



Moving Beyond the Native-Speaker Bias in the Analysis of Variable Gender Marking

Aarnes Gudmestad^{1*}, Amanda Edmonds² and Thomas Metzger³

¹Virginia Polytechnic Institute and State University, Department of Modern and Classical Languages and Literatures, Blacksburg, VA, United States, ²Université Paul-Valéry Montpellier 3, Département D'études Anglophones, Montpellier, France, ³The Ohio State University, Department of Statistics, Columbus, OH, United States

In the current study, we respond to calls for reform in second language acquisition that center on the field's preoccupation with native-speaker and prescriptive targets as a benchmark for additional-language learning. In order to address these concerns, we examine the use and development of grammatical gender marking in additional-language Spanish in a prescriptive-independent manner. Specifically, we depart from previous analyses that have centered on accuracy and targetlikeness and we shift the object of analysis to the linguistic forms (i.e., feminine and masculine modifiers) that additional-language participants use. We adopt a variationist approach to explain how participants vary their use of modifier gender and how this use changes longitudinally. We argue that such an approach to studying additional languages allows us to offer new insights about the acquisition of grammatical gender marking in additional-language Spanish. We end by critically reflecting on some of the challenges that we encountered in trying to integrate this paradigm shift into the examination of a well-studied grammatical structure.

Keywords: native speaker bias, variationist approaches, grammatical gender, Spanish, variation

INTRODUCTION

Scholars in applied linguistics have raised concerns about the native-speaker bias that has shaped the field. Ortega (2014), p. 32 writes:

The bias results from the assumption that monolingualism is the default for human communication and from valuing nativeness as a superior form of language competence and the most legitimate relationship between a language and its users. These critiques are poignant in unmasking deeply negative consequences for research and praxis . . . In many of these critiques, the field of second language acquisition (SLA) has been targeted explicitly as suffering in its very core, and in particularly acute ways, from the ailments that result from taking nativeness and monolingualism as natural organizing principles for the study of additional-language learning.

Following these criticisms, Cook and Wei (2016), The Douglas Fir Group (2016), Ortega (2013), Ortega (2017), among others, have advocated for conceptual and methodological reform regarding the role of native-speaker targets in SLA. In the current study, we respond to the call for reform by offering a concrete example of how additional-language¹ data may be profitably analyzed in a prescriptive-independent manner. We conduct an analysis of the development of grammatical

¹We use this term to refer to any language learned after the first language (L1, cf. The Douglas Fir Group, 2016).

OPEN ACCESS

Edited by:

Pedro Guijarro-Fuentes,
University of the Balearic Islands,
Spain

Reviewed by:

Jonathan R. Kasstan,
University of Westminster,
United Kingdom
Ana de Prada Pérez,
Maynooth University, Ireland

*Correspondence:

Aarnes Gudmestad
agudmest@vt.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 10 June 2021

Accepted: 26 July 2021

Published: 09 August 2021

Citation:

Gudmestad A, Edmonds A and
Metzger T (2021) Moving Beyond the
Native-Speaker Bias in the Analysis of
Variable Gender Marking.
Front. Commun. 6:723496.
doi: 10.3389/fcomm.2021.723496

gender marking behavior in additional-language Spanish that crucially does not use a native-speaker or prescriptive norm² as a benchmark against which learner behavior is measured. Namely, rather than analyze accuracy in gender marking, we focus on the forms that participants use – namely, feminine and masculine modifiers – and we use a variationist approach to model the variability in the use of these forms at three points over a 21-month period. We conclude not only with a discussion of the specific ways in which SLA stands to benefit from explanations of developmental trajectories that are independent from considerations of a native-speaker target but also with a reflection on the challenges that come with this paradigm shift.

BACKGROUND

The Native-Speaker Norm in Second Language Acquisition

For years now, many researchers working within SLA have been criticizing the field's preoccupation with native-speaker targets (e.g., Bley-Vroman, 1983; Cook, 1992; Cook, 1997; Cook, 2016; The Douglas Fir Group, 2016; Kinginger, 2009; Klein, 1998; Ortega, 2014; Ortega, 2016; Ortega, 2017; see also; Bachman, 1990; Mauranen, 2012).³ Studies using such targets tend to conduct error analyses or other assessments of accuracy, nativelikeness, or targetlikeness that involve comparing additional-language users to native speakers or prescriptive norms. Some of the arguments against the native-speaker bias are that it implicitly takes a deficit view of additional-language speakers and that it ignores the reality of multilingualism, because “when researchers and educators insist on a monolingual native-speaking golden rule for their interpretations of development, progress, or success, they are setting up L2 [second-language] learners for failure, since multilingual competence is simply different in nature from monolingual competence” (The Douglas Fir Group, p. 35; cf. Hall et al., 2006). In an earlier critique, Bley-Vroman called the comparison of additional-language users to native speakers of the target language the “comparative fallacy” and asserted that it negatively impacts descriptions of additional-language users' linguistic behavior (p. 2). Bachman warned against evaluating additional-language behavior based on native speakers because “native speakers show considerable variation in ability” (p. 39; see also Dabrowska (2012) and Mulder and Hulstijn (2011), for research that offers evidence of variability in linguistic knowledge, oral proficiency, etc. among native speakers). Moreover, another concern about the impact that the native-speaker bias has had

on SLA is that because of this bias, SLA has little standing in the language sciences. Klein argued that with a focus on “learners' utterances as deviations from a certain target, instead of genuine manifestations of underlying language capacity ... [SLA] analyses them in terms of what they are not rather than what they are” (p. 527). The consequence of this focus is that the observations that emerge from SLA research are not essential to theoretical advancements in linguistics more generally (Klein, 1998, p. 530).

