adjectives was accurate in all learning conditions but inconsistent in the gender pre-test with real nouns and in the oral production task with nonce words.

Although the presentation of gender-mismatching noun–adjective stimuli did not affect the native speakers' gender agreement accuracy in the gender pre-test with real nouns, all groups were more accurate with the gender-matching than with the gender-mismatching condition in the oral production task with nonce nouns. Still, the L2 learners showed a different pattern of results because they used the adjective more than the other two groups as a predictive cue for gender with nonce nouns. These results are reminiscent of those of Foote (2014), who tested sensitivity to gender in word repetition with gender-matched and gender-mismatched input with real nouns. Foote's experimental results showed that the native speakers and the heritage speakers were sensitive to gender agreement between the determiner and the noun in both gender match (*el mejor libro* “the best book”) and mismatch (**la mejor libro* “the best book”) trials. The L2 learners of Spanish, by contrast, only slowed down, i.e., were able to make use of gender cues in word recognition, when those cues conflicted with the gender of the noun. So, gender mismatch input affects L2 learners differently during gender processing.

Similarly, Hopp (2016) found that when native speakers of German were exposed to incorrect determiner–noun sequences in an experimental task, they stopped using predictive gender processing. A possible interpretation of our results is that heritage speakers and native speakers were indeed paying attention to all cues, which is why they noticed the inconsistent noun–adjective pairs and therefore stopped using the adjective as a predictive cue with nonce nouns in input condition 2. However, as discussed above, L2 learners are not very affected by inconsistent input. Another possibility suggested by an anonymous reviewer is that adjectives are longer and more perceptually salient than determiners. Therefore, these results confirm previous findings that L2 learners may process gender differently from native (Carroll 1989; Faber 2017; Hopp 2016; Martoccio 2022) and heritage speakers (Montrul et al. 2013, 2014) at initial stages of learning. Given that all learners must build grammatical representations from input, and in this study all learners

were exposed to the same input, these differences may stem from their language experience as second language learners who learn the L2 later in life and in an instructed setting. What the present study makes more explicit than previous research is precisely how L2 learners go about this inference task differently.

Confirmatory evidence that the L2 learners focused on different morphosyntactic cues and may learn words differently from native speakers and heritage speakers comes from the picture-naming vocabulary task with nonce words, a secondary task used to test their memory or retrieval of the nonce words right after the learning sessions. The L2 learners were more accurate than the other two groups naming the nonce words overall and particularly more accurate remembering the words presented in the input condition with no morphosyntactic cues (condition 1) and with transparent adjectives (condition 2) (Figure 9).

The reason why the L2 learners were more accurate than the other two groups remembering nouns (word learning) in these conditions in the vocabulary task may be related to their enhanced metalinguistic training and classroom-instructed learning experience. L2 learners typically learn and memorize decontextualized nouns in lists when they study. Although all nonce nouns appearing in the learning sessions were presented in sentences, those presented in input condition 1 were not contextualized in a way that indicated gender in any of the elements of the DP. Since there were no cues to gender (except for word markers on the transparent nouns), at this initial stage of acquisition, the L2 learners were able to focus their attention on remembering the meaning of the words. In the oral production task with nonce nouns where the input was inconsistent (noun–adjective match and mismatch trials), the L2 learners performed the best on the words presented with transparent adjectives (input condition 2), unlike the other two groups. Whereas [Arnon and Ramscar \(2012\)](#) indicated that initial exposure to isolated nouns may block later learning of grammatical gender in L2 speakers, our results from the picture-naming vocabulary task in Figure 4 indicate that input with additional morphosyntactic gender cues on determiners (input conditions 3 and 4) may cause processing overload to L2 learners in learning new noun labels themselves. Since young monolingual children ([Pérez-Pereira 1991](#)), adult Spanish native speakers and heritage speakers use all cues, having fewer cues is less helpful for them.

