Gender agreement among Russian learners of Spanish in an instructed versus naturalistic learning environment

Olga Tararova[1] https://orcid.org/0000-0002-1606-9497 otararov@uwo.ca
Gabriela Martínez Loyola[1] https://orcid.org/0009-0002-1759-8013 gmarti49@uwo.ca
Martha Black[1] https://orcid.org/0000-0002-5363-8644 mblack85@uwo.ca
Nadav Benjamin[1] https://orcid.org/0009-0005-1628-8074 ebenjam4@uwo.ca
Qiyao Wang[1] qwang262@uwo.ca

Abstract The present study examines Spanish gender agreement among beginner and non-beginner naturalistic and instructed native Russian learners of L3 Spanish. The project has two goals: first, to investigate whether the above groups differed in their target production and comprehension of gender agreement according to a series of morphological variables (gender class, type, and congruency) and secondly, to determine whether there was a relationship between accuracy and task completion times. A total of 49 native speakers of Russian learning Spanish as an L3, divided across two learner groups (24 instructed in Canada and 25 naturalistic in Mexico) and two proficiency levels (28 beginners and 21 non-beginners), along with a control group of 15 native Spanish speakers, completed several tasks. Results demonstrate that regardless of learning environment, native-like proficiency for gender agreement can be achieved at advanced levels. Differences occurred at the beginner level with the naturalistic group performing better with more difficult forms (e.g., feminine, non-canonical, and incongruent), indicating that at initial stages there is an advantage of naturalistic acquisition. Naturalistic learners had faster task completion times, though this did not correspond to higher accuracy levels. This study has important implications for the field of applied linguistics as it places importance on assessing gender acquisition across distinct learning environments.

Keywords: gender agreement, naturalistic and instructed learning, Russian, Spanish, L3 acquisition

[1] Western University, London, ON, Canada

Gender agreement among Russian learners of Spanish in an instructed versus naturalistic learning environment

1. Introduction

There has been extensive research on second language classroom learners' acquisition of both phonology and morphosyntax (e.g., Festman 2021; Montrul et al. 2008). However, less research has focused on examining third language (L3)/additional language (L*n*) learners and those learning a language in an immersive setting, particularly in the domain of morphosyntax (e.g., Tararova et al., 2023). The main contributions in the literature that have been noted in language acquisition of such learners include those examining language typology and level of proficiency. Research that follows this line of study claims that learners whose first (L1) or second (L2) language is similar to the third (L3) language acquire the L3 at a faster rate, in comparison to learners with three typologically different languages (e.g., Festman 2021; Montrul et al. 2008; Tararova et al. 2023).

Most research on language development in non-classroom settings has focused on gains in oral proficiency, overall linguistic competence, and phonological abilities. However, a noticeable gap remains with regards to morphosyntactic abilities developed in non-classroom settings, specifically among multilingual learners. Furthermore, very few studies (e.g., Pliatsikas and Marinis 2013) have examined incidental or implicit learners who immigrated to another country and learned the target language in a naturalistic environment. In our paper, we adopt Leow's (2019) definition of learning *implicitly*, which he refers to as "learning without awareness or incidentally", while learning *explicitly* refers to the notion of learning the target language intentionally by teachers placing an importance on form and meaning (Leow 2019: 480–481).

Our study examines the production and comprehension of gender agreement in Spanish (e.g., *la comid-a ric-a*, the.F food-F delicious-F, 'the delicious food') as an L3 among L1 Russian learners who are proficient in English as their L2. Both Spanish and Russian exhibit morphological gender and gender is inherently assigned on all nouns in both languages. Gender agreement must occur between adjectives, articles, and nouns in Spanish and between adjectives and nouns in Russian, since there are no articles in Russian[2]. The aim of this study is to investigate how learning context (naturalistic vs. instructed) and proficiency level (beginner vs. non-beginner) affect performance with gender agreement in L3 Spanish. Here, by the label 'non-beginner' we are referring to individuals with an intermediate or advanced level of proficiency in Spanish.

This study has important implications for the fields of applied linguistics and language acquisition by examining gender acquisition across distinct learning environments. The process of acquisition specifically in immersive contexts has previously been understudied; therefore, this study aims to better understand the role of learning context by performing a comparative analysis of classroom and naturalistic learners, both at the initial and more advanced stages of non-native language development. Furthermore, by including different tasks of varying levels of difficulty, we aim to

[2] Please note, there are determiners in Russian, such as demonstratives, quantifiers, possessives, cardinal, and ordinal numbers, and indefinites, but there are no definite and indefinite articles as is the case with Spanish.

shed light on the acquisitional stages of adult learning within the domain of morphosyntax (i.e. grammatical gender).

2. Literature review

2.1 Background

Grammatical gender is a common, typological feature, that is present in approximately twenty-five percent of world languages (Corbett 1991). Grammatical gender forms part of a larger noun class system that allows agreement between nouns and other linguistic structures such as adjectives, articles, pronouns, and/or verbs. According to Hockett (1958) and, later, Corbett (1991), "[…] genders are classes of nouns reflected in the behavior of associated words". Therefore, gender in languages is just one way of dividing nouns into classes and, according to some linguists, "grammatical gender" and "noun class" are synonymous terms.

Before proceeding to a discussion on the acquisition of morphological gender in Spanish in instructed and naturalistic environments, it is necessary to illustrate the morphological differences and similarities between the two languages under study, namely, Russian and Spanish[3]. Although the two languages are considered to be typologically different due to their distinct language families, Russian and Spanish nonetheless share many similarities, one of which is gender marking and agreement. Additionally, the acquisition of gender tends to occur at similar developmental stages with gender acquisition occurring at the age of two in Russian monolingual children and at the age of three in typically developing Spanish monolingual children (see Hernández-Piña [1984] for Spanish gender acquisition and Schwartz et al. [2015] for developmental stages of Russian gender acquisition). To better facilitate an understanding of gender marking and agreement in the two languages under study, Sections 2.1.1 and 2.1.2 discuss grammatical gender in Russian and Spanish, respectively.

2.1.1 Grammatical gender in Russian

Russian consists of three noun-gender class forms, namely, masculine, feminine, and neuter. Gender agreement occurs between nouns and adjectives, but not with articles as there are no articles in Russian. Gender marking in Russian is assigned by formal and semantic rules and patterns, whereby nouns referring to a human are semantically assigned to a grammatical gender based on their biological sex (e.g., отец [otets] father.M, сестра [sestra] sister.F) (Corbett 1982 and Corbett 1991; Corbett and Fraser 1999; Wang 2014). Nouns that do not fall into the semantic category are assigned to a grammatical gender class based on formal rules. According to Corbett (1991), these formal rules refer to distinct declinational classes or patterns that determine the gender of a noun. While many nouns abide by the general rules proposed by the four classes, exceptions do exist. The four declinational classes in Russian plus a potential fifth type are indicated below:

[3] In both Russian and Spanish, the noun and its corresponding article and adjective need to not only agree in gender but also in number. For the purposes of this paper, we will not be discussing number. For more information on number agreement, please see Lightbown and Spada 2021; Sarnecka et al. 2007; Arias-Trejo et al. 2014.

(1) Declension I: Nouns that end in a soft or hard consonant are masculine.
(2) Declension II: Nouns that end in -a are feminine, with the exception of nouns that are semantically feminine.
(3) Declension III: Nouns that end in -’ are feminine.
(4) Declension IV: Nouns that end in -o or -e are neuter.
(5) Declension V: Other[4]

Previous studies such as Wang (2014) have expressed the difficulty that learners face when acquiring gender in a language such as Russian. Participants in this study overgeneralized nonce nouns ending in *-a* as feminine, those ending in *-o* as neuter, and confounding results were obtained for nonce words ending in *-i* and *-ju* (Wang 2014). Studies like these demonstrate that learners require a certain level of familiarity with the word and notion of the grammatical gender rules in the target language in order for correct gender assignment to occur, especially at the beginner level.

2.1.2 Grammatical gender in Spanish

Grammatical gender in Spanish, like most Romance languages, is a binary system where all nouns are assigned to a masculine or feminine category. Although gender assignation is a lexical property of nouns, grammatical gender is realized syntactically where there must be agreement between a noun and its determiner and modifier(s), thus resulting in two domains of grammatical gender in Spanish: assignment (lexical) and agreement (syntactic) (Alarcón 2009, 2011). Animate nouns are nouns whose gender is assigned in accordance with the biological sex/social gender of the referent and thus is semantically motivated (e.g., *maestr-o* teacher-M, *maestr-a* teacher-F), whereas inanimate nouns are those whose gender is not semantically motivated and is purely grammatical (e.g., *el plat-o* the.M plate-M, *la mes-a* the.F. table-F), making their classification arbitrary (Montrul et al. 2008). Finally, noun morphology is also a relevant linguistic feature of grammatical gender in Spanish in which nouns can also be classified based on their level of morphological transparency for gender. Most nouns in Spanish follow a canonical or prototypical pattern in which their inherent morphology reveals their grammatical gender class, such that masculine nouns tend to end in /-o/ and feminine nouns tend to end in /-a/ (e.g., *el zapat-o* the.M shoe-M; *la cas-a* the.F house-F) (Montrul et al. 2008). This morphologically prototypical group of nouns can be described as canonical or overt. According to Teschner and Russell (1984), 99.87% of all nouns that end in /-o/ are masculine and 96.30% of all nouns ending in *-a* are feminine in the Royal Spanish Academy’s *Diccionario de la Lengua Española*. Therefore, the canonical pattern for gender marking on Spanish nouns is extremely common. However, there are nouns that do not follow this prototypical pattern and therefore are classified as *non-canonical*, *non-overt*, or *marked* because their morphology does not directly reveal information about their grammatical gender class. These morphological variants, allomorphs, include nouns that end in *-e* (e.g., *el coch-e* the.M car-M) and in consonants (e.g., *la nariz* the.F nose-F). There is also a subclass of non-overtly marked nouns which can be classified as exceptional in that they directly contradict the canonical pattern, such that masculine nouns end in *-a* (e.g., *el problem-a* the.M problem-M) and feminine nouns end in *-o* (e.g., *la man-o* the.F hand-F) (Montrul et al. 2008; Alarcón 2011; Foote 2015).

[4] Corbett and Fraser (2000) only focus on discussing four declensional classes and only one reference is made towards the fifth type, which can be neuter for non-animate nouns and mostly masculine if the noun is semantically masculine. For more information refer to Corbett and Fraser (2000: 67-69).

2.2 L2/L3 acquisition in naturalistic vs. instructed environments

Though our study does not directly test for acquisition and learning differences, Krashen's (1982) Monitor Model is relevant to our project when distinguishing between naturalistic and instructed modes of learning (Lightbown and Spada 2021). For Krashen, "the acquisition process is subconscious and 'effortless', given that the learner processes the language with minimal amount of cognitive or mental effort" (cited in Leow 2019: 481). Moreover, *acquisition* is treated as an implicit, informal, or natural form of learning. This is similar to Ellis et al. (2009) and Leow's (2019) point, who stress that incidental learning is an absence of awareness with low depth of metalinguistic processing (Leow 2019: 481). The term "learning environment", on the other hand, refers to *explicit* instruction where teachers provide explicit grammatical or instructional information in a target language. These definitions raise the issue of whether "native-like" proficiency can be attained in languages learned in late adulthood.

There has been an ongoing debate regarding whether late (adult) nonnative speakers are able to implicitly process language and reach native-like competence (Pliatsikas and Marinis 2013). There are many factors that can play a role in a learner's acquisition of additional languages, including their working memory, proficiency, and the type of learning environment (e.g., instructed/classroom or naturalistic) to which the learner is exposed, among other factors. In a classroom-based environment, classes are usually held twice a week for a total of three to four hours per week; in more intensive programs, students engage in classroom study for multiple hours a day. Muñoz (2008) describes a classroom environment to be formal and highly structured with input limited to what the teacher provides and not much interaction between the learner and the target language outside of the classroom. Additionally, Loewen (2015), defines "traditional" or "classroom" instruction as focused heavily on explicit attention to language, rather than using language to communicate. One important aspect to consider is the language of instruction. This refers specifically to whether the target language is taught in the dominant language of the university or the target language itself. A naturalistic environment is unstructured, and the learner is exposed to virtually unlimited native-speaker input.