Despite the concerns raised by various leaders in the field with the native-speaker bias in SLA, comparisons of additional-language data to native-speaker or prescriptive baselines remain commonplace (Ortega, 2016). Ortega (2014) has advocated for the need to “replace SLA's existing research goal of explaining why late bi/multilinguals are not native speakers ... with the goal of understanding the process and consequences of becoming bilingual or multilingual later in life” (p. 33). This conceptual shift arguably necessitates methodological changes. Perhaps the best-known proposal for navigating this change has come from Cook (e.g., 2016), whose multicompetence approach is offered as a concrete way to carry out SLA research without referencing a native or prescriptive norm. One of the hallmarks of this approach is the need to study the full linguistic repertoire of bi/multilingual speakers – not only any additional languages that speakers may use, but also their L1. Thus, the multicompetence approach necessitates a dramatic shift in research design, which may in part explain why it has not yet been fully embraced in the field. Alongside proposals for more radical design changes, we believe that there is great potential to move beyond the preoccupation with prescriptive and native-speaker comparisons by drawing on frameworks currently in place within SLA. In this vein, we note a call for reform by Ortega (2017) who welcomes “empirical research on all alternatives ... because it will make us unearth new knowledge about L2 development” (see also Ortega, 2016). Thus, the overarching goal of the current study is to respond to this call for conceptual and methodological reform with an analysis that takes advantage of one theoretical framework that has been fruitfully applied to SLA research, namely variationism. As we will detail in the following section, the variationist approach offers many strengths that make it compatible with the current endeavor. In the present study, we draw on certain aspects of variationism to study grammatical gender in Spanish, a linguistic phenomenon that to our knowledge has yet to be investigated without comparing additional-language users to a native-speaker target or prescriptive baseline. The analysis we offer is a reanalysis of the data examined in Gudmestad et al. (2019), in which grammatical gender marking was analyzed with reference to a prescriptive norm. With this reanalysis, we hope to show how existing tools in SLA can be valuable in helping the field to fully realize this conceptual and methodological shift.

Variationist Second Language Acquisition

Variationist sociolinguistics is an area of scholarship that focuses on variation and change in language (Labov, 1966). Variable structures, which refer to cases where a single language function

²A native-speaker target can either refer to an idealized norm or a target that reflects real-world language behavior of native speakers, whereas a prescriptive norm refers to an idealized norm. In the current article, we use both terms because the native-speaker target and prescriptive norm for grammatical gender marking in Spanish, the linguistic phenomenon under investigation, are largely identical (cf. Gudmestad et al., 2019).

³Research on language revitalization and new speakers has voiced similar concerns (e.g., O'Rourke et al., 2015) and addressed issues such as speaker motivation and identity construction (e.g., Nance et al., 2016).

can be expressed by two or more forms, are the object of study. One example is subject expression in Spanish, where a speaker can use either a pronoun (*yo*) or an unexpressed subject to express the first-person singular (see example 1). Researchers typically use multivariate and quantitative analyses in order to identify the multitude of linguistic and social (extra-linguistic) factors that influence the variable occurrence of a given form. Returning to subject expression in Spanish, de Prada Pérez (2015) examined first-person subjects in a native-speaker dataset. When she analyzed all of her data together,⁴ she found that three linguistic factors and two social factors simultaneously predicted the use of first-person subjects. For example, the linguistic factor of verb form ambiguity⁵ was significant. The probability of using a personal pronoun was higher with ambiguous verbs, whereas the probability of using an unexpressed subject was higher with unambiguous verbs. The results for the social factor of age showed that older speakers favored the use of *yo* and younger speakers favored the use of unexpressed subjects.

- 1) *Yo no tengo tiempo.* ‘I don’t have time.’
 Ø *no tengo tiempo.* ‘(I) don’t have time.’

As an extension of this work, variationist SLA “explores the relationship between contextual variables (both social and linguistic contextual variables) and variation in the form of learner language” (Tarone, 2007, p. 845). Given its connection with sociolinguistics, most variationist scholarship in SLA has examined linguistic phenomena that are sociolinguistically variable among native speakers (e.g., Mougeon et al., 2010). This has been called Type 2 variation (Rehner, 2002). However, this approach has also been adopted to study linguistic structures for which learners exhibit variability but native speakers do not (i.e., Type 1 variation, Rehner, 2002), such as preposition + article contractions in Portuguese (Picoral and Carvalho, 2020) and plural marking in English (Young, 1991). Researchers often examine cross-sectional (e.g., Kanwit, 2017) or longitudinal (e.g., Regan et al., 2009) data to understand how additional-language learners’ variable behavior changes along the developmental trajectory. Because there exists a substantial body of variationist SLA research (cf. Geeslin and Long, 2014), we offer an example of a study that used this approach to illustrate what a variationist SLA analysis looks like and to demonstrate what such an analysis of learner data can contribute to SLA.