The debriefing questionnaire also revealed confirmatory information about how the different speakers approached the learning task, which complements the quantitative results of the oral production task testing gender production. Like [Martoccio \(2022\)](#) found for L2 learners and native speakers, all three groups in the present study used noun ending and determiners to figure out the gender of nouns. However, more L2 learners than native and heritage speakers reported relying on adjective cues to learn the gender of non-transparent nouns, whereas more native speakers and heritage speakers than L2 learners reported relying on determiners. More heritage speakers than L2 learners reported resorting to chunking (repeating the determiner with the noun), which is the mechanism assumed to operate in L1 acquisition. The different strategies were confirmed visually and inferentially with a classification tree (Figure 7).

Another result of our study is that length of experience and frequency of use of words play a role in gender accuracy, a conclusion we reached by comparing the results of the gender proficiency task with real words and the oral production task with nonce words. Accuracy on gender with real nouns correlated with proficiency scores for the heritage speakers and the L2 learners: the L2 learners were of lower proficiency and scored lower on gender accuracy in the gender proficiency pre-test than the heritage speakers and the native speakers, confirming the results of several previous studies ([Alarcón 2011](#); [Montrul et al. 2008, 2014](#)). As found by [Martoccio \(2022\)](#), the L2 learners assigned incorrect gender to high-frequency nouns more than the other groups. Together, the results of the gender pre-test with real nouns and of the oral production task with nonce nouns support the weaker links hypothesis ([Gollan et al. 2008](#)) because there is a relationship between amount of experience with the language (in terms of length of exposure and type of input) and accuracy on vocabulary: the native speakers have

been using the language for many years and in rich linguistic environments. Their lexical links are strong, such that inconsistent input presentation in the form of gender-mismatching noun-adjective pairs in the gender-pretest with real nouns did not affect their accuracy with gender production at all. The heritage speakers had longer exposure and experience with real words in Spanish than the L2 learners, but they have weaker lexical links than native speakers. That is why they were more affected by the inconsistent input presentation in the gender-pretest than the native speakers, but they were still more accurate with gender production than the L2 learners, confirming the findings of Montrul et al. (2013). But with the nonce nouns used in the experiment, all speakers had the same length of experience and were impacted by the inconsistent input presentation in the oral production task with nonce nouns to the same extent because their lexical links with nonce words are still weak compared to lexical links of the real nouns they know. That is why we were able to isolate, in some way, the mechanisms guiding gender learning in the different groups based on their language processing experience with the language.

Based on previous research with native and non-native speakers, we predicted that the gender of nouns with transparent endings would be easier to learn than non-transparent nouns. We found that noun transparency plays a role for L2 learners and heritage speakers with real nouns, but less so for native speakers, confirming previous findings with L2 learners and different types of bilinguals (Bordag et al. 2006; Caffarra et al. 2017; Foote 2014). By contrast, in the oral production task with nonce nouns, there were no overall differences between the native speakers, the heritage speakers, and the L2 learners. However, when gender production errors occurred with the nonce words, these were more frequent with non-transparent, irregular-ending nouns, and all groups made more errors of gender assignment (**el_{Masc} tida_{Fem} negro_{Masc}*) than of gender agreement (**la_{Fem} tida_{Fem} negro_{Mas}*), including the native speakers. All groups were equally affected by the inconsistent input presentation, performing more accurately on the gender-matching than on the gender-mismatching noun-adjective condition. This suggests that when language experience with specific words and input presentation are controlled for, all learners, regardless of linguistic background, perform similarly, and word transparency matters for everybody.

Finally, we confirmed, with real nouns and nonce nouns, that errors with gender production in noun phrases tend to be lexical rather than syntactic in both L2 learners (Grüter et al. 2012; Hopp 2013, 2016) and heritage speakers (Montrul et al. 2013; cf. Fuchs et al. 2015): of all possible non-target patterns for a DP like *la-FEM serpiente-FEM negra-FEM* (**la serpiente negro*, **el serpiente negro*, **el serpiente negra*), more than 85% of the errors are of correct agreement but incorrect assignment (**el serpiente negro*). However, more gender assignment than gender agreement errors in our study could also arise from our methodology of presenting gender-matching and gender-mismatching adjectives in the gender pre-test with real nouns and in the oral production task with nonce nouns because the native speakers also made this error with nonce nouns. Critical-period-inspired accounts of gender difficulties in L2 learners (Carroll 1989; Hawkins and Franceschina 2004; Franceschina 2001, 2005) have suggested that syntactic mechanisms for gender agreement are intact in infancy but may be faulty in post-puberty learners. If that is the case, then lexical assignment errors are precisely the kind of error that is expected of native speakers when learning new words.