Researchers argue that the amount of input is crucial for processing and competence in a non-native language (e.g., Flege and Liu 2001; Dussias 2003; Festman 2021). Based mostly on semi-artificial language data, research shows that adult L2 learners may incidentally learn aspects of non-native morphosyntax while processing the target language for meaning and without being explicitly exposed to learning the language in a classroom setting (e.g., Williams and Kuribara 2008; Rebuschat and Williams 2012; Grey et al. 2014). In their study on L3 acquisition of word order and case marking in a semiartificial language, Grey et al. (2014) examined data from 36 undergraduate L1 English-L2 Spanish incidental learners who were exposed auditorily to an artifical L3 for two weeks. Twenty-one of the participants were advanced learners of Spanish with 9.7 semesters of prior exposure to Spanish, while the rest of the participants were beginner Spanish learners with 3.3 semesters of exposure in L2. Their results pointed to a significant task effect on the initial post-test such that participants performed better with the acceptability-grammaticality judgement task, compared to the picture naming task. The results on the delayed post-test showed significant improvements on the second task, where participants were able to discriminate with

higher and consistent accuracy. Overall, these results align with other previous work (Williams and Kuribara 2008; Rebuschat and Williams 2012), which suggests that adults can learn grammatical aspects of language without explicit instruction in a short period of time without additional practice in the target language.

Other studies also suggest that learners with less than nine months of naturalistic exposure tend to produce errors that stem from erroneous L1 transfer or interference, while learners with more than five-year residence in the country where the language is spoken no longer demonstrate these types of errors and are able to accurately resemble the speech of native speakers. In a study on highly proficient Greek learners of English, Pliatsikas and Marinis (2013) compared how L2 English learners in a classroom environment in Greece and L2 English learners in a naturalistic setting in the UK of nine years of exposure processed wh-dependencies in declarative sentences. Their results showed that the two groups, irrespective of type of exposure, performed at a similar rate of accuracy with comprehension questions (over 70% accuracy in two groups). However, significant differences were found with processing sentences with intermediate gaps where the learners in the UK performed better than the other group. This suggests that linguistic immersion can lead to native-like processing and acquisition among highly proficient learners. Llanes and Muñoz (2013), conducted a study on Spanish-speaking children and adults who studied English at home and abroad in Ireland or England. The authors found that oral competence significantly increased for the study abroad groups, but at-home students performed better with writing tasks, suggesting that explicit instruction of grammatical structures was necessary for knowledge of the grammar, while the procedural and implicit knowledge that the study abroad groups received did not augment the knowledge of L2 grammar in writing. This is consistent with other studies which argue that oral fluency, especially at initial levels of learning, increases while studying abroad, whereas the knowledge of grammatical structures is facilitated more by the traditional classroom setting (see Valls-Ferrer 2012; Du 2013; di Silvio et al. 2016). In a study on the development of L2 fluency in Mandarin, Russian, and Spanish during study abroad programs in China, Russia, Peru and Chile, di Silvio et al. (2016) found that Mandarin and Spanish learners made significant gains in speech rate and mean sentence length. The pause rates also decreased among advanced learners, indicating speedier and more fluid speech.

Finally, when examining individuals learning a language in a traditional classroom vs. those in an immersed environment it is important to tease apart whether the setting has an effect on how quickly an individual is able to perform a grammatical task and whether this rate of response is comparable to that of a native speaker. Klassen et al. (2023) conducted a study on Spanish gender agreement (determiner-noun concord) among intermediate English learners of Spanish prior and after a short-term study abroad program in Spain. Using a self-paced reading task and a picture selection task, the results showed that the participants overused the masculine article as a default form, specifically in the first task prior to immersive learning, corroborating previous studies (e.g., McCathy 2008). After immersion, the number of errors decreased. As for the second task, L2 learners selected the images faster after the study abroad experience, but errors still persisted with gender marking. Again, similar to previous studies (Llanes and Muñoz 2013; di Silvio et al. 2016)

examining adult language acquisition in the study abroad context, participants demonstrated more communicative and lexical gains, while losing sensitivity to grammatical errors.

Overall, studies on modes of acquisition have shown that at the advanced level, regardless of learning environment, learners are able to acquire the target language and resemble the speech of native speakers. At the beginner level, most studies observe fluency and lexical gains with participants who study the target language in the immersive setting, yet there is still debate regarding whether grammatical structures are easier to acquire in the instructed or naturalistic environment.

2.3 Acquisition of morphological gender

In the past two decades, much previous research has focused on the acquisition of L2 gender (see Sabourin et al. 2006; Montrul et al. 2008; Montrul et al. 2014). Specifically, one of the main questions that researchers have been concerned with is whether typological similarity/proximity between two languages plays a role in L2 gender production. Sabourin et al. (2006) studied the assignment and agreement of grammatical gender in Dutch as an L2 among three groups of adult speakers: German, English, and Romance languages (Spanish, Italian, and French). German and Romance languages are groups of languages that exhibit a gender marking system. German, however, is typologically similar to Dutch due to similar ancestry, resulting in congruency between the two languages. Though Romance languages also exhibit a binary gender system, the systems are not congruent with Dutch. Seventy participants were recruited to partake in two experiments: a gender assignment task and a grammaticality judgement task. The results showed that all three groups had an accuracy score over 80% on the gender assignment task. Out of the three groups, the German group performed the best, suggesting that this group directly transferred morphologically similar forms between their L1 and L2, known as surface transfer. The Romance group also performed well, specifically with higher frequency nouns (over 90% accuracy). The English learners performed better with higher frequency nouns but overall performed the worst in comparison to the other two groups (83%). Regarding the second task, similar results were found with the German group performing the best, specifically with higher frequency items, followed by the Romance group, suggesting that typological congruency plays a bigger role than the existence of abstract gender features. As for the middle frequency items, there was no significant difference between the two gender groups. This indicates that speakers of gendered languages have an advantage in the acquisition of grammatical gender over learners that have previous knowledge in languages that do not exhibit a gender marking system. In another study on L2 gender acquisition, Montrul et al. (2014) analyzed data of L2 Spanish intermediate-advanced learners and compared them to the heritage Spanish learners using an oral elicitation task.[5] The L2 group was more accurate with masculine and canonically non-marked nouns, suggesting that the L2 group overgeneralized feminine nouns and by default assigned and produced a masculine form, corroborating the results of previous studies (e.g., Bruhn de Garavito and White 2002; Gamboa

[5] Though we agree the study by Montrul et al. (2014) provides a comparative analysis between L2 and heritage speakers, we chose to mention it in the present study because it shows common errors of gender production among L2 learners, which is relevant to the research carried out here.

Rengifo 2012). Furthermore, more L2 errors were found with noun-determiner agreement than adjective-noun agreement.

When it comes to L3 acquisition, previous studies have demonstrated the advantage of learning an L3/Ln when the languages are typologically related and comprise overlapping properties (see Rothman 2011; Rothman 2013; Rothman 2015; Jarvis 2015; Rothman et al. 2019; Krepka et al. 2020; Festman 2021; Długosz 2023). Additionally, both Cenoz (2003) and later Festman (2021) in her review paper claim that bilingual and multilingual speakers are said to be equipped with a greater degree of metalinguistic awareness, which allows them to process language systems, structures, and rules quicker compared to monolingual speakers. This has been shown to be true with previous studies focusing on gender agreement. Specifically, Długosz (2023) studied gender agreement among two groups of intermediate and advanced participants: Polish native speakers of L2-English/- L3-Swedish and Polish native speakers of L2-English, L3-German, acquiring Swedish as a fourth language (L4). Though Polish is a gendered language, Swedish and German share similarity in gender marking, so it was expected that the L4 group would outperform the L3 group due to positive transfer. Using a speeded grammaticality judgement task, the results showed that the groups did *not* actually differ in terms of their judgement accuracy (roughly 81% accuracy for both groups). Regarding task completion times, the L4 group processed noun phrases faster than the other group, but only at the intermediate level, which could be explained by a surface transfer of similar gender agreement marking that allows learners to automatize the gender agreement processes earlier. As for the sensitivity to ungrammatical structures, the learners of the two groups developed sensitivity to ungrammatically when they reached an advanced level of proficiency. As for acquisition of grammatical gender at the initial stages of L3, Brown (2020) investigated the topic using data from beginner L3 German learners with L1 English and L2 Spanish and the corresponding mirror group, L1 Spanish-L2 English-L3 German learners. The procedure consisted of an untimed grammaticality judgement task, in which 32 participants were presented with 68 sentences consisting of a singular noun in a nominative case and a definite article, which either matched or mismatched the gender of its corresponding noun. The results showed that the first group (L1 English-L2 Spanish-L3 German) significantly outperformed the other group (by 33%), indicating that the transfer of gender occurred from L2 rather than L1, supporting the L2 Status Factor Hypothesis (Bardel and Falk 2007). Also, since both the L2 and the L3 were learned in an instructed setting, participants of that group received explicit metalinguistic knowledge of grammatical gender which facilitated their learning and equipped them with contexts of use. Regarding the matching-mismatching factor, participants were significantly less likely to correctly identify a grammatical gender error if the gender mismatch was “match” or “opposite”; in other words, both matching and mismatching nouns only resulted in 44-47% accuracy, indicating that there was not much difference in terms of congruency. Sá-Leite et al. (2019), on the other hand, while performing a meta-analysis of a variety of language pairings among bilinguals, found that bilinguals were able to more quickly process nouns that were congruent in both their languages compared to those that were incongruent. The gender of incongruent nouns was often assigned based on their L1.

In a more recent study, directly relevant to our research, Tararova et al. (2023) investigated gender agreement among L1 English-L2 Spanish and L1 Russian/L1 Mandarin speakers who were proficient learners of English as L2 and were learning Spanish as their L3. All three learner groups (n=45) were enrolled either in a beginner or advanced level Spanish course at a Canadian institution. In addition to these three learner groups, 15 native Spanish speakers were also included as a baseline for comparison. The participants completed two online (speeded) tasks, including a picture identification and a grammatically judgment task. The first task included 24 target items where participants saw a series of images and were asked to describe the picture pairings orally. In the second task, participants were asked to select the option that best answered the question. Four answers were provided to them with only one being the grammatically correct answer. Both tasks tested for gender form (canonical and non-canonical), gender class (masculine and feminine), and number[6] (singular or plural), as well as task effect. The results showed that advanced learners of all three groups performed at or near ceiling, especially on Task 1. The Russian group outperformed the other two learner groups in both tasks converging on native-like performance. Regarding the beginner groups, again, the Russian group outperformed the other two learner groups, suggesting that L1 learners of gendered languages perform significantly better than the learners of non-gendered languages. In terms of gender form accuracy, participants were more accurate with canonical forms, which corroborates prior research (Montrul et al. 2008; Gamboa Rengifo 2012; Foote 2015). Regarding gender class, all three groups performed better with masculine target items than feminine forms. One explanation, provided by the authors, is based off of McCathy's (2008) hypothesis which states that nouns tend to be overgeneralized to the masculine form, as a general cognitive processing strategy. Even though Russian exhibits both masculine and feminine forms, the beginner group was more accurate with the masculine form (over 80%) than with the feminine form (under 70%). Given that the researchers did not analyze congruency between Russian and Spanish nouns, it is hard to interpret the results with much certainty. As for task effect, only the beginner groups showed more accuracy with the second task, suggesting that orthographic input during testing can yield higher accuracy (e.g., a written question with four answers available in Task 2 in comparison to a picture with only the determiner in Task 1). Overall, the study provides evidence that at advanced levels, participants of gendered languages (e.g., Russian) can resemble native-like speech, independent of the specific task and linguistic variables. At the beginner level, gender class, morphological form, as well as the nature of the task all play a role in L3 gender production.

3. Methodology

3.1 Research questions and hypotheses

Based on the main objectives, the current study is guided by the following research questions:

1. How does context of acquisition (naturalistic vs. instructed) and proficiency level (beginner vs. non-beginner) impact grammatical gender accuracy and agreement in L3 Spanish?

[6] The researchers also tested for number, but since this variable is not studied in the current project, the discussion is omitted.

2. How do noun type (canonical vs. non-canonical; masculine vs. feminine) and gender assignment congruency with Russian affect accuracy rates in each learner group (naturalistic and instructed)?

3. Is there a correlation between task completion time and accuracy scores across all four tasks, and, if so, is this correlation different for naturalistic and instructed learners across proficiency groups?

Based on our research questions and previous studies, we predict the following:

H1**:** Regarding learner type, we expect to find differences in accuracy scores between our naturalistic and instructed learners. Though the research is extremely limited on incidental learning, previous studies find that at the initial stages, immersion where the target language is spoken can be beneficial for fluency and potentially grammar acquisition due to increased and enhanced exposure and input (e.g., Pliatsikas and Marinis 2013). However, at later stages of acquisition, both groups will show similar behaviour and perform equally well.

H2**:** Regarding the linguistic variables, we predict that participants of both groups will be more accurate with canonical nouns compared to non-canonical forms (Sabourin et al. 2006; Montrul et al. 2012; Tararova et al. 2023). Regarding masculine versus feminine forms, we expect participants to be more accurate with masculine forms than feminine, as shown in previous studies (e.g., Gamboa Rengifo 2012; Foote 2015; Tararova et al. 2023), since this will be more of a default form and, therefore, easier to retrieve. Regarding the effect of congruency, we expect to see more accuracy with nouns that are matching in gender between Russian and Spanish.