In Geeslin et al. (2012), the researchers examined the development of perfective past-time reference (i.e., the variation between the preterit and the present perfect to express reference to the perfective past) among additional-language learners who studied in Spain for 7 weeks and a group of native speakers of Peninsular

Spanish (a case of Type 2 variation). They analyzed data from a written contextualized task in which participants selected whether they liked the use of the preterit, present perfect, or both forms in specific contexts. The additional-language learners completed the task three times (during weeks 1, 4, and 7 of the study-abroad program) and the native speakers completed it once. Although each item on the written contextualized task presented learners with three response options (preterit, present perfect, both forms), for most of the analysis, the researchers analyzed the present perfect and both responses together as a “present perfect allowed” category. We focus here on the part of the analysis that centered on the predictive factors. In order to examine the factors that predicted the variable selection of perfective verb forms, the researchers performed four regression models – one for each data-collection point for the learners and one for the native speakers. The three learner models enabled the researchers to make observations about additional-language development longitudinally, and assessments of targetlikeness were made by comparing the learner and native-speaker models. One finding, for example, was that years of study, an extra-linguistic variable, impacted verb selection on the task at weeks 1, 4, and 7 for the learners, such that participants who had been studying Spanish for 5 years or more selected more present perfect than those who had studied Spanish for 4 years or fewer. Another result was that the telicity⁶ variable, a linguistic factor, was significant for learners at weeks 1 and 4 but not at week 7 or for the native speakers; at weeks 1 and 4 the preterit was more likely than the present perfect in telic contexts.

It is important to note that Geeslin et al. (2012) analyzed native-speaker data, which they then used as a baseline against which to compare learner data. Although the use of native-speaker (or bilingual, e.g., Kanwit, 2017) benchmarks is typical in variationist SLA, we argue that it is not an essential component and that variationism can be fruitfully used to study additional languages without reference to a native or prescriptive norm. Three important characteristics of variationist SLA, all exemplified in Geeslin et al. (2012), show the potential of this approach. The first is that linguistic forms are the object of study; for example, Geeslin et al. (2012) focus on perfective past verb forms. This attention to understanding the occurrence of different forms (instead of, for example, investigating accuracy) makes variationism particularly promising for additional-language research that attempts to avoid comparisons with a prescriptive norm. In this vein, we aim to apply variationist tools in order to analyze the development of grammatical gender marking in additional-language Spanish by focusing on the forms that participants use. We return to the question of the forms that we analyze in the next session.

Second, at the heart of the variationist approach to SLA is the goal of understanding systematic variability in language. In Geeslin et al. (2012), the researchers showed that, rather than selecting a single verb form categorically in certain contexts, the learners’ behavior was variable. For example, at weeks 1 and 4, when telicity significantly influenced verb selection, the learners did not select the preterit 100% of the time in telic contexts. Instead, they were more likely to choose the preterit in these contexts but the present perfect was also possible.

⁴She began her analysis by examining four speaker groups together: Spanish L1 bilinguals, Catalan L1 bilinguals, a Spanish control group, and a Catalan control group.

⁵de Prada Pérez (2015), p. 125 coded a verb form “as ambiguous if it was morphologically ambiguous with another [grammatical-person] form”.

⁶Telicity refers to whether a predicate has an endpoint (Geeslin et al. (2012) p. 203).

We argue that the focus on variation that characterizes variationist SLA is a strength of this approach that makes it an excellent candidate for attempts to respond to current calls for reform in SLA using established frameworks. The Douglas Fir Group contends, “[v]ariability is not measurement error begging for better control. Acknowledging inter-as well as intra-individual variation helps counter deficit orientations in the description of linguistic development in an L2 . . . and focus on what learners can do rather than what they cannot do” (2016, p. 30). Variationist SLA not only recognizes the presence of variability, it provides conceptual and methodological tools for explaining the complex systematicity and dynamicity of additional-language variation, which leads us to the third characteristic.

The variationist approach to SLA results in detailed observations about the dynamicity of learner language. This is accomplished through the use of multivariate analyses that offer explanations of how the linguistic (e.g., telicity) and extra-linguistic (e.g., years of Spanish study) factors that predict the occurrence of linguistic forms can change over time, thus furthering knowledge about additional-language development. In Geeslin et al. (2012), whereas the impact that years of study had on perfective past reference was stable over time, change in learner behavior was observed with telicity between weeks 4 and 7. Importantly for the current study, both this stability and this change can be observed without relying on a native-speaker baseline. Therefore, in the current study, we aim to examine how the additional-language use of grammatical-gender forms changes (or not) longitudinally. Crucially, however, we do not compare the additional-language data to a native-speaker benchmark or prescriptive norm.

Grammatical Gender in Spanish

In Spanish nouns have either feminine or masculine gender. Gender assignment is arbitrary for most nouns (e.g., *bicicleta*_{fem} ‘bike’, *coche*_{masc} ‘car’), though biological sex determines the gender of some nouns (e.g., *hija*_{fem} ‘daughter’, *hijo*_{masc} ‘son’). Descriptively, whereas some adjectives and determiners have a single form that is used with nouns of both genders (e.g., *mi* ‘my’ and *verde* ‘green’, as in *mi manzana*_{fem} *verde* ‘my green apple’ and *mi melón*_{masc} *verde* ‘my green melon’), most have different feminine and masculine forms (e.g., *una*_{fem} *manzana*_{fem} *amarilla*_{fem} ‘a yellow apple’ and *un*_{masc} *melón*_{masc} *amarillo*_{masc} ‘a yellow melon’). Some endings are linked with one gender. The canonical endings for nouns and modifiers are *-a* for feminine and *-o* for masculine. However, there are exceptions, such that nouns ending in *-a* can be masculine (e.g., *poema*_{masc} ‘poem’) and those ending in *-o* can be feminine (*nao*_{fem} ‘ship’). There are other endings that are either strongly connected to one gender (e.g., *-dad* as in *libertad*_{fem} ‘freedom’ and *-e estante*_{masc} ‘shelf’) or that are not linked with a particular gender (e.g., *-s* as in *tos*_{fem} ‘cough’ and *mes*_{masc} ‘month’; Teschner and Russell, 1984).