To conclude, using nonce nouns is a good way to gain new information on the classification of nouns at the very initial stages of learning. We found differences in how learners with different types of language learning experience learn the gender of nouns in Spanish. Heritage speakers are like native speakers in how they process input for learning gender in Spanish, while L2 learners show different patterns. Heritage speakers and native speakers pay attention to all informative cues whereas L2 learners rely more on adjectives, even when these may be used inconsistently in the experiment.

Although our results suggest that language experience throughout the lifetime affects input processing and initial learning, these results must be interpreted with caution because our study is not immune to methodological limitations. First, although all participants were exposed to all the nonce words, they did so in two different conditions and in two

different sessions. Ideally, all participants should learn all words in only one session and in the different conditions in a full Latin square design, or twice the number of participants must be recruited to learn all the words in all learning conditions in a between-subjects design. Second, it was difficult to tell what type of gender learning could take place in the learning condition with no cues, so future studies must re-evaluate the inclusion of this input manipulation. Third, the L2 group could focus only on late L2 learners or be expanded to include a more balanced sample of early and late L2 learners to see whether age and proficiency within this group makes a difference in how L2 learners approach noun learning initially. Finally, a gender production task that avoids presenting matching and mismatching noun–adjective pairs may be more appropriate to assess gender assignment accuracy. As suggested by an anonymous reviewer, a gender assignment test would be a good solution to test gender processing without production (see also Montrul et al. 2014).

Author Contributions: Conceptualization, S.M.; methodology, S.M., S.A.M. and A.A.; software, S.A.M. and A.A.; validation, S.A.M. and A.A.; formal analysis, S.M.; investigation, S.M.; writing—original draft preparation, S.M.; writing—review and editing, A.A. and S.A.M.; visualization, A.A. and S.A.M.; supervision, S.M.; project administration, S.M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. This research was funded by funds from the University of Illinois at Urbana-Champaign awarded to Silvina Montrul.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Illinois at Urbana Champaign (protocol code IRB#15650 initially approved on 9 March 2015) for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data files and codes for all tests reported can be found at https://osf.io/j6p5b/?view_only=9794107ee2884da99aaf8623d76537ca.

Acknowledgments: We are grateful to Guilherme D. Garcia with help with statistics and statistical consulting.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

- ¹ Not all adjectives show gender agreement either, depending on their morpho-phonological form. If we consider colors, for example, the colors *negro/a* “black”, *blanco/a* “white”, *rojo/a* “red”, and *amarillo/a* “yellow” can be inflected in the feminine or masculine form, but those ending in “e” such as *verde* “green” and *celeste* “light blue” or in a consonant such as *azul* “blue” cannot be visibly inflected for gender and remain invariant (*El auto celeste/azul* “the light blue/blue car,” *la luz verde/azul* “the green/blue light.”).
- ² These data were collected before the COVID-19 pandemic, when in-person testing was still the norm for experimental studies.
- ³ Five participants had all or some data lost through technical error. Five participants were removed due to variation in language background. Six participants were removed due to inability to complete the task because of low proficiency, reduced attention, or difficulty reading.
- ⁴ One native Spanish speaker, age 29, reported being born in the United States to Colombian parents. However, this participant was raised and educated in Colombia without childhood exposure to English. They completed school and university coursework in Colombia. Only Spanish was spoken at home, and they began studying English as an adult.
- ⁵ One-way ANOVAs comparing masculine and feminine nouns: (1) Frequency: $F(1,22) = 0.9, p = 0.35$; (2) Length in letters: $F(1,22) = 3.3, p = 0.09$; (3) Length in syllables: $F(1,22) = 0.63, p = 0.45$. One-way ANOVAs comparing transparent and non-transparent nouns: (1) Frequency: $F(1,22) = 1.0, p = 0.35$; (2) Length in letters: $F(1,22) = 1.23, p = 0.28$; (3) Length in syllables: $F(1,22) = 3.14, p = 0.091$.
- ⁶ Given the number of participants (3 groups), words (24), and learning conditions (4), it was not possible to do a full Latin square design so that all participants learned all words in all learning conditions. This would have required twice the number of participants or twice the number of sessions or twice the length of each learning session, making this study way too costly in terms of human and financial resources.