H3**:** Based on Sá-Leite et al. 2019, we predict that participants of all groups will be quicker and more accurate with non-marked forms (masculine, canonical, and congruent), than marked types (feminine, non-canonical, and incongruent). For task effect, we expect to find differences in task completion times with the oral translation task (Task 1) taking the longest and the grammaticality judgement task (Task 4) taking the least amount of time, specifically among our beginner learners. This is due to the orthographic cues that will be available in Task 4 and not in Task 1, since during the translation of a text, the participants will have to retrieve their knowledge of Spanish grammar and vocabulary in the absence of any cues.

3.2 Participants

A total of 64 adult participants were recruited between December 2020 and March 2023[7]. This included 15 native Spanish speakers which served as a baseline for comparison and 49 L1 Russian

[7] The 64 participants mentioned here were recruited over such an extended period of time since this participant pool is part of a larger project where we test multiple groups of speakers (Mandarin, English, Russian). From 2020-2021, we only recruited instructed Spanish learners, specifically the beginner groups. In November 2021, we expanded the pool to include non-beginner groups. The new ethics to test naturalistic groups was submitted and approved in June 2022, so we began our recruitment then. Additionally, we faced challenges in recruiting this unique subset of participants due to limited inclusion criteria.

speakers who had English as their L2 and were learning Spanish as their L3[8]. Of these, 24 learned Spanish in an instructed learning environment and 25 learned Spanish in a naturalistic environment while working/living in Mexico. Both L1 Russian groups were composed of speakers of differing levels of proficiency in Spanish and were subsequently placed into groups designated as either "beginner" or "non-beginner" (see Table 1 for a full breakdown of participant groups). The participants who learned Spanish in an instructed setting were students enrolled in a beginner-level Spanish course at a university in Ontario (beginner group) or alumni, graduate students or current upper-year students enrolled in Spanish courses across Ontario universities (non-beginner group). To ensure the beginner-level instructed learners would be able to complete and understand the tasks, we recruited individuals from this specific group after they had completed at least one semester (i.e., 4 months) of Spanish. The participants who had learned Spanish in a naturalistic setting included individuals who had been living in Mexico for at least one year, had not had any form of explicit or formal education in Spanish (e.g., through tutors or university classes), and were immersed in the language in their daily lives either through work, study, or family. Many of the naturalistic Russian participants had jobs in real estate, motivational careers, or were employed as personal trainers, among other occupations. Most used English in their work, but some did report using Spanish. See Table 2 for a more detailed breakdown on the average age, average age of acquisition of Spanish (AoA), and average length of residency (LoR) in an English-speaking country for the instructed learners and a Spanish-speaking country for the naturalistic learner group. To ensure correct placement of beginner and non-beginner participants, depending on their level of proficiency in Spanish, all Spanish learners completed a proficiency test which will be further described in the *Study tasks* and *Procedures* section.

Based on the participants' self-reported data provided in the linguistic questionnaire and a brief conversation in Russian at the start of the session (see Section 4.3), all participants were verified as proficient in their native language in addition to English. In Ontario, French is mandatory until Grade 9. Therefore, participants were excluded if they were enrolled in French immersion or took it beyond the mandatory level. This was of particular importance for the instructed learning group, since they had completed schooling in Canada prior to enrolling in Spanish classes at the postsecondary level.

Table 1. Participant profile.

Type of Speaker		
Control Group *L1 Spanish*	**Experimental Group** *L1 Russian - L2 English - L3 Spanish* N = 49	
	Instructed (n = 24)	**Naturalistic** (n = 25)
n = 15	Beginner: n = 16 Non-beginner: n = 8	Beginner: n = 12 Non-beginner: n = 13

[8] The relatively small sample size was due to the fact that Russian learners in Canada constitute a rare and heterogeneous population. Specifically, many potential participants were excluded due to their knowledge of other languages (e.g., Hebrew, French, etc.), which made finding representative groups with only three languages and with the required order of acquisition, a difficult task. Regarding our naturalistic learners living in Mexico, some potential participants were excluded due to their low proficiency in English.

Table 2. Participant sociolinguistic data.

Learner Group					
Instructed			**Naturalistic**		
Beginner			**Beginner**		
Age	**AoA**	**LoR***	**Age**	**AoA**	**LoR****
31.2 (18-50)	20.8 (10-50)	17.1 (0-34)	37.3 (21-52)	33.6 (19-52)	2.2 (0-8)
Non-Beginner			**Non-Beginner**		
Age	**AoA**	**LoR***	**Age**	**AoA**	**LoR****
38.6 (32-54)	19.1 (10-33)	16.4 (3-34)	33.2 (26-53)	22.5 (16-52)	4.8 (1-15)

*Values represent mean scores for average length of residency in an English-speaking country
**Values represent mean scores for average length of residency in a Spanish-speaking country

3.3 Study tasks

The present study included a total of four untimed tasks, namely, a translation, picture identification, oral description, and grammaticality forced choice (GFCT) tasks. Both experimental groups completed all four tasks, while the native control group did not complete Task 1 (Translation Task). All participants completed the study tasks online via the *Zoom* video conferencing platform. We used *Zoom* to conduct the study as it would allow us to reach a larger proportion of potential participants, especially for the naturalistic group who were living in Mexico at the time of the study. Additionally, the study was initiated during COVID-19, therefore, conducting the experimental procedure online was the easiest method for recruitment. All the target tokens for each of the tasks were analyzed and divided into three categories to examine accuracy, namely, morphological gender (masculine/feminine), canonical/non-canonical, and congruency of gender between Russian and Spanish. This categorization will be discussed further in the *Data Analysis* section. For the translation task (Task 1), participants were required to translate a passage from Russian into Spanish (see Example 1). Example 1 contains the first few sentences of the passage, but the entire passage consisted of 25 targets. Note that participants received a version of the passage with no underlining or identification of specific phrases to avoid thereby diverting their attention from the variables under study.

(Example 1) *Soy una chica italiana y tengo una familia pequeña. Vivo con mi madre, mi papá y mi hermano. Mi papá es doctor y trabaja en un hospital nuevo.*
'I am an Italian girl and I have a small family. I live with my mother, my father, and my brother. My father is a doctor, and he works in a new hospital'.

In Example 1, the underlined phrases represent the target phrases that were later analyzed for potential errors. The participants were not required to read the passage out loud in Russian, but rather provide their oral translation in Spanish sentence by sentence. The passage remained on the screen for the entirety of the task.

For the picture identification task (Task 2), participants were given a sentence containing a blank followed by a set of 3 images (see Figure 1, Example 2). Participants chose the image they believed

was correct, with only 1 of the 3 options being grammatically correct. The task included 3 practice examples, followed by a total of 34 questions, 24 of which were the target items (12 masculine and 12 feminine nouns, divided into canonical/non-canonical and congruent/incongruent forms) and 10 distractor items, which included infinitive verb forms and did not test gender agreement.

Figure 1. Example 2

In Example 2, *la* 'the' refers to the feminine singular definite article in Spanish. The first option, *falda* 'skirt' is the only grammatically correct option since the noun is feminine and singular.

For the oral description task (Task 3), participants saw one of two scenarios: a number + an object or a colour + an object. They were then asked to combine the two images to create a short phrase (see Figure 2, Example 3). The task included 2 examples (with answers provided by the researcher) and 2 practice prompts (to be completed by the participant in preparation for the real trials), followed by a total of 34 prompts: 24 of these prompts were the target items (same division as Task 2) and 10 distractors which did not assess gender concord were also included. The target nouns used in this task were similar to that of Task 2 mentioned previously.

Figure 2. Example 3

In Example 3, *rojo* 'red' and *barco* 'boat' would grammatically produce *el barco rojo* 'the red boat' since boat is a masculine noun in Spanish and, therefore, the article and corresponding adjective must also correspond to the noun's masculine form.

Finally, for the GFCT (Task 4), participants were given a question followed by four possible answers (see Figure 3, Example 4). Here, participants were instructed to choose one of the four options they believed was grammatically correct. The task included 3 practice items, followed by a total of 34 test questions. 24 of these questions were the target items (same division as Task 2), in addition to 10 distractor items which were composed of questions that did not test for gender concord. The target nouns used in this task were similar to that of Task 2 mentioned previously.

4. ¿Cuáles quieres comprar?	1. los chocolates más amargos _____ 2. las chocolates más amargos _____ 3. los chocolates más amargas _____ 4. las chocolates más amargas _____

Figure 3. Example 4

In Example 4, option 1 is the only grammatically correct option, since *chocolates* 'chocolates' is a masculine noun; therefore, the article *los* and the adjective *amargos* 'bitter' must agree with the masculine noun.

Though all four tasks focused on gender marking, they differed with regards to the amount of Spanish orthographic cues they provided. For instance, in Task 1, the participants only saw a paragraph in Russian, therefore, no orthographic cues were available in Spanish, in contrast to Task 4, where learners were explicitly provided with four different written options in Spanish. Besides orthographic cues, the nature of each task varied, each aiming to test a distinct feature to determine the role of task effect. More specifically, Task 1 assessed overall competence with gender assignment and agreement since participants had to produce sentences with no access to Spanish cues. Task 2 examined to what extent participants were able to correctly assign and form agreement between a provided article and noun (pictured in the images provided), while Task 3 determined whether participants were able to detect and assign the gender of an article, noun, and corresponding adjective based on the combination of noun and colours provided as images. Finally, Task 4 tested whether participants were able to choose the correct option based on gender agreement between an article, a noun, and an adjective. All the target items were controlled for their relative frequency and familiarity, as we used the most frequent nouns found in recently studied chapters of the beginner Spanish textbooks used in the Spanish language program at a post-secondary institution in Ontario. Additionally, we made sure to not include any targets in Spanish that belonged to the following categories: (i) nouns that when translated would pertain to the neuter category in Russian or (ii) any of the exceptions mentioned in Section 2.1.2 (Grammatical Gender in Spanish) such as ***la*** *man**o*** which ends in -*o* and would most often be masculine, but in this particular case is feminine. We also made sure to use vocabulary that beginner level learners would have access to through their textbook. For the naturalistic participants, we corroborated their knowledge of the vocabulary used in the target nouns through the proficiency test (Section 4.3) since they do not have access to the same textbooks the instructed learners did.

3.4 Procedure

Prior to the experimental session, all participants were vetted with a series of questions to ensure they met the eligibility criteria. They were then provided with a letter of information and a consent form to be signed. This was followed up by a Linguistic Background Questionnaire which required participants to answer questions regarding their gender identity, native language, language use, etc. This questionnaire was administered to gain insight on the participants' language abilities and to further confirm their eligibility. Participants were then provided with a Zoom link to participate in the study. Sessions began with a short interview in the participant's L1 to ensure their

proficiency in the language. This was followed by a Language Proficiency Test in Spanish to determine the individual's level of Spanish and place them into the appropriate proficiency group. This test was composed of two parts: (i) a fill-in-the-blank section where participants had to choose the best option to complete the sentence, and (ii) a cloze test where participants were required to choose the verb conjugation that best completed the story provided. The test was composed of easier questions that were appropriate for beginner learners, followed by questions increasing in difficulty, meant for more advanced participants. The native Spanish control group also completed this proficiency test to obtain a baseline of comparison for the experimental groups. Next, all the participants were administered the four tasks in the following and same order: Task 1-Task 4. In the interest of confidentiality, all data were anonymized. At the end of the session, participants were compensated for their time spent in the study with a $20 (CAD) Amazon e-gift card. The study took place over one study session lasting approximately 45 minutes to an hour, where participants' responses on Tasks 1-4 were recorded using the *Zoom* record function.

3.5 Data analysis

A total of 6205 tokens were analyzed (i.e., 1600 tokens for Task 1 and 1535 tokens each for Tasks 2, 3, and 4). For Task 1, any productions that were given in a language other than Spanish (e.g., some participants produced English utterances) and productions that deviated too much from the target noun and did not permit for the analysis of gender concord (e.g., *la universidad es muy* ***popular*** 'the university is very popular' instead of *"la universidad es muy famos****a****"* 'the university is very famous') were excluded from the final analysis. In the case of *popular*, items such as these could not be analyzed for gender concord since *popular* can be used to describe nouns that are masculine or feminine. The recordings for each participant were segmented per task using *Audacity* (version 3.3.3). The translation task was analyzed using the interview annotation program *ELAN* (version 6.7). The remainder of the tasks were carefully transcribed, and errors and accuracy scores were recorded in an Excel spreadsheet. For all four tasks, target data was analyzed and coded according to 3 categories: overall, canonical/noncanonical, masculine/feminine, and congruent/incongruent, and accuracy rates. Canonical targets included those ending in a vowel (e.g., *un-a chic-a italian-a,* a-F girl-F Italian-F, 'an Italian girl'), while noncanonical targets were those ending in a consonant (e.g., *la Universidad*, the.F university-F). Masculine targets were mainly comprised of nouns/adjectives ending in an *-o* (e.g., *el pel-o cort-o* the.M hair-M short-M 'the short hair') but also included words that do not follow these rules such as those that are not canonically marked for gender (e.g., *el jardín* the.M garden-M). Feminine targets were mainly comprised of nouns ending in *-a* (e.g., *la sill-a* the.F chair-F) but also included some exceptions. Matching targets included those that had the same gender in both Russian and Spanish (e.g., *el barc-o* the.M boat-M), while mismatching targets included those that were one gender in Russian and the opposite gender in Spanish (e.g., *el cuadern-o* the.M notebook-M, masculine in Spanish, feminine in Russian). For further examples of target tokens, please refer to Section 3.3.