Grammatical gender in additional-language Spanish has been studied extensively, and to our knowledge all of this work has been oriented toward a native-speaker or prescriptive benchmark (cf. Alarcón, 2014). This means that the focus has been on accuracy or targetlikeness, where a mismatch in the gender of a noun and its modifier constitutes an error or an instance of non-targetlike behavior.

Previous research has sought to better understand how additional-language users of Spanish produce and process gender marking (Alarcón, 2014). We limit our review, however, to production studies since we analyze language use in the present investigation. We focus on the variables identified in this research that explain the development of targetlike grammatical gender marking. Noun gender, noun ending, modifier type, and noun class have been the most widely studied factors in gender-marking research. Findings have shown that learners tend to exhibit more accurate gender marking with masculine nouns (e.g., Finnemann, 1992; White et al., 2004; Montrul et al., 2008), which has been interpreted to indicate that masculine modifiers are a default form that develop more quickly than feminine modifiers. Studies have also shown that nouns that have the prototypical *-o* ending for masculine nouns and *-a* ending for feminine nouns are connected to higher rates of accuracy (e.g., Fernández-García, 1999; Alarcón, 2011). Various investigations (Bruhn de Garavito and White, 2002; White et al., 2004; Alarcón, 2010), though not all (e.g., Montrul et al., 2008; Alarcón, 2011), have found that learners mark gender more accurately on determiners than adjectives (i.e., modifier type). Regarding noun class, studies have demonstrated differing results, with some reporting higher accuracy rates with arbitrary gender (e.g., Bruhn de Garavito and White, 2002) and others showing that learners are more accurate with biological gender (e.g., Fernández-García, 1999). Other factors investigated include noun number, where Finnemann (1992) found that learners exhibited fewer errors with singular compared to plural nouns, and course level or proficiency, about which it was revealed that learners became more accurate with grammatical gender as course level or proficiency increased (e.g., Bruhn de Garavito and White, 2002; Montrul et al., 2008).

Moreover, a recent study examined the development of targetlikeness in grammatical gender marking longitudinally by bringing together the previously studied independent variables and five new factors. In addition to the aforementioned six variables, Gudmestad et al. (2019) analyzed the number of syllables between the noun and the modifier, task (oral interview, oral narration, written essay), time (before study abroad, during study abroad, after study abroad), and two factors that assessed noun frequency: noun log-frequency (language) and noun frequency (individual). The factor noun log-frequency (language) provided a measure of noun frequency in Spanish using the *Corpus del español* (Davies, 2016-), whereas the factor noun frequency (individual) provided a measure of noun frequency that depended on each individual’s use (see the *Methods* section for details on the participants, the data collection, and the full set of variables, as the current study constitutes a reanalysis of Gudmestad et al. (2019)). A generalized linear mixed-effects model revealed that noun ending, task, noun gender, noun frequency (individual), syllable distance, modifier type, initial proficiency, time, and the interaction between noun ending and time simultaneously predicted targetlike gender marking in language production. Similar to previous investigations, we found that learners were more likely to be targetlike with gender marking with canonical *-o/-a* endings, masculine nouns, and determiners and that higher scores on a proficiency test were also linked to higher rates of targetlike use. Several novel findings came out of this investigation: Learners exhibited higher log-odds of targetlike use on a written essay compared to oral tasks, at the in-stay and post-stay data-collection points compared to the pre-stay time, as the distance between the

noun and the modifier decreased, and as the frequency with which each individual used a noun with a modifier overtly marked for gender increased. Finally, the interaction between noun ending and time indicated that the learners made gains in marking gender with nouns that have what were called “deceptive” endings, that is, nouns that ended in either *-o* or *-a*, but whose gender did not correspond to the canonical gender for that ending (e.g., *poema*_{masc} ‘poem’, *nao*_{fem} ‘ship’).

In the current study, we reanalyze our previous work in order to offer a reconceptualization of additional-language data in a prescriptive-independent manner by shifting the focus of analysis from targetlikeness to the forms that participants use. More specifically, instead of analyzing how targetlike learners are in their marking of gender, we analyze their use of feminine and masculine modifiers through the lens of systematic variation. Indeed, gender-marking behavior consists of making a (conscious or unconscious) choice between the masculine and feminine forms of a given modifier; previous research has demonstrated that learners’ use of modifier gender is not categorical, which means that certain nouns may be used variably with both feminine and masculine marked modifiers (Dewaele and Véronique, 2001). This variability constitutes a case of Type 1 variation, as gender marking in Spanish (with the exception of certain nouns whose gender varies by geographical region) has not been shown to vary sociolinguistically among native speakers (Gudmestad et al., 2019). Thus, in line with variationist SLA, there is value in modeling variability in the use of these linguistic forms. With the present analysis, we shift the focus from how correct additional-language users are to what predicts their use of modifier gender.

THE CURRENT STUDY

We address the following question in the present investigation: What linguistic and extra-linguistic factors predict the variable use of modifier gender over time? In order to answer this question, we conduct a variationist analysis of gender marking in additional-language Spanish that attempts to be independent from a native-speaker or prescriptive norm. Instead of determining to what extent learners approximate a targetlike norm, the findings contribute insight into factors that predict the use of modifier gender and whether and how these factors change over time. After presenting the results, we discuss the new knowledge about grammatical gender that emerges from this type of analysis, and then we reflect on what an approach to additional-language gender marking that moves away from a native-speaker norm brings to SLA. Specifically, we consider the more general impact and some of the challenges of such an approach on research within the field of SLA.