- 7 As in the overall results, all participants were more accurate on the adjective-noun gender-matching items than on the gender-mismatching ones (native = 50.6% matching vs. 41.6% mismatching, HS = 45.7% matching vs. 36.2% mismatching, L2 = 60.3% matching vs. 54.0% mismatching, $\beta = 6.32$, $SE = 3.16$, $z = 4.287$, $p = 0.002$).

References

- Afonso, Olivia, Alberto Domínguez, Carlos J. Álvarez, and David Morales. 2014. Sublexical and lexico-syntactic factors in gender access in Spanish. *Journal of Psycholinguistic Research* 43: 13–25. [[CrossRef](#)] [[PubMed](#)]
- Alarcón, Irma. 2011. Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition* 14: 332–50. [[CrossRef](#)]
- Alemán Bañón, José, Robert Fiorentino, and Alison Gabriele. 2014. Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research* 30: 275–306. [[CrossRef](#)]
- Anderson, Raquel, and Alison Lockowitz. 2009. How do children ascribe gender to nouns? A study of Spanish-speaking children with and without specific language impairment. *Clinical Linguistics & Phonetics* 23: 489–506.
- Arias-Trejo, Natalia, Alberto Falcón, and Elda Arias-Canto. 2013. The gender puzzle: Toddlers use of articles to access noun information. *Psicologica: International Journal of Methodology and Experimental Psychology* 34: 1–23.
- Arnon, Inbal, and Michael Ramscar. 2012. Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition* 122: 292–305. [[CrossRef](#)]
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [[CrossRef](#)]
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. Parsimonious mixed models. *arXiv* arXiv:1506.04967.
- Bates, Elisabeth, Antonella Devescovi, Arturo Hernández, and Luigi Pizzamiglio. 1996. Gender priming in Italian. *Perception and Psychophysics* 58: 992–1004. [[CrossRef](#)]
- Bates, Elisabeth, Antonella Devescovi, Luigi Pizzamiglio, Simonetta D'amico, and Arturo Hernandez. 1995. Gender and lexical access in Italian. *Perception & Psychophysics* 57: 847–62.
- Bordag, Denisa, Andreas Opitz, and Thomas Pechmann. 2006. Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32: 1090. [[CrossRef](#)]
- Bowles, Melissa A. 2011. Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition* 33: 247–71. [[CrossRef](#)]
- Caffarra, Sandy, Horacio Barber, Nicola Molinaro, and Manuel Carreiras. 2017. When the end matters: Influence of gender cues during agreement computation in bilinguals. *Language, Cognition and Neuroscience* 32: 1069–85. [[CrossRef](#)]
- Carroll, Susanne. 1989. Second language acquisition and the computational paradigm. *Language Learning* 39: 535–94. [[CrossRef](#)]
- Carroll, Susanne. 1999. Input and SLA: Adults' sensitivity to different sorts of cues to French Gender. *Language Learning* 49: 37–92. [[CrossRef](#)]
- Carstens, Vicki. 2000. Concord in minimalist theory. *Linguistic Inquiry* 31: 319–55. [[CrossRef](#)]
- Chomsky, Noam. 1995. Language and nature. *Mind* 104: 1–61. [[CrossRef](#)]
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- Domínguez, Alberto, Fernando Cuetos, and Juan Seguí. 1999. The processing of grammatical gender and number in Spanish. *Journal of Psycholinguistic Research* 28: 485–98. [[CrossRef](#)]
- Faber, Andie. 2017. Assigning Grammatical Gender to Novel Nouns in L1 and L2 Spanish. Unpublished. Doctoral dissertation, University of Massachusetts, Amherst, MA, USA.
- Falk, Julia. 1978. *Linguistics and Language. A Survey of Basic Concepts and Implications*. New York: John Wiley & Sons.
- Foote, Rebecca. 2014. Age of acquisition and sensitivity to gender in Spanish word recognition. *Language Acquisition* 21: 365–85. [[CrossRef](#)]
- Foote, Rebecca. 2015. The production of gender agreement in native and L2 Spanish: The role of morphophonological form. *Second Language Research* 31: 343–73. [[CrossRef](#)]
- Franceschina, Florencia. 2001. Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research* 17: 213–47. [[CrossRef](#)]
- Franceschina, Florencia. 2005. *Fossilized Second Language Grammars*. Amsterdam: John Benjamins Publishing Company.
- Fuchs, Zuzanna. 2021. Facilitative use of grammatical gender in heritage Spanish. *Linguistic Approaches to Bilingualism* 12: 845–71. [[CrossRef](#)]
- Fuchs, Zuzanna. 2022. Eyetracking evidence for heritage speakers' access to abstract syntactic agreement features in real-time processing. *Frontiers in Psychology* 13: 960376. [[CrossRef](#)] [[PubMed](#)]
- Fuchs, Zuzanna, Maria Polinsky, and Gregory Scontras. 2015. The differential representation of number and gender in Spanish. *The Linguistic Review* 32: 703–37. [[CrossRef](#)]
- Gillon Downes, Margaret, Marta Vergara, Horacio A. Barber, and Manuel Carreiras. 2010. Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience* 22: 1870–87. [[CrossRef](#)] [[PubMed](#)]
- Gollan, Tamar, Rosa Montoya, Cynthia Cera, and Tiffany Sandoval. 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. [[CrossRef](#)] [[PubMed](#)]