Task completion times were calculated for all tasks, considering task completion time to be the time elapsed between the participant's voice onset for the first word in a task and the end of the participant's response for the final stimulus in a task. This metric allows for uniformity of data analysis across all tasks, which encompass various response modalities.

Statistical analyses were performed in Jamovi (version 2.3.21), an open-source spreadsheet-based interface for using the R programming language. All statistical tests provided are two-tailed independent samples Student's *t*-tests, with Welch's *t*-test also provided when samples failed Levene's test of equal variances. A confidence interval (CI) of 95% was used. Please see Appendix A for a summary of all the variables analyzed in the study.

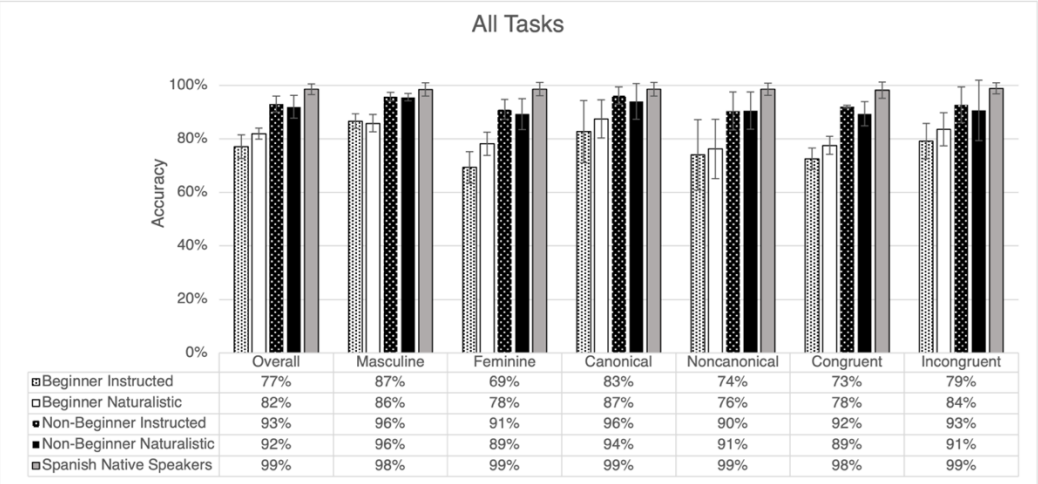
4. Results

We present the results of the present study per task, as outlined in the Methodology section. First, Section 4.1 discusses the overall results of two groups among both proficiency levels across all four tasks. Sections 4.2-4.6 describe the main results obtained for the four tasks individually.

4.1 Overall accuracy scores across all four tasks

Figure 4 illustrates the results of overall accuracy across the four tasks[9] among the four experimental groups (two beginner and two non-beginner) and the Spanish control group. As seen from Figure 4, the control group performed at ceiling for all linguistic variables. Regarding our experimental groups, both non-beginner naturalistic and instructed learners performed at ceiling and, overall, we do not see many differences between these two groups regarding gender production. Where the main differences lie is with the performance of the beginner groups. In general, the beginner naturalistic group performed better than the beginner instructed group, but the differences were minimal and there were few statistically significant findings across the four tasks employed. Regarding the linguistic variables analyzed, we find that all participants performed better with the masculine form and the canonical type, as predicted in Hypothesis 1. Interestingly, regarding congruency, we notice a slight increase in the accuracy of incongruent forms among the beginner groups. We now turn our discussion to Sections 4.2-4.6, where we present each task separately, while discussing the effect of linguistic factors on accuracy and task completion times with gender production.

[9] Keep in mind, the control data included in three tasks only.

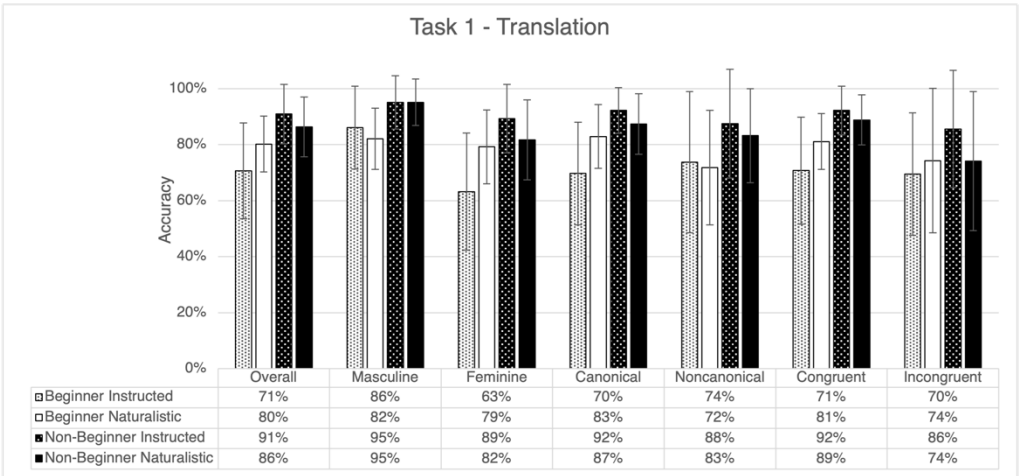


	Overall	Masculine	Feminine	Canonical	Noncanonical	Congruent	Incongruent
Beginner Instructed	77%	87%	69%	83%	74%	73%	79%
Beginner Naturalistic	82%	86%	78%	87%	76%	78%	84%
Non-Beginner Instructed	93%	96%	91%	96%	90%	92%	93%
Non-Beginner Naturalistic	92%	96%	89%	94%	91%	89%	91%
Spanish Native Speakers	99%	98%	99%	99%	99%	98%	99%

Figure 4. Overall rate of accuracy across all four tasks for four groups of Russian learners of Spanish.

4.2 Translation task

Figure 5 demonstrates how each group of Russian learners of Spanish performed on the translation task. No differences were found amongst the non-beginner groups for either method of learning (i.e., instructed, or naturalistic). However, differences were detected among the beginner instructed and naturalistic groups. Specifically, the beginner instructed group performed better than the naturalistic group with feminine and incongruent nouns between Russian and Spanish.



	Overall	Masculine	Feminine	Canonical	Noncanonical	Congruent	Incongruent
Beginner Instructed	71%	86%	63%	70%	74%	71%	70%
Beginner Naturalistic	80%	82%	79%	83%	72%	81%	74%
Non-Beginner Instructed	91%	95%	89%	92%	88%	92%	86%
Non-Beginner Naturalistic	86%	95%	82%	87%	83%	89%	74%

Figure 5. Accuracy scores for Task 1 for four groups of Russian learners of Spanish.

In addition to analyzing overall trends and individual statistical analyses per variable for Task 1, we also ran an independent samples *t*-test to determine if there were any significant differences across proficiency groups and linguistic variables examined in the study. These results are presented in Table 3. Significant differences (marked by *) were not found for the non-beginner groups of either learning environment, but they were found for the beginner groups. Specifically, significant differences were observed for the production of feminine, canonical, and congruent

nouns, with the naturalistic group outperforming the instructed group. Additionally, almost significant results were obtained for congruent nouns (p=0.058).

Table 3. Accuracy scores of all learners (combined, beginner, non-beginner) – independent samples *t*-test for Task 1

Task 1 – Translation Task

Accuracy Scores – Independent Samples T-Test

		All Levels				Beginner				Non-Beginner		
		t		df	p	t		df	p	t	df	p
Overall	Student's t	1.722	[a]	46.0	0.092	1.758		25.0	0.091	-0.961	19.0	0.349
	Welch's t	1.797		41.0	0.080							
Masculine	Student's t	0.506		46.0	0.615	-0.332		25.0	0.743	-0.009	19.0	0.993
Feminine	Student's t	1.807	[a]	46.0	0.077	2.190		25.0	0.038 *	-1.249	19.0	0.227
	Welch's t	1.877		42.4	0.067							
Canonical	Student's t	1.964	[a]	46.0	0.056	1.728	[a]	25.0	0.096	-0.855	19.0	0.403
	Welch's t	2.068		37.7	0.046 *	2.131		25.0	0.043 *			
Noncanonical	Student's t	0.008		46.0	0.993	0.584		25.0	0.565	-1.094	19.0	0.288
Congruent	Student's t	2.079	[a]	46.0	0.043	1.574	[a]	25.0	0.128	-0.151	19.0	0.882
	Welch's t	2.206		35.0	0.034 *	1.988		24.9	0.058			
Incongruent	Student's t	0.850		46.0	0.400	1.331		25.0	0.195	-1.154	19.0	0.263

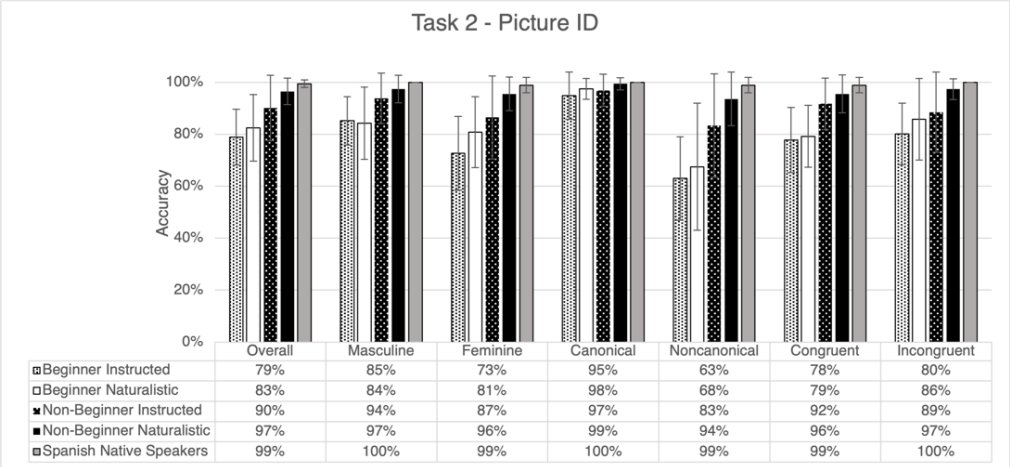
Note. $H_a\ \mu_{Naturalistic} \neq \mu_{Instructed}$

[a] Levene's test is significant ($p \leq 0.05$), suggesting a violation of the assumption of equal variances

* Test statistic is significant ($p \leq 0.05$)

4.3 Picture identification task

Figure 6 demonstrates the accuracy results of the five participant groups (divided by proficiency level and learning context) in the picture identification task according to the three linguistic variables (gender assignment, noun morphology, and cross-linguistic gender congruency). As predicted from Hypothesis 1, the masculine form was easier to produce than the feminine form: even at the non-beginner level, instructed learners performed better with the masculine gender class than the feminine one. Regarding differences between canonical and non-canonical forms, participants of all levels performed at ceiling while producing the canonical form.



	Overall	Masculine	Feminine	Canonical	Noncanonical	Congruent	Incongruent
Beginner Instructed	79%	85%	73%	95%	63%	78%	80%
Beginner Naturalistic	83%	84%	81%	98%	68%	79%	86%
Non-Beginner Instructed	90%	94%	87%	97%	83%	92%	89%
Non-Beginner Naturalistic	97%	97%	96%	99%	94%	96%	97%
Spanish Native Speakers	99%	100%	99%	100%	99%	99%	100%

Figure 6. Accuracy scores for Task 2 for four groups of Russian learners of Spanish, compared to Spanish native speakers.

We also ran an independent samples *t*-test to determine significance levels for naturalistic-instructed differences of means according to our linguistic variables. As seen in Table 4, there was no significant difference when beginner groups are separated from the non-beginner groups. However, when we merged the proficiency levels for naturalistic and classroom learners, we notice significant effects on all three linguistic variables. Specifically, naturalistic learners performed significantly better with feminine, non-canonical and incongruent nouns than instructed learners.