METHODS

Corpus and Participants

Our data come from LANGSNAP, a publicly available corpus (<http://langsnap.soton.ac.uk>, e.g., Mitchell et al., 2017). For this corpus the research team collected data from additional-language

speakers of French and Spanish over a period of 21 months, which included an academic year abroad in a French- or Spanish-speaking country. The data were collected at six points in time: Before the participants went abroad, three occasions while they were abroad, and twice after returning to the United Kingdom.⁷ At each data-collection period, the participants completed an oral interview in which they talked about their lives, an oral picture-based narration task, and a written argumentative essay. We report on half of data-collection periods for the Spanish data. We analyzed all three tasks at three different data-collection periods: Before the participants went abroad (henceforth, pre-stay), the third data collection while abroad that occurred at the end of their stay abroad and 1 year after pre-stay (in-stay), and the final data collection in the United Kingdom that took place 21 months after the initial data collection and about 8 months after returning home (post-stay).

Our dataset consists of 21 of the 27 undergraduate students included in the corpus who were pursuing an undergraduate degree in Spanish in the United Kingdom.⁸ They ranged in age from 20 to 25 years ($M = 20.8$, $SD = 1.6$). Fifteen were women and six were men. When the project began, they had been studying Spanish for an average of 5.4 years ($SD = 3.4$, $range = 2-14$). The L1 of the participants was English ($n = 19$), Polish ($n = 1$), or both Polish and English ($n = 1$). In terms of other languages that the participants had studied, 18 indicated that they had studied French, German and/or Italian, two had not learned another language, and one opted not to share this information. During the participants’ academic year abroad, they were teaching assistants ($n = 10$), exchange students, ($n = 9$), and workplace interns ($n = 2$), and five lived in Mexico whereas 16 were in Spain.

Data Coding and Analysis

We began the coding by identifying every referent that was modified by a determiner or an adjective ($K = 16,357$).⁹ The tokens retained for the dataset that we analyzed in the present investigation ($k = 11,351$)¹⁰ shared three characteristics: 1) the referent for each token was a noun (instances involving pronouns were coded, but not analyzed in this project), 2) only nouns that occurred more than once in the dataset were analyzed (a total of 482 nouns occurred a single time and were thus removed from our dataset), and 3) each token involves a modifier that exhibited overt gender marking, meaning that the modifier had distinct

⁷The three in-stay periods were collected five, nine and 12 months after the pre-stay data collection. Additionally, see Tracy-Ventura and Huensch (2018) for a presentation of later phases of the project.

⁸We have coded and analyzed a subset of the data-collection points and of the participants because the coding, which was done entirely by hand, was very labor intensive. Additionally, the three data-collection periods that we coded enabled us to make observations about possible change over the course of an academic year abroad (in-stay) and whether any changes held after the participants had returned home (post-stay).

⁹For the oral data, we relied on the transcripts provided online by the LANGSNAP team. After the transcription was completed initially, the transcripts were checked by at least one other member of the LANGSNAP team.

¹⁰In Gudmestad et al. (2019), our dataset consisted of more observations because we analyzed nouns that occurred once and those that occurred multiple times. This previous analysis did not include a random effect for noun type.

masculine and feminine forms. Examples from the data are available in (2), with the nouns underlined and the modifiers in bold. The examples in (2) illustrate variable use of modifier gender (both feminine and masculine marked modifiers) with a single noun.

2) *Voy a estudiar por **la**_{fem} día* ‘I’m going to study during the day’ (participant 166, pre-stay, interview)

*Iban a casa **todos**_{masc} **los**_{masc} días* ‘they went home every day’ (participant 152, in-stay, narrative).

Whereas we originally analyzed this dataset using targetlikeness as the dependent variable (Gudmestad et al., 2019), in the current analysis, we sought to move away from using the prescriptive norm as a yardstick. For this reason, the current dependent variable was modifier gender (feminine or masculine). We coded for nine independent, fixed-effect variables that had been studied previously in research on grammatical gender marking in additional-language Spanish (see the **Supplementary Materials** for a table that lists the variables and their categories). Although these factors have been examined in prior investigations in order to better understand targetlike use, we explored their potential impact on the use of modifier gender in the current study.

Five factors pertained to characteristics of the noun. First, noun number differentiated between singular and plural nouns. Second, noun class distinguished between nouns that have biological gender (e.g., *mujer*_{fem} ‘woman’) and those that have arbitrary gender (e.g., *lápiz*_{masc} ‘pencil’). Third, each token was coded according to the ending seen on the noun. We distinguished four categories for the variable of noun ending. Canonical *-o/-a* endings were masculine nouns ending in *-o* and feminine nouns ending in *-a*. Non-canonical *-o/-a* endings were the opposite – feminine nouns ending in *-o* and masculine nouns ending in *-a*. Predictive endings were those that, according to Teschner and Russell (1984), were strongly linked with one gender (e.g., *-ción*, as in *educación*_{fem} ‘education’ and *-e*, as in *pie*_{masc} ‘foot’, as feminine and masculine endings, respectively). Other endings were those that were not strongly connected to one gender (e.g., *-s* as in *país*_{masc} ‘country’ and *tesis*_{fem} ‘thesis’; Teschner and Russell, 1984). The final two factors that targeted characteristics of the noun were included in order to examine the possible role that noun frequency plays in the use of modifier gender. The factor noun log-frequency (language) provided a measure of noun frequency in Spanish and, as such, is taken as a proxy for possible input. For this factor, we identified the frequency per million words with which each noun in our dataset occurred in the *Corpus del español* (Davies, 2016-). Because of the skew in the distribution of frequency scores, we used the natural logarithm of noun frequency in our analysis. The factor noun frequency (individual) provided a measure of noun frequency that depended on each individual’s use. Usage-based research, which demonstrates that an individual’s language use shapes her/his internal grammar (e.g., Bybee, 2006), motivates this variable. For this factor, we counted how often each participant produced a given noun with a gender-marked