- Grüter, Theres, Casey Lew-Williams, and Anne Fernald. 2012. Grammatical gender in L2A: A production or a real-time processing problem? *Second Language Research* 28: 191–216.
- Guillelmon, Delphine, and François Grosjean. 2001. The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition* 29: 503–11.
- Harris, James W. 1991. The exponence of gender in Spanish. *Linguistic Inquiry* 22: 27–62.
- Hawkins, Roger, and Florencia Franceschina. 2004. Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. *Language Acquisition and Language Disorders* 32: 175–206.
- Hernández Pina, Fuensanta. 1984. *Teorías psicopsicolingüísticas y su aplicación a la adquisición del español como lengua materna*. Guatemala: Siglo veintiuno.
- Holmes, Virginia, and Dejean B. de la Bâtie. 1999. Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics* 20: 479–506. [[CrossRef](#)]
- Hopp, Holger. 2013. Grammatical Gender in Adult L2 Acquisition: Relations between Lexical and Syntactic Variability. *Second Language Research* 29: 33–56. [[CrossRef](#)]
- Hopp, Holger. 2015. Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching* 53: 277–306. [[CrossRef](#)]
- Hopp, Holger. 2016. Learning (not) to predict: Grammatical gender agreement in non-native processing. *Second Language Research* 32: 277–307. [[CrossRef](#)]
- Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15: 651–74. [[CrossRef](#)]
- Hur, Esther, Julio Cesar Lopez Otero, and Liliana Sanchez. 2020. Gender agreement and assignment in Spanish heritage speakers: Does frequency matter? *Languages* 5: 48. [[CrossRef](#)]
- Keating, Gregory D. 2009. Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59: 503–35. [[CrossRef](#)]
- Kempe, Vera, and Patricia J. Brooks. 2001. The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? *Language Learning* 51: 221–56. [[CrossRef](#)]
- Larson-Hall, Jenifer, and Richard Herrington. 2010. Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics* 31: 368–90. [[CrossRef](#)]
- Lemhöfer, Kristin, Herbert Schriefers, and Peter Indefrey. 2014. Idiosyncratic grammars: Syntactic processing in second language comprehension uses subjective feature representations. *Journal of Cognitive Neuroscience* 26: 1428–44. [[CrossRef](#)]
- Lemmerth, Natalia, and Holger Hopp. 2019. Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition* 26: 21–45. [[CrossRef](#)]
- Lew-Williams, Casey, and Anne Fernald. 2007a. How first and second language learners use predictive cues in online sentence interpretation in Spanish and English. In *Proceedings of the 31st Annual Boston University Conference on Language Development*. Edited by H. Caunt-Nulton, S. Kulatilake and I. Woo. Somerville: Cascadilla Press, pp. 382–93.
- Lew-Williams, Casey, and Anne Fernald. 2007b. Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science* 33: 193–98. [[CrossRef](#)] [[PubMed](#)]
- Lew-Williams, Casey, and Anne Fernald. 2010. Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language* 63: 447–64. [[CrossRef](#)] [[PubMed](#)]
- Lleó, Conxita. 1998. Proto-articles in the acquisition of Spanish: Interface between phonology and morphology. *Models of Inflection*, 175–95.
- López Ornat, Susana. 1997. What lies in between a pre-grammatical and a grammatical representation: Evidence on nominal and verbal form-function mappings in Spanish from 1;7 to 2;1. *Contemporary Perspectives on the Acquisition of Spanish* 1: 3–20.
- MacWhinney, Brian. 1987. The competition model. In *Mechanisms of Language Acquisition*. Edited by B. MacWhinney. Hillsdale: Erlbaum.
- Mariscal, Sonia. 2009. Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language* 36: 143. [[CrossRef](#)] [[PubMed](#)]
- Martoccio, Alyssa. 2022. Gender assignment strategies used by L1 and L2 speakers of Spanish. *International Review of Applied Linguistics in Language Teaching* 60: 229–54. [[CrossRef](#)]
- McCarthy, Corrine. 2008. Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research* 24: 459–86. [[CrossRef](#)]
- Montrul, Silvina. 2011. First language retention and attrition in an adult Guatemalan adoptee. *Language, Interaction and Acquisition* 2: 276–311. [[CrossRef](#)]
- Montrul, Silvina, Israel de la Fuente, Justin Davidson, and Rebecca Foote. 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. [[CrossRef](#)]
- Montrul, Silvina, Justin Davidson, Israel De La Fuente, and Rebecca Foote. 2014. Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism* 17: 118. [[CrossRef](#)]
- Montrul, Silvina, Rebecca Foote, and 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: 3–53. [[CrossRef](#)]