Table 4. Accuracy scores of all learners (combined, beginner, non-beginner) – independent samples *t*-test for Task 2

Task 2 – Picture ID Task

Accuracy Scores – Independent Samples T-Test												
		All Levels				**Beginner**			**Non-Beginner**			
		t		**df**	**p**	**t**		**df**	**p**	**t**		**df**	**p**
Overall	Student's t	2.36		47.0	0.023 *	0.785		26.0	0.440	1.64	[a]	19.0	0.117
	Welch's t									1.365		8.41	0.208
Masculine	Student's t	1.23		47.0	0.224	-0.233		26.0	0.818	1.14	[a]	19.0	0.270
	Welch's t									0.989		9.57	0.347
Feminine	Student's t	2.97		47.0	0.005 *	1.472		26.0	0.153	1.83	[a]	19.0	0.083
	Welch's t									1.524		8.43	0.164
Canonical	Student's t	1.66	[a]	47.0	0.104	0.851		26.0	0.403	1.32	[a]	19.0	0.202
	Welch's t	1.74		33.3	0.092					1.088		8.21	0.308
Noncanonical	Student's t	2.21		47.0	0.032 *	0.596		26.0	0.556	1.56	[a]	19.0	0.134
	Welch's t									1.350		9.34	0.209
Congruent	Student's t	1.72		47.0	0.092	0.287		26.0	0.777	1.02		19.0	0.320
Incongruent	Student's t	2.66		47.0	0.011 *	1.093		26.0	0.284	2.01	[a]	19.0	0.059
	Welch's t									1.602		7.59	0.150

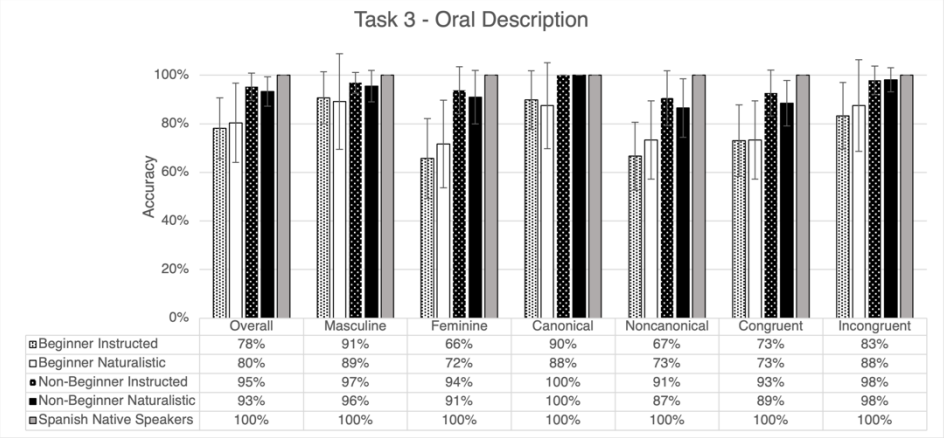
Note. $H_a\ \mu_{Naturalistic} \neq \mu_{Instructed}$

[a] Levene's test is significant ($p \leq 0.05$), suggesting a violation of the assumption of equal variances

* Test statistic is significant ($p \leq 0.05$)

4.4 Oral description task

Regarding the Oral Description Task, similar to the other two tasks, the masculine and canonical forms resulted in higher accuracy rates. Regarding the effect of congruency, as shown in Figure 7, participants of all levels were more accurate with incongruent forms than congruent ones, which contradicts our initial hypothesis.



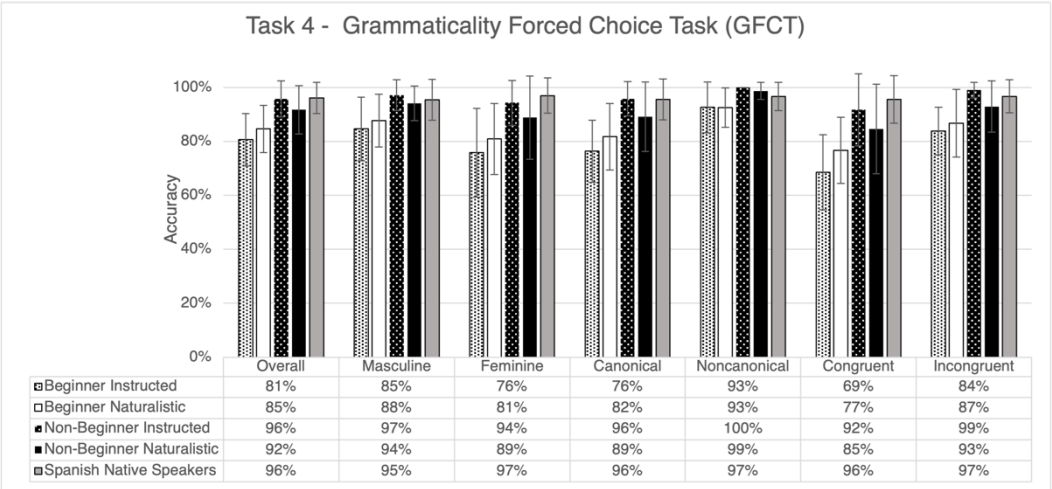
	Overall	Masculine	Feminine	Canonical	Noncanonical	Congruent	Incongruent
Beginner Instructed	78%	91%	66%	90%	67%	73%	83%
Beginner Naturalistic	80%	89%	72%	88%	73%	73%	88%
Non-Beginner Instructed	95%	97%	94%	100%	91%	93%	98%
Non-Beginner Naturalistic	93%	96%	91%	100%	87%	89%	98%
Spanish Native Speakers	100%	100%	100%	100%	100%	100%	100%

Figure 7. Accuracy scores for Task 3 for four groups of Russian learners of Spanish, compared to Spanish native speakers.

However, when analyzing the independent *t*-test results, we see no significant effect between our two learner groups. See Appendix B.

4.5 Grammaticality forced choice task (GFCT)

The results for the last task, the Grammaticality Forced Choice Task (GFCT), are presented in Figure 8. Similar to the other three tasks described previously, both non-beginner groups performed at or near ceiling. This task also shows similarities to the other tasks in that the beginner naturalistic group performed better than the beginner instructed group with nouns that are feminine and those that are incongruent with Russian. However, in contrast to the results obtained in some of the other tasks, particularly Task 1, all groups had an overall score above 80%. This could potentially indicate that having the orthographic representations of each option helps learners in selecting the correct sentence more often. This will be discussed further in the next section.



	Overall	Masculine	Feminine	Canonical	Noncanonical	Congruent	Incongruent
Beginner Instructed	81%	85%	76%	76%	93%	69%	84%
Beginner Naturalistic	85%	88%	81%	82%	93%	77%	87%
Non-Beginner Instructed	96%	97%	94%	96%	100%	92%	99%
Non-Beginner Naturalistic	92%	94%	89%	89%	99%	85%	93%
Spanish Native Speakers	96%	95%	97%	96%	97%	96%	97%

Figure 8. Accuracy scores for Task 4 for four groups of Russian learners of Spanish, compared to Spanish native speakers.

Although we see these trends, we did not observe any significant differences when performing independent samples *t*-tests for Task 4 (see Appendix C).

4.6 Task completion times

Figures 9A-D below demonstrate the task completion times for each of the four tasks for the four groups of Russian learners. Looking at Figure 9A, we see that the non-beginner groups of both learner types took a similar amount of time to complete Task 1 (the translation task), both groups being significantly faster than the instructed beginner learners.

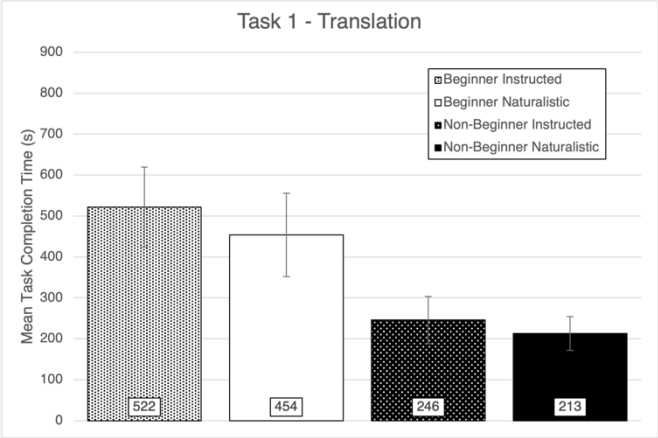


Figure 9A. Task completion times for Task 1 for four groups of Russian learners of Spanish.

Table 5. T-test table for task completion times of all four tasks

All Tasks

Task Completion Times – Independent Samples T-Test

		All Levels				Beginner				Non-Beginner			
		t	**df**	**p**		**t**	**df**	**p**		**t**		**df**	**p**
Task 1	Student's t	2.20	47.0	0.033	*	.955	26.0	0.348		1.06		19.0	0.303
Task 2	Student's t	0.797	47.0	0.429		–0.975	26.0	0.338		0.721	[a]	19.0	0.479
	Welch's t									0.619		9.2	0.551
Task 3	Student's t	1.58	47.0	0.121		0.761	26.0	0.454		0.174	[a]	19.0	0.864
										0.143		8.19	0.890
Task 4	Student's t	0.349	47.0	0.729		–1.37	26.0	0.182		0.489	[a]	19.0	0.631
	Welch's t									0.406		8.42	0.695

Note. H_a $\mu_{Naturalistic} \neq \mu_{Instructed}$

[a] Levene's test is significant ($p \leq 0.05$), suggesting a violation of the assumption of equal variances

* Test statistic is significant ($p \leq 0.05$)

Graphs 9A-9D also include their 95% confidence interval. Regarding Task 2, we see an advantage in task completion time for both non-beginner learner groups as compared to the beginner-level participants, both naturalistic and instructed learners.

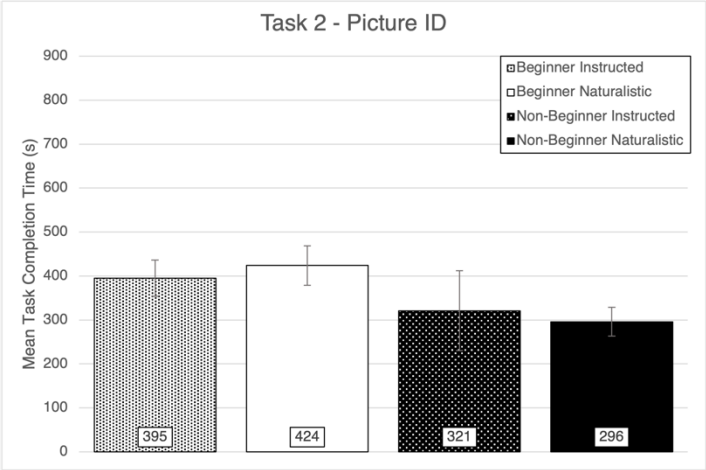


Figure 9B. Task completion times for Task 2 for four groups of Russian learners of Spanish.

Regarding Task 3, the Oral Description Task, the only significant between-group difference was found between the non-beginner naturalistic learners and the instructed beginners. One of the possible explanations could be the nature of task itself, which will be further explained in the Discussion section. Regarding our non-beginner groups, similar to the previous tasks, we do not observe a significant difference, which suggests that both learning contexts equally prepare adult learners with knowledge of the Spanish gender system.

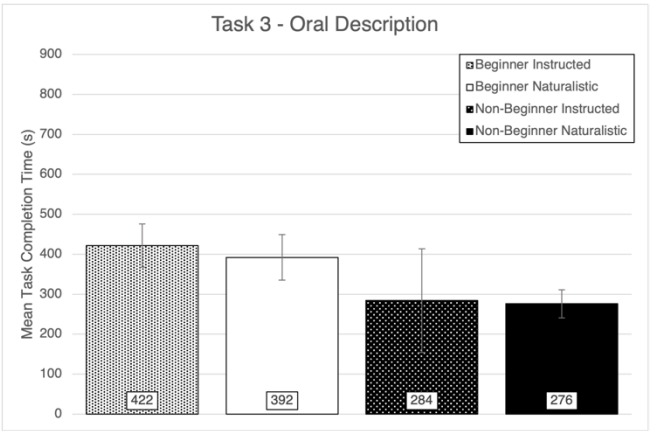


Figure 9C. Task completion times for Task 3 for four groups of Russian learners of Spanish.

Concluding the results section is Figure 9D which illustrates the task completion times of each of the four groups for the linguistic variables analyzed in Task 4 (Grammaticality Forced Choice Task). Similar to Task 2, as demonstrated in Figure 9D, significant advantages are observed for the non-beginner naturalistic learners against both beginner groups, though there is no advantage between learner type for either beginners or non-beginners. This means that regardless of the learning environment, the participants were able to complete this specific task both relatively

quickly and with a higher rate of accuracy as all groups performed above a score of 80% overall, as seen in Section 4.5, indicating a potential effect of task type.