modifier. Because individual speakers can change their use of nouns and modifiers as a function of task and time, we calculated this score for every individual, each task, and at each data-collection period. Therefore, this coding gives the possibility of nine different frequency scores (three tasks x three data-collection periods) for a given noun for every participant. We examined the possible role that frequency plays in the use of modifier gender because previous research suggests that noun frequency influences additional-language gender marking (e.g., Sabourin et al., 2006); two different frequency factors were included because the operationalization of frequency is complex (Hashimoto and Egbert, 2019).

The remaining four variables pertained to characteristics that did not concern solely the noun. Syllable distance measured the number of syllables between the modifier and the noun. Modifier type differentiated between determiners and adjectives. For the final fixed effects, time distinguished between the pre-stay, in-stay, and post-stay data-collection periods and task analyzed possible differences among the oral interview, oral narration, and written essay. It is important to note that, while noun gender has been widely studied in investigations on grammatical gender, we have not included it in the present analysis because it can be interpreted to represent a native-speaker or prescriptive benchmark. Namely, the examination of noun gender as a fixed effect and modifier gender as the dependent variable would allow for observations about targetlike use, as the results would show, for example, whether feminine modifiers were more likely to occur with feminine nouns (i.e., targetlike use) or masculine nouns (i.e., non-targetlike use). In other words, including noun gender as a fixed effect could be considered an indirect inclusion of a native-speaker or prescriptive norm. Given the overarching goal of the current study, which is to move beyond assessments of targetlikeness or accuracy in the study of grammatical gender, we elected not to analyze this factor.

Finally, we included participant and noun type as random effects in the analysis. The participant random effect enables us to account for variability among the participants and the noun-type random effect recognizes that language behavior with individual nouns may differ. By including these two variables as random effects, we treated participants as part of a larger population of speakers and noun types as a part of a larger vocabulary. The inclusion of noun type as a random effect explains why the current dataset is limited to nouns that were used more than once, as we cannot distinguish how much variability in usage can be attributed to nouns that occur only once in the dataset.

We analyzed the data quantitatively using R (R Core Team, 2019) and SAS software, Version 9.4 of the SAS System for Windows (Copyright © 2018 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, United States). First, with R, we examined whether there were strong correlations among any of the fixed effects using chi-square methods, with the intention to remove variables from further analysis when strong correlations were observed. With bootstrapping, we also explored whether any of the fixed effects appeared to be important for explaining variation in the dependent variable. This step enabled us to remove variables that

were not important and helped us to avoid overfitting the model. Next, we fit one generalized linear mixed-effects model employing a backward selection strategy with SAS. The model was fit with respect to feminine usage. We focused on feminine modifiers because previous research has indicated that the use of feminine and masculine modifiers develops at different rates, with feminine modifiers developing more slowly (e.g., Finnemann, 1992; Montrul et al., 2008). However, it is worth clarifying that all effect estimates in this model refer to how feminine modifiers behave in relation to masculine modifiers. Thus, because the dependent variable is binary, information about the use of masculine modifiers is also present in the models. This regression examined multiple independent variables concurrently and determined which ones conditioned the use of modifier gender. If a variable was found to be non-significant, we removed it from the analysis and reran the model. For nominal independent variables, one category is selected as the reference point and compared to the other category or categories of the same variable. The reference points for the categorical fixed effects were singular (noun number), arbitrary (noun class), canonical *-o/-a* (noun ending), determiner (modifier type), pre-stay (time), and essay (task). Noun log-frequency (language), noun frequency (individual), and syllable distance were continuous factors, so they had no reference point. Once we identified the significant fixed effects, we examined interactions between these factors and time, in order to make observations about language development over the 21-month period covered by the LANGSNAP corpus. After fitting the model, we assessed whether any of the fixed effects were highly correlated, which would have led to instability of effect estimates. We considered a magnitude of greater than 0.6 to be highly correlated. We also identified the McFadden's R^2 (with R) and the Bayesian Information Criterion (BIC, calculated with SAS), two metrics that indicate whether the generalized linear mixed-effects model does a good job of modeling the data.

RESULTS

We began the analysis by examining whether any of the fixed effects were highly correlated. No such strong correlations were found. We then moved on to the bootstrapping phase, which revealed that noun number and noun log-frequency (language) appeared not to be important for the use of modifier gender, so we removed them from further investigation. Then we fit a generalized linear mixed-effects model. Noun frequency (individual) and noun class were not significant in this model, so we removed them and reran the model. **Table 1** shows an overview of the fixed effects included in the final regression model. The significant fixed effects were noun ending, syllable distance, task, and modifier type. Time is also included in the model, though it is not significant, because we were interested in exploring interactions between time and the fixed effects. Interactions between time and other fixed effects allow us to make observations about longitudinal development. One significant interaction was identified: Time x noun ending.

TABLE 1 | Overview of the fixed effects in the regression model.

Effect	Df	F	p
Time	2	0.41	0.6064
Noun ending	3	4.23	0.005
Syllable distance	1	9.16	0.003
Task	2	3.60	0.027
Modifier type	1	25.45	<0.0001
Noun ending x Time	6	4.439	0.0002

After fitting the model, we examined whether there were strong correlations between any of the fixed effects and found none.