- Morales, Luis, Daniela Paolieri, Paola E. Dussias, Jorge R. Valdés Kroff, Chip Gerfen, and María Teresa Bajo. 2016. The gender congruency effect during bilingual spoken-word recognition. *Bilingualism: Language and Cognition* 19: 294–310. [[CrossRef](#)] [[PubMed](#)]
- Morgan-Short, Kara, Cristina Sanz, Karsten Steinbauer, and Michael T. Ullman. 2010. Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning* 60: 154–93. [[CrossRef](#)] [[PubMed](#)]
- Pérez-Pereira, Miguel. 1991. The acquisition of gender: What Spanish children tell us. *Journal of Child Language* 18: 571–90. [[CrossRef](#)] [[PubMed](#)]
- Sabourin, Laura, and Laurie A. Stowe. 2008. Second language processing: When are first and second languages processed similarly? *Second Language Research* 24: 397–430. [[CrossRef](#)]
- Savickienė, Ineta, and Wolfgang U. Dressler. 2007. The Acquisition of Diminutives. In *A Cross-Linguistic Perspective*. Amsterdam: Benjamins.
- Sánchez-Walker, Noelia, and Silvina Montrul. 2020. Language experience affects comprehension of Spanish passive clauses: A study of heritage speakers and second language learners. *Languages* 6: 2. [[CrossRef](#)]
- Taraban, Roman, and B. Roark. 1996. Competition in learning language-based categories. *Applied Psycholinguistics* 17: 125–48. [[CrossRef](#)]
- Taraban, Roman, and Vera Kempe. 1999. Gender processing in native and nonnative Russian speakers. *Applied Psycholinguistics* 20: 119–48. [[CrossRef](#)]
- Teschner, R. V., and W. M. Russell. 1984. The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics* 1: 115–32.
- White, L., Elena Valenzuela, Martyna Kozłowska-Macgregor, and Yan-Kit Ingrid Leung. 2004. Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics* 25: 105–33. [[CrossRef](#)]
- Wicha, Nicole, Eva Moreno, and Marta Kutas. 2004. Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience* 16: 1272–88. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.