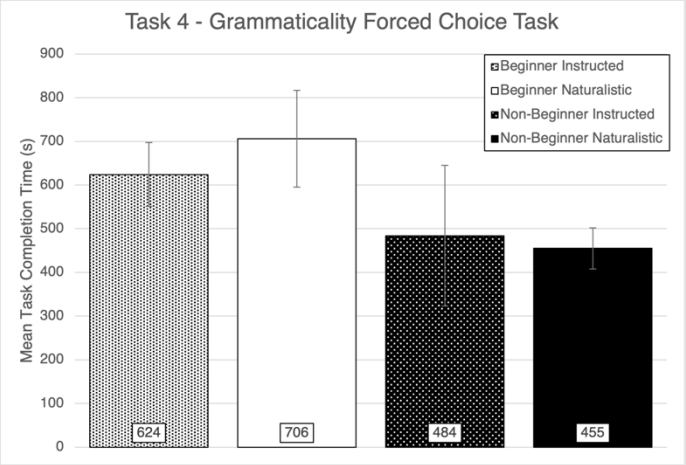


Figure 9D. Task completion times for Task 4 for four groups of Russian learners of Spanish.

Linear correlation (*r*) (Figure 10) and a follow-up omnibus ANOVA test (Table 6) revealed a negative correlation between accuracy (calculated as accuracy across all four tasks) and task completion times (R^2 = -0.491), indicating that individuals who responded faster in gender production tasks tended to be more accurate. Proficiency level was a nearly significant predictor of accuracy (p = 0.058), though this result is unsurprising given that accuracy and proficiency are, for most purposes, indistinguishable. Notably, learner type did not impact the relationship between task completion time and accuracy.

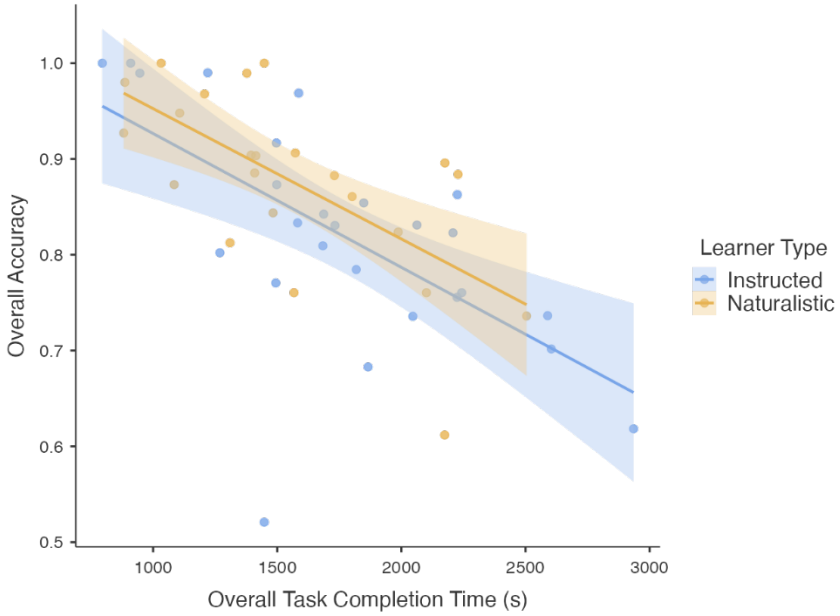


Figure 10. Correlation of participant task completion times and accuracy scores.

Table 6. Omnibus ANOVA test for all four tasks combined

Model Fit Measures

Model	R	R^2
1	0.700	0.491

Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
Level	0.02514	1	0.02514	3.781	0.058
Learner Type	0.00486	1	0.00486	0.731	0.397
Overall Task Completion Time (s)	0.06015	1	0.06015	9.046	0.004
Residuals	0.29925	45	62.9		

Note. Type 3 sum of squares

Model Coefficients – Overall Accuracy

Predictor	Estimate	SE	t	p
Intercept	1.0340	0.0477	21.676	<0.001
Level: Beginner – Non-Beginner	-0.0637	0.0328	-1.944	0.058
Learner Type: Naturalistic – Instructed	0.0207	0.0242	0.855	0.397
Overall Task Completion Time (s)	-9.65e-5	3.21e-5	-3.008	0.004

5. Discussion

To recap, this study analyzed the production and comprehension of grammatical gender agreement among L3 Russian learners of Spanish across two proficiency levels (i.e., beginner and non-beginner) and two learning environments (i.e., instructed, and naturalistic). In this section, we discuss the results of our study in conjunction with previous studies and our predictions. In general, the results indicate that regardless of learning environment, native-like proficiency in terms of gender agreement can be achieved at advanced levels as observed with the non-beginner groups of both instructed and naturalistic learning settings. However, where we saw the majority of differences was at the beginner level, specifically among the naturalistic group who performed better at harder-to-acquire/marked noun gender forms. Furthermore, less transfer between Russian and Spanish was observed in gender marking for these noun classes for this group, indicating that language learning in an immersive environment at initial stages of learning can lead to linguistic gains for this phenomenon. To further discuss some of the results obtained and their main implications, we present our research questions once again in the following sections.

5.1 Research question 1

RQ1: *How does context of acquisition (naturalistic vs. instructed) and proficiency level (beginner vs. non-beginner) impact grammatical gender accuracy and agreement in L3 Spanish?*

Regarding the first research question, we compared learner type (naturalistic vs. instructed) to investigate whether the context of acquisition affected accuracy of gender production and comprehension in Spanish. Non-beginner learners of two modes of acquisition performed almost at ceiling, which indicates that beyond the initial stages, learners of either environment are able to acquire grammatical gender. In this regard, we can assume that both learner groups have subconsciously learnt the syntactic phrasal agreement (Leow 2019).

As seen from the four tasks, the main differences were noticeable at initial stages, with naturalistic beginners performing better overall than instructed beginners, which contradicts previous studies (e.g., Valls-Ferrer 2012; Du 2013; Llanes and Muñoz 2013; di Silvio et al. 2016). One explanation for this could be the quality and quantity of input and exposure in the target-language environment for the naturalistic participants specifically. Though our beginner participants have been residing in Mexico anywhere from one to three years, during an informal interview with them, they confirmed that many are married to Mexican speakers and use Spanish and English on a daily basis in a variety of contexts and domains. Those who are not married, work in a Spanish-speaking environment, again increasing their amount of incidental exposure to Spanish over time, in comparison to our instructed participants. As previous research shows, those who are immersed in a naturalistic environment seem to demonstrate more fluency in their rate of speech compared to those who receive instruction in a classroom (e.g., di Silvio et al. 2016). We noticed this specifically in Task 1, where participants had to translate a passage from Russian to Spanish. This observation suggests that incidental learning at initial stages has certain benefits for language acquisition. Another explanation of naturalistic learner's advantage can be linked to the years of exposure in the target country. In our study, our beginner participants have resided in Mexico at least a year, in comparison to other studies, where the exposure was less than nine months. For example, though Klassen et al. (2023) studied gender agreement among intermediate learners of Spanish, their data have shown that a short-term study abroad experience (25 days) did not lead to gains in grammar, but rather in lexicon and communicative competence. Similarly, Llanes and Muñoz's (2013) study on Spanish children and adults learning English found that during two- to three-month stays in the UK, adults performed better orally and gained more lexical complexity rather than writing gains, compared to the other in-home/instructed adult group. Therefore, in the future, it will be interesting to compare participants with less than a year of naturalistic exposure to Spanish to those with more time to see at what point there is an advantage in grammatical accuracy.

5.2 Research question 2

RQ2: *How do noun type (canonical vs. non-canonical; masculine vs. feminine) and gender assignment congruency with Russian affect accuracy rates in each learner group (naturalistic and instructed)?*

Regarding the linguistic variables, as we predicted in our second hypothesis, our learner groups produced prototypical (masculine, canonical) and incongruent forms with greater accuracy, which supports previous studies (e.g., Montrul et al. 2008; Gamboa Rengifo 2012; Foote 2015; Klassen et al. 2023; Tararova et al. 2023). In our study, we also found that participants performed better with masculine forms than feminine, which corroborates other studies that indicate that the masculine form is used by default. We also observed the canonical form to be easier, again similar to previous studies, but it is interesting to point out that in Task 4, instructed participants performed better with non-canonical forms than canonical ones, which indicates that tasks with orthographic cues lead to better accuracy. Regarding the congruency effect, our results are different from Sá-Leite et al. (2019), who found that congruent forms would be easier to acquire, especially in initial stages of language acquisition. In our study, we did not see these results; moreover, in most of our tasks, we found that our participants performed better with non-congruent forms. One of the possible explanations for this finding could be vocabulary fluency and retrieval of more frequently used words. Though we controlled and included the vocabulary items used from our textbook, we did not test participants' vocabulary familiarity. It is possible that they could have picked the forms (from Tasks 2 and 4) by guessing, rather than knowing noun gender. It is also possible that genetically- and typologically distant gender systems may not be as strongly integrated as closely related languages are. Paolieri et al. (2019) compared gender congruency effects between L1 Russian-L2 Spanish and L1 Italian-L2 Spanish bilinguals and found the presence of a gender congruency effect for both groups, but the magnitude of the effect in the Italian-Spanish bilinguals was greater; for example, Italian-Spanish bilinguals produced gender-congruent concrete noun phrases an average of 88ms faster than Russian-Spanish bilinguals did in the same condition, with similar advantages for abstract nouns and for bare noun conditions. These findings indicate that typological similarity between languages might have precedence when considering a congruency effect. Since this study has not investigated any typologically similar languages, this observation merits future work.

5.3 Research question 3

RQ 3: *Is there a correlation between task completion time and accuracy scores across all four tasks, and, if so, is this correlation different for naturalistic and instructed learners across proficiency groups?*

We examined the accuracy scores of each group and their task completion times for the four tasks analyzed in addition to whether speed was correlated with accuracy level. Regardless of learning environment, we found that our non-beginner groups completed all tasks at a similar pace and had results near ceiling, comparable to that of our native speaker controls. Regarding the beginner learners, we found that both groups had similar task completion times for Tasks 2 and 4 and

accuracy rates near or above 80% overall. For the other tasks, differences in task completion times between the two groups were much larger with the instructed group taking, on average, 294s more than the instructed group on Task 1 and 80s more on Task 3. Some general observations that we made while participants were completing Task 1 specifically are that among the naturalistic group, beginners included, they tended to translate very quickly the passage given in Russian into Spanish and sounded more confident overall when providing their responses. However, as demonstrated in Section 5.2, the speed or rate of answering does not necessarily signify equal or heightened accuracy, as the instructed beginner group performed better than their naturalistic counterpart. The slightly more delayed responses from the beginner instructed group could, in part, be due to the nature of these two tasks. For both Tasks 1 and 3, learners were required to produce a response based on a passage in Russian (Task 1) or a combination of pictures with no words (Task 3). Learners were not given any orthographic input in Spanish as was the case with Tasks 2 and 4. This lack of orthographic cues in Tasks 1 and 3 required participants to retrieve the information from their own prior knowledge, thereby adding to the overall response rate for the beginner instructed learners. These results are in-line with Sá-Leite (2019), who found that learners were quicker at identifying the gender of nouns that were high-frequency, canonically marked, and congruent between their L1 and L2. It is important to keep in mind, however, that the above study did not perform a comparative analysis between reaction times of classroom-based and naturalistic learners. Additionally, the participants in these studies were bilinguals and not trilinguals, as is the case with the current investigation. However, when we consider the total time that beginner instructed learners took compared to the naturalistic learners, we must also consider their relative accuracy rates. In Task 1, while beginner naturalistic learners were quicker, the beginner instructed learners had a higher rate of accuracy (i.e., 80% overall accuracy for beginner instructed learners in comparison to 71% for the naturalistic group). For Task 3, while naturalistic learners were also quicker, they only performed 2% above the instructed group. Therefore, the takeaway here is that speed of responding does not necessarily yield a higher level of accuracy when it comes to gender production at the initial stages of learning, overall supporting Hypothesis 2.

6. Conclusions and future work

This project investigated gender agreement in Spanish among two types of L3 learners: naturalistic learners who immigrated from Russia and have been residing in Mexico, and classroom Russian learners who have been studying or completed their education in a Canadian university. To the best of our knowledge, this is the first study which examined this language grouping in a trilingual setting, while comparing two types of learning environments. Our findings have revealed that there is a clear advantage of immersed incidental learning at initial stages, in which natural informal input and exposure have played a role in gender acquisition. Overall, our naturalistic participants performed better on the four tasks, specifically with more difficult forms. It is important to note, however, that in some cases, they were not as accurate as our instructed participants, but were more fluent and rapid overall, suggesting that while living in a country where the target language is spoken, adults are able to learn aspects of grammar without explicit instruction. Regarding our non-beginner groups, both classroom and naturalistic learners performed at ceiling, resembling the native controls. This suggests that beyond the initial stages of learning, the type of learning environment is not as important since native-like competence can be achieved regardless.