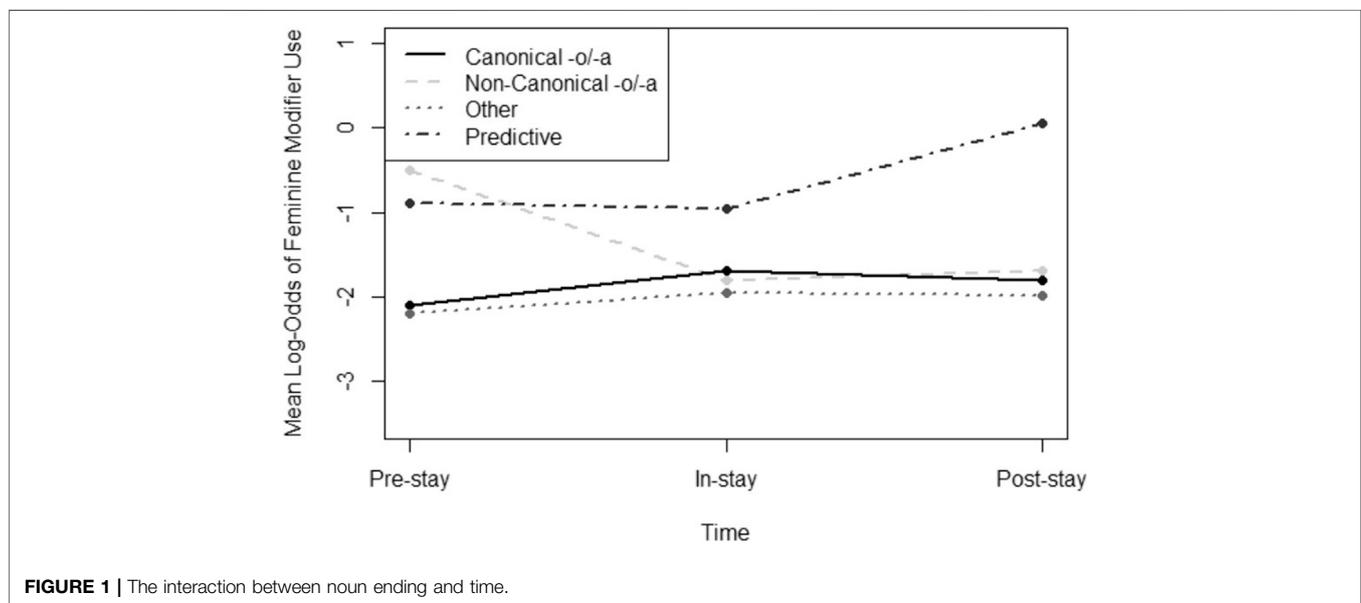
Table 2 provides the details of the results for fixed effects and the interaction; the results for the two random intercepts (participant and noun type) are available as **Supplementary Materials**. We focus our presentation of the findings on three pieces of information in the tables: The estimate, p value, and confidence interval (CI). A positive estimate indicates a higher log-odds of using a feminine modifier compared to a masculine modifier and a negative estimate means that the log-odds of using a feminine modifier are lower. Results for the use of masculine modifiers can be inferred from these results: If we were to rerun the analysis where masculine modifiers were treated as the reference point, we would obtain effect estimates that are of the same magnitude as what we report in **Table 2**, but of opposite sign. In other words, a positive estimate for feminine modifiers would be a negative estimate for masculine modifiers and vice versa. The p value shows whether the result is significant (in the current analysis $\alpha = 0.05$ and significance is when $p < \alpha$). For nominal fixed effects that have more than two categories (time, noun ending, and task), we also look to the CIs to see whether the non-reference point categories are similar to or different from each other. Overlap in CIs indicates a similarity, whereas the lack of overlap points to an important difference.

Beginning with time, although we saw that the overall F test is not significant (**Table 1**), the pairwise comparison of pre-stay versus in-stay was significant (**Table 2**). This indicates that collectively the variability in data-collection periods was similar, but individually two of the points were different from each other. Namely, the log-odds of using a feminine modifier were higher at in-stay than at pre-stay. There was no significant difference in the use of feminine modifiers between post-stay and pre-stay. Moreover, the overlap in the CIs for in-stay and post-stay demonstrate that the use of feminine versus masculine modifiers was similar between these two data-collection periods. For noun ending, the log-odds of using a feminine modifier were significantly higher with both non-canonical *-o/-a* and predictive endings compared to canonical *-o/-a* endings. Other endings were not significantly different from canonical *-o/-a* endings. The examination of the CIs of the non-reference point categories showed that the log-odds of feminine modifier use were similar to each other. Regarding syllable distance, the log-odds of using a feminine modifier decreased as the distance between the noun and the modifier increased. For task, the log-odds of feminine-modifier use were

TABLE 2 | Details of the fixed effects in the regression model.