Secondly, though we have not explicitly analyzed the learning-acquisition distinction, which Krashen (1982) puts forward, we can ascertain that the feature of gender has been acquired by both groups, since they have demonstrated accuracy with all the morphological forms.

One of the main limitations of this study was the small participant pool. Given difficulties with finding Russian learners of Spanish with only three languages in Canada and abroad, future projects will expand the criteria and include participants of other Slavic languages (e.g., Polish, Serbian, Ukrainian). Given the large influx of Ukrainian individuals in Canada and other parts of the world, including Latin America, we hope to be able to recruit more participants at a quicker rate. Furthermore, given that the above Slavic languages all exhibit gender, it will be interesting to compare whether the results are similar to those of the current study, specifically whether similar linguistic and non-linguistic factors will be shown to be significant in the production of grammatical gender.

Though this project controlled for participants to be only trilingual, some of our participants, specifically those in Mexico, showed some degree of variation in their use of Spanish. Particularly, among our participants, some were married to Mexican spouses and spoke Spanish with them, while others spoke English at home. Some participants also had jobs where they predominantly spoke Spanish, others regularly communicated in English at work, while others used all three languages regularly. In other words, their use of Spanish at home and at work and the input they received could have affected their accuracy and production of gender in Spanish. Therefore, future work should analyze language use as a social factor to see whether there is a correlation between gender accuracy and percentage of Spanish use in different settings.

Finally, another important factor to consider is motivation. Both Krashen (1982) and later Gardner (1985) discuss motivation and attitudinal factors as a variable for success in language acquisition. Gardner (1985) differentiates between integrative and instrumental motivation, in which integrative motivation is considered to be positive motivation, associated with integration into the culture and the community of the language under study, while instrumental orientation refers to the motivation to learn a target language for a specific goal, such as getting better pay, receiving a higher grade at school, etc. In his socio-education model, Gardner (2005) clarifies that *integrativeness* refers to a learner's openness and ability to "take on characteristics of another linguistic and cultural group" (p. 7), while *instrumentality* refers more to learning a language for practical reasons. Krashen (1982) in his "Monitor Model" introduces the affective filter hypothesis, which can either be facilitative or non-facilitative for learners to acquire a new language. Specifically, learners who are extroverts with high motivation, self-confidence, and less stress are able to acquire the language faster than those who experience a high level of stress and anxiety. Therefore, based on the above research, future work should consider motivation as one of the variables for our trilingual learners. Specifically, a future survey/questionnaire could test learners' reasons for studying Spanish, whether it is to pass a specific language requirement or integrate with the culture and become part of the community. We predict differences in accuracy rates between gender production and learners' personal attitudes towards studying the language, which could either positively affect their acquisition or slow it down, depending on the valence of their language-learning attitudes.

To conclude, this study has important implications for the fields of applied linguistics and language acquisition. First, regarding pedagogical contributions, so far, the process of acquisition in immersive contexts has been understudied; therefore, this study helps to shed light on this group of speakers by performing a comparative analysis of classroom and naturalistic learners. As this study has shown, immersive contexts are beneficial at the initial stages of language learning since grammar can be learned without explicit instruction in a classroom. Secondly, by including different tasks of varying levels of difficulty, we can better understand the acquisitional stages of learning and how easy or difficult it is for adult learners from naturalistic and instructed learning contexts to produce the correct gender form based on the given task.

Ethics and consent

The study was approved by the Ethics Committee of Western University, Canada (Protocol code: 116433, date of initial approval: 5 November 2020. The ethics protocol has been active since then).

Funding

This research was funded by the SSHRC – Seed Grant (Grant #: R6087A05) and Faculty Research Development Fund (Grant #: 0000049826).

Conflicts of interest

The authors declare no conflict of interest.

Appendices

Appendix A: Summary of variables analyzed.

Study Tasks	1. Translation task 2. Picture identification task 3. Oral description task 4. Grammaticality forced choice task (GFCT)
Linguistic Variables	1. Gender (masculine/feminine) 2. Ending (canonical/non-canonical) 3. Congruency (congruent/incongruent)
Extralinguistic Variables	1. Proficiency level 2. Task 3. Task completion time

Appendix B: Accuracy scores of all learners (combined, beginner, non-beginner) – independent samples *t*-test for Task 3 (Oral Description Task)

Task 3 – Oral Description Task

Accuracy Scores – Independent Samples T-Test

		All Levels			**Beginner**			**Non-Beginner**		
		t	**df**	**p**	**t**	**df**	**p**	**t**	**df**	**p**
Overall	Student's t	-1.104	47.0	0.275	0.395	26.0	0.695	-0.772	19.0	0.449
Masculine	Student's t	-0.037	47.0	0.970	-0.276	26.0	0.784	-0.525	19.0	0.605
Feminine	Student's t	-1.558	47.0	0.126	0.883	26.0	0.385	-0.574	19.0	0.572
Canonical	Student's t	-0.472	47.0	0.639	-0.412	26.0	0.683	NaN [a]		
Noncanonical	Student's t	-1.450	47.0	0.154	1.144	26.0	0.263	-0.772	19.0	0.450
Congruent	Student's t	-0.618	47.0	0.540	0.031	26.0	0.976	-1.009	19.0	0.325
Incongruent	Student's t	-1.447	47.0	0.155	0.672	26.0	0.507	0.066	19.0	0.947

Note. $H_a\ \mu_{Naturalistic} \neq \mu_{Instructed}$

[a] All observations are tied

* Test statistic is significant ($p \leq 0.05$)

Appendix C: Accuracy scores of all learners (combined, beginner, non-beginner) – independent samples *t*-test for Task 4 (GFCT)

Task 4 – Grammaticality Forced Choice Task (GFCT)

Accuracy Scores – Independent Samples T-Test

		All Levels			**Beginner**			**Non-Beginner**			
		t	**df**	**p**	**t**	**df**	**p**	**t**		**df**	**p**
Overall	Student's t	1.111	47.0	0.272	1.087	26.0	0.287	-1.127		19.0	0.274
Masculine	Student's t	0.960	47.0	0.342	0.701	26.0	0.489	-1.096		19.0	0.287
Feminine	Student's t	0.863	47.0	0.393	0.846	26.0	0.405	-0.927		19.0	0.366
Canonical	Student's t	0.920	47.0	0.362	1.134	26.0	0.267	-1.369	[a]	19.0	0.187
	Welch's t							-1.60		18.4	0.127
Noncanonical	Student's t	0.534	47.0	0.596	-0.026	26.0	0.979	-1.147	[a]	19.0	0.266
	Welch's t							-1.48		12.0	0.165
Congruent	Student's t	1.181	47.0	0.244	1.547	26.0	0.134	-1.013		19.0	0.324
Incongruent	Student's t	0.574	47.0	0.569	0.710	26.0	0.484	-1.719	[a]	19.0	0.102
	Welch's t							-2.12		15.4	0.051

Note. $H_a\ \mu_{Naturalistic} \neq \mu_{Instructed}$

[a] Levene's test is significant ($p \leq 0.05$), suggesting a violation of the assumption of equal variances

***** Test statistic is significant ($p \leq 0.05$)

Appendix D: Measures of Mean Dispersion

Task 1 – Translation Task

Accuracy Scores – Measures of Mean Dispersion

		Beginner		**Non-Beginner**	
		Instructed	**Naturalistic**	**Instructed**	**Naturalistic**
Overall	Mean	0.706	0.802	0.909	0.863
	St. Error	0.0404	0.0318	0.0376	0.0297
	C.I.	{0.621, 0.792}	{0.73, 0.874}	{0.82, 0.998}	{0.798, 0.928}
Masculine	Mean	0.861	0.821	0.951	0.951
	St. Error	0.0348	0.0344	0.0337	0.023
	C.I.	{0.788, 0.935}	{0.744, 0.899}	{0.871, 1.03}	{0.9, 1}
Feminine	Mean	0.632	0.792	0.893	0.817
	St. Error	0.0493	0.0418	0.0431	0.0395
	C.I.	{0.528, 0.736}	{0.698, 0.887}	{0.791, 0.995}	{0.731, 0.903}
Canonical	Mean	0.697	0.829	0.922	0.874
	St. Error	0.0431	0.0359	0.0289	0.03
	C.I.	{0.606, 0.788}	{0.748, 0.91}	{0.853, 0.99}	{0.809, 0.939}
Noncanonical	Mean	0.737	0.718	0.875	0.832
	St. Error	0.0595	0.0647	0.0686	0.0467
	C.I.	{0.612, 0.862}	{0.572, 0.865}	{0.713, 1.04}	{0.73, 0.934}
Congruent	Mean	0.707	0.811	0.922	0.888
	St. Error	0.045	0.0317	0.0308	0.0248
	C.I.	{0.613, 0.802}	{0.739, 0.882}	{0.849, 0.995}	{0.834, 0.942}
Incongruent	Mean	0.695	0.743	0.856	0.741
	St. Error	0.0516	0.0815	0.0741	0.0688
	C.I.	{0.587, 0.804}	{0.559, 0.928}	{0.681, 1.03}	{0.591, 0.891}

Note. The C.I. (95%) of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

Task 2 – Picture Identification Task

Accuracy Scores – Measures of Mean Dispersion

		Beginner		**Non-Beginner**	
		Instructed	**Naturalistic**	**Instructed**	**Naturalistic**
Overall	Mean	0.789	0.825	0.901	0.965
	St. Error	0.0253	0.0406	0.0445	0.014
	C.I.	{0.736, 0.843}	{0.733, 0.917}	{0.796, 1.01}	{0.934, 0.995}
Masculine	Mean	0.852	0.842	0.938	0.974
	St. Error	0.0219	0.0438	0.0343	0.0146
	C.I.	{0.806, 0.898}	{0.743, 0.941}	{0.856, 1.02}	{0.943, 1.01}
Feminine	Mean	0.727	0.808	0.865	0.955
	St. Error	0.0336	0.0431	0.0566	0.0179
	C.I.	{0.656, 0.798}	{0.711, 0.906}	{0.731, 0.999}	{0.916, 0.994}
Canonical	Mean	0.949	0.975	0.969	0.994
	St. Error	0.0215	0.0127	0.0219	0.00641
	C.I.	{0.904, 0.994}	{0.946, 1}	{0.917, 1.02}	{0.98, 1.01}
Noncanonical	Mean	0.63	0.675	0.833	0.936
	St. Error	0.0376	0.077	0.0704	0.0285
	C.I.	{0.55, 0.709}	{0.501, 0.849}	{0.667, 1}	{0.874, 0.998}
Congruent	Mean	0.778	0.792	0.917	0.955
	St. Error	0.0294	0.0378	0.0352	0.0203
	C.I.	{0.716, 0.84}	{0.706, 0.877}	{0.833, 1}	{0.911, 0.999}
Incongruent	Mean	0.801	0.858	0.885	0.974
	St. Error	0.0279	0.0498	0.0544	0.0111
	C.I.	{0.742, 0.86}	{0.746, 0.971}	{0.757, 1.01}	{0.95, 0.999}

Note. The C.I. (95%) of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

Task 3 – Oral Description Task

Accuracy Scores – Measures of Mean Dispersion

		Beginner		**Non-Beginner**	
		Instructed	**Naturalistic**	**Instructed**	**Naturalistic**
Overall	Mean	0.782	0.804	0.953	0.933
	St. Error	0.0295	0.0516	0.02	0.0167
	C.I.	{0.72, 0.845}	{0.687, 0.921}	{0.906, 1}	{0.896, 0.969}
Masculine	Mean	0.907	0.892	0.969	0.955
	St. Error	0.0251	0.0622	0.0152	0.0179
	C.I.	{0.854, 0.96}	{0.751, 1.03}	{0.933, 1}	{0.916, 0.994}
Feminine	Mean	0.657	0.717	0.938	0.91
	St. Error	0.0386	0.0572	0.0343	0.0305
	C.I.	{0.576, 0.739}	{0.587, 0.846}	{0.856, 1.02}	{0.844, 0.977}
Canonical	Mean	0.898	0.875	1	1
	St. Error	0.0282	0.0559	0	0
	C.I.	{0.839, 0.958}	{0.749, 1}	{1, 1}	{1, 1}
Noncanonical	Mean	0.667	0.733	0.906	0.865
	St. Error	0.033	0.0509	0.04	0.0334
	C.I.	{0.597, 0.736}	{0.618, 0.849}	{0.812, 1}	{0.793, 0.938}
Congruent	Mean	0.731	0.733	0.927	0.885
	St. Error	0.0347	0.0509	0.0332	0.0259
	C.I.	{0.658, 0.805}	{0.618, 0.849}	{0.849, 1.01}	{0.828, 0.941}
Incongruent	Mean	0.833	0.875	0.979	0.981
	St. Error	0.0323	0.0599	0.0208	0.0138
	C.I.	{0.765, 0.902}	{0.74, 1.01}	{0.93, 1.03}	{0.951, 1.01}

Note. The C.I. (95%) of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

Task 4 – Grammaticality Forced Choice Task (GFCT)

Accuracy Scores – Measures of Mean Dispersion

		Beginner		**Non-Beginner**	
		Instructed	**Naturalistic**	**Instructed**	**Naturalistic**
Overall	Mean	0.806	0.846	0.958	0.917
	St. Error	0.0228	0.0278	0.0236	0.025
	C.I.	{0.757, 0.854}	{0.783, 0.909}	{0.902, 1.01}	{0.862, 0.971}
Masculine	Mean	0.846	0.877	0.971	0.941
	St. Error	0.0278	0.0308	0.0202	0.0177
	C.I.	{0.787, 0.905}	{0.807, 0.947}	{0.923, 1.02}	{0.902, 0.98}
Feminine	Mean	0.758	0.809	0.943	0.888
	St. Error	0.0389	0.0417	0.0294	0.0426
	C.I.	{0.676, 0.84}	{0.715, 0.903}	{0.874, 1.01}	{0.795, 0.981}
Canonical	Mean	0.764	0.817	0.958	0.891
	St. Error	0.0272	0.0389	0.0223	0.0358
	C.I.	{0.707, 0.821}	{0.729, 0.905}	{0.906, 1.01}	{0.813, 0.969}
Noncanonical	Mean	0.926	0.925	1	0.987
	St. Error	0.0222	0.0231	0	0.00868
	C.I.	{0.879, 0.973}	{0.873, 0.977}	{1, 1}	{0.968, 1.01}
Congruent	Mean	0.685	0.767	0.917	0.846
	St. Error	0.0327	0.0389	0.0472	0.0461
	C.I.	{0.616, 0.754}	{0.679, 0.855}	{0.805, 1.03}	{0.746, 0.947}
Incongruent	Mean	0.838	0.867	0.99	0.929
	St. Error	0.0207	0.0397	0.0104	0.0264
	C.I.	{0.794, 0.882}	{0.777, 0.956}	{0.965, 1.01}	{0.872, 0.987}

Note. The C.I. (95%) of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

All Tasks

Task Completion Times – Measures of Mean Dispersion

		Beginner		Non-Beginner	
		Instructed	**Naturalistic**	**Instructed**	**Naturalistic**
Task 1	Mean	522	454	246	213
	St. Error	46.2	45.1	24.5	19.1
	C.I.	{424, 619}	{352, 556}	{188, 304}	{172, 255}
Task 2	Mean	395	424	321	296
	St. Error	19.3	19.5	38.6	15.2
	C.I.	{354, 435}	{379, 468}	{230, 413}	{263, 329}
Task 3	Mean	422	392	284	276
	St. Error	25.6	25.1	55	15.9
	C.I.	{368, 476}	{335, 449}	{154, 414}	{241, 310}
Task 4	Mean	624	706	484	455
	St. Error	34.9	49.1	67.8	21.4
	C.I.	{551, 698}	{595, 817}	{323, 644}	{408, 502}

Note. The C.I. (95%) of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

References

Alarcón, Irma. 2009. The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data. *Hispania* 92(4). 814–828. https://www.jstor.org/stable/40648463.

Alarcón, Irma V. 2011. Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition* 14(3). 332–850. https://doi.org/10.1017/S1366728910000222.

Bardel, Camilla, & Ylva Falk. 2007. The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research* 23(4). 459–484. https://doi.org/10.1177/0267658307080557.

Brown, Megan M. 2020. Grammatical gender acquisition in sequential trilinguals: influence of a gendered L1 vs. L2. *Linguistic Society of America* 5(1). 331–344. https://doi.org/10.3765/plsa.v5i1.4712.

Bruhn de Garavito, Joyce, & Lydia White. 2002. The second language acquisition of Spanish DPs: The status of grammatical features. In Ana Teresa Pérez Leroux & Juana M. Liceras (eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection*, 153–178. Dordrecht: Kluwer Academic. https://doi.org/10.1007/978-94-010-0291-2_6.

Cenoz, Jasone. 2003. The role of typology in the organization of the multilingual lexicon. In Jasone Cenoz, Britta Hufeisen & Ulrike Jessner (eds.), *The multilingual lexicon*, 103–106. Dordrecht: Kluwer Academic.

Corbett, Greville G. 1982. Gender in Russian: An account of gender specification and its relationship to declension. *Russian Linguistics* 6(2). 197–232. https://www.jstor.org/stable/40160034.

Corbett, Greville G. 1991. *Gender*. Cambridge Textbooks in Linguistics. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139166119.

Corbett, Greville G. & Norman M. Fraser. 2000. Default genders. In Barbara Unterbeck, Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*, 55–98. De Gruyter Mouton. https://doi.org/10.1515/9783110802603.55.

Długosz, Kamil. 2023. Processing gender agreement in an additional language: The more languages the better? *Second Language Research* 39(4). 997–1026. https://doi.org/10.1177/02676583221113333.

Du, Hang. 2013. The development of Chinese fluency during study abroad in China. *The Modern Language Journal* 97(1). 131–143. https://doi.org/10.1111/j.1540-4781.2013.01434.x.

Dussias, Paola E. 2003. Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition* 25(4). 529–557. https://www.jstor.org/stable/44486698.

Ellis, Rod, Shawn Loewen, Catherine Elder, Hayo Reinders, Rosemary Erlam & Jenefer Philp. 2009. *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847691767.

Festman, Julia. 2021. Learning and processing multiple languages: The more the easier?" *Language Learning* 71(S1). 121–162. https://doi.org/10.1111/lang.12437.

Flege, James Emil & Serena Liu. 2001. The effect of experience on adults' acquisition of a second language." *Studies in Second Language Acquisition* 23(4). 527–552. https://www.jstor.org/stable/44486960.

Foote, Rebecca. 2009. Transfer in L3 acquisition: The role of typology. In Yan-kit Ingrid Leung (ed.), *Third language acquisition and Universal Grammar*, 89–114. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847691323-008.

Gamboa Rengifo, Arnold. 2012. *The acquisition of grammatical gender in Spanish by English-speaking L2 learners*. West Lafayette, IN: Purdue University MA thesis. https://www.proquest.com/docview/1239142014/abstract/431E2A3CAA294B80PQ/1.
Gardner, Robert C. 1985. *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold. https://publish.uwo.ca/~gardner/docs/SECONDLANGUAGE1985book.pdf.
Gardner, Robert C. 2005, May 30. Integrative motivation and second language acquisition. *Canadian Association of Applied Linguistics/Canadian Linguistics Association.* London, ON, Canada. https://publish.uwo.ca/~gardner/docs/caaltalk5final.pdf.
Grey, Sarah, John N. Williams & Patrick Rebuschat. 2014. Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition* 36(4). 611–645. https://www.jstor.org/stable/26324754.
Hernández Pina, Fuensanta. 1984. *Teorías psicosociolingüísticas y su aplicación a la adquisición del Español como lengua materna*. Madrid: Siglo XXI de España Editores. https://www.researchgate.net/publication/39068399_Teorias_psicosociolinguisticas_y_su_aplicacion_a_la_adquisicion_del_espanol_como_lengua_materna.
Jarvis, Scott. 2015. Influences of previously learned languages on the learning and use of additional languages." In Maria Juan-Garau and Joana Salazar-Noguera (eds.), *content-based language learning in multilingual educational environments*, 69–86. Cham: Springer. https://doi.org/10.1007/978-3-319-11496-5_5.
Krashen, Stephen D. 1982. *Principles and practice in second language acquisition*, 1st edn. Oxford: Pergamon Press.
Leow, Ronald P. 2019. ISLA: how implicit or how explicit should it be? theoretical, empirical, and pedagogical/curricular issues. *Language Teaching Research* 23(4). 476–493. https://doi.org/10.1177/1362168818776674.
Lightbown, Patsy & Nina Spada. *How languages are learned*, 5th edn. Oxford: Oxford University Press, 2021.
Llanes, Àngels & Carmen Muñoz. 2013. Age effects in a study abroad context: Children and adults studying abroad and at home." *Language Learning* 63(1). 63–90. https://doi.org/10.1111/j.1467-9922.2012.00731.x.
Loewen, Shawn. 2015. *Introduction to instructed second language acquisition*. New York: Routledge.
Montrul, Silvina. 2008. *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.
Montrul, Silvina, Justin Davidson, Israel de la Fuente & Rebecca Foote. 2014. Early language experience facilitates the processing of gender agreement in Spanish heritage speakers." *Bilingualism: Language and Cognition* 17(1). 118–138. https://doi.org/10.1017/S1366728913000114.
Montrul, Silvina, Rebecca Foote & Silvia Perpiñán. 2008. Gender Agreement in Adult Second Language Learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58(3): 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x.
Muñoz, Carmen. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning." *Applied Linguistics* 29(4): 578–596. https://doi.org/10.1093/applin/amm056.
Paolieri, Daniela, Francisca Padilla, Olga Koreneva, Luis Morales & Pedro Macizo. 2019. Gender congruency effects in Russian–Spanish and Italian–Spanish bilinguals: the role of

language proximity and concreteness of words." *Bilingualism: Language and Cognition* 22(1): 112–129. https://doi.org/10.1017/S1366728917000591.
Pliatsikas, Christos & Theodoros Marinis. 2013. Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition* 16(1): 167–182. https://doi.org/10.1017/S136672891200017X.
Puig-Mayenco, Eloi, Jorge González Alonso & Jason Rothman. 2020. A systematic review of transfer studies in third language acquisition. *Second Language Research* 36(1). 31–64. https://doi.org/10.1177/0267658318809147.
Rebuschat, Patrick & John N. Williams. 2012. Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics* 33(4). 829–856. https://doi.org/10.1017/S0142716411000580.
Rothman, Jason. 2011. L3 syntactic transfer selectivity and typological determinacy: The Typological Primacy Model." *Second Language Research* 27(1). 107–127. https://doi.org/10.1177/0267658310386439.
Rothman, Jason. 2013. Cognitive economy, non-redundancy and typological primacy in L3 Acquisition Initial Stages of L3 Romance and Beyond." In Sergio Baauw, Frank Drijkoningen, Luisa Meroni & Manuela Pinto (eds.) *Romance languages and linguistic theory 2011: Selected papers from "Going Romance" Utrecht 2011*. Amsterdam: John Benjamins Publishing Company.
Rothman, Jason. 2015. Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered." *Bilingualism: Language and Cognition* 18(2). 179–190. https://doi.org/10.1017/S136672891300059X.
Rothman, Jason, Jorge González Alonso & Eloi Puig-Mayenco. 2019. *Third language acquisition and linguistic transfer*. Cambridge: Cambridge University Press, 2019. https://doi.org/10.1017/9781316014660.
Sabourin, Laura, Laurie Stowe & Ger de Haan. 2006. Transfer effects in learning a second language grammatical gender system. *Second Language Research* 22(1): 1–29. https://doi.org/10.1191/0267658306sr259oa.
Sá-Leite, Ana Rita, Isabel Fraga & Montserrat Comesaña. 2019. Grammatical gender processing in bilinguals: An analytic review." *Psychonomic Bulletin & Review* 26(4). 1148–1173. https://doi.org/10.3758/s13423-019-01596-8.
Sarnecka, B.W., V.G. Kamenskaya, Y. Yamana, T. Ogura & Yulia B. Yudovina. 2007. From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in English, Russian, and Japanese." *Cognitive Psychology* 55(2). 136–168. https://doi.org/10.1016/j.cogpsych.2006.09.001.
Schwartz, Mila, Miriam Minkov, Elena Dieser, Ekaterina Protassova, Victor Moin, & Maria Polinsky. 2015. Acquisition of Russian gender agreement by monolingual and bilingual children. *International Journal of Bilingualism* 19(6). 726–752. https://doi.org/10.1177/1367006914544989.
di Silvio, Francesca, Wenhao Diao & Anne Donovan. 2016. The development of L2 fluency during study abroad: A cross-language study. *The Modern Language Journal* 100(3). 610–624. https://www.jstor.org/stable/44135009.
The jamovi Project. 2022. Version 2.3.21. https://jamovi.org.

Wang, Qiang. 2014. Gender assignment of Russian indeclinable nouns. Eugene: University of Oregon MA thesis. https://www.proquest.com/docview/1614473606/abstract/5CBD8E6CA17C4C45PQ/1.
Williams, John N. & Chieko Kuribara. 2008. comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling." *Lingua* 118(4). 522–553. https://doi.org/10.1016/j.lingua.2007.03.003.