Effect	Estimate	SE	Df	t	p	CI
Intercept	-1.588	0.391	20	-4.06	0.0006	[-2.404, -0.772]
Time [pre-stay]						
in-stay	0.537	0.161	10,605	3.34	0.0008	[0.222, 0.852]
post-stay	0.354	0.191	10,605	1.85	0.0643	[-0.021, 0.728]
Noun ending [canonical -o/-a]						
non-canonical -o/-a	1.699	0.798	10,605	2.13	0.0333	[0.134, 3.263]
predictive	1.504	0.583	10,605	2.58	0.0099	[0.361, 2.647]
other	-0.036	0.531	10,605	-0.07	0.9455	[-1.076, 1.004]
syllable distance	-0.058	0.019	10,605	-3.03	0.0025	[-0.096, -0.021]
Task [written]						
interview	-0.325	0.177	10,605	-1.84	0.0660	[-0.671, 0.022]
narrative	-0.662	0.248	10,605	-2.66	0.0077	[-1.149, -0.175]
Modifier type [determiner]						
adjective	-0.583	0.116	10,605	-5.04	<0.0001	[-0.809, -0.356]
Noun ending x time [canonical -o/-a and pre-stay]						
non-canonical -o/-a x in-stay	-1.831	0.417	10,605	-4.39	<0.0001	[-2.642, -1.013]
non-canonical -o/-a x post-stay	-1.547	0.529	10,605	-2.92	0.0035	[-2.584, -0.509]
other x in-stay	-0.143	0.278	10,605	-0.52	0.6065	[-0.688, 0.402]
other x post-stay	-0.116	0.346	10,605	-0.34	0.7375	[-0.794, 0.562]
predictive x in-stay	-0.600	3.462	10,605	-1.73	0.0832	[-1.278, 0.079]
predictive x post-stay	0.375	0.433	10,605	0.87	0.3862	[-0.474, 1.224]

Note. The model fits the log-odds of the usage of feminine modifiers. The reference points for the nominal fixed effects and the interaction are in brackets.



lower in the narrative task than in the written essay. The difference between the interview and the essay was not significant, and a comparison of the CIs for the narrative task and the interview show that the use of modifier gender was similar between the two. Additionally, the log-odds of using a feminine modifier were lower with adjectives compared to determiners (the modifier type factor).

Turning to the significant interaction, the plot in **Figure 1** illustrates the findings shown in **Table 2** for the noun ending \times time interaction. A significant change over time was observed in the log-odds of feminine modifier use with non-canonical -o/-a

nouns: The comparisons between pre-stay and in-stay and between pre-stay and post-stay show that use of feminine modifiers with such nouns (as compared to the reference points) became less likely. In sum, this participant group's variability in the use of modifier gender was conditioned by time, noun ending, syllable distance, task, modifier type, and the interaction between time and noun ending.

Finally, both the McFadden's R^2 (Smith and McKenna, 2013) and the BIC (Kass and Raftery, 1995) show that this generalized linear mixed-effects model does a good job of fitting the data for modifier gender. McFadden's R^2 is a measurement of the relative likelihood of

the fitted and null models. This metric indicates a strong model fit ($R^2_{\text{McFadden}} = 0.672$). Next, the BIC is a metric that compared our model to a null model using their log-likelihood. It also penalizes models that have potentially extraneous parameters, which helps to protect against overfitting a model. Our model has a BIC of 5,107.73 and the null model has a BIC of 15,569.31. A difference between the two models that is larger than 10 is deemed to be strong evidence for the model with the smaller BIC. Thus, we can conclude that our model, with its lower BIC, has a higher probability of being the true model when considered against the null model.

DISCUSSION

We now return to the research question: What linguistic and extra-linguistic factors predict the variable use of modifier gender over time? We first answer this question with a discussion of the insights into the acquisition of grammatical gender marking that have emerged from the variationist analysis. We then turn to a reflection on implications of and challenges associated with SLA research that seeks to divorce itself from native-speaker and prescriptive biases.

Insights into the Use of Modifier Gender Using a Variationist Approach

Unlike previous research on the acquisition of grammatical gender marking, the present investigation did not incorporate a native-speaker or prescriptive target (cf. Alarcón, 2014). In particular, we sought to understand what factors influence the variable use of modifier gender as opposed to understanding what factors influence targetlike behavior. This methodological and conceptual decision has resulted in a new knowledge base that crucially differs from the one that has been built on a focus on accuracy. Collectively, we found that participants' variable use of modifier gender is complex and conditioned by multiple factors simultaneously. The mixed-effects model, which accounted for variability in gender-marking behavior among individual participants and noun types by including them as random effects, revealed that the use of feminine modifiers was predicted by 1) the linguistic factors of noun ending, syllable distance, and modifier type, 2) the extralinguistic factors of task and time, 3) and the interaction between noun ending and time. Among those factors, we see several indications of stability over time. Over the 21 months examined in the present investigation, participants were more likely to use feminine modifiers when little distance separated the noun from the modifier, when the modifier was a determiner (versus an adjective), and on the written essay (versus the oral narrative). Taken together, these findings show that, despite a change in learning context, there was stability over time in the factors influencing their use of modifier gender. This stability echoes research that has found that additional-language learners do not always show changes in grammar during a stay abroad (Llanes, 2011).

Stability, however, is not the whole story, as the results also showed evidence of change over time in two ways. One is that the participants were more likely to use feminine modifiers at in-stay compared to pre-stay, which points to an increase in the use of feminine modifiers over the course of an academic year in a

Spanish-speaking country. Keeping in mind previous research that observed feminine-modifier use to develop more slowly than masculine-modifier use, leading researchers to suggest that the masculine modifier is the default (e.g., Montrul et al., 2008), our finding may be indicative of the fact that during an academic year abroad, the strength of the masculine default weakened as the participants became more likely to use feminine modifiers. The other evidence of development is seen with the interaction between noun ending and time: The participants were less likely to use feminine modifiers with nouns that have non-canonical *-o/-a* endings over time. In other words, the log-odds of using feminine modifiers with nouns with non-canonical *-o/-a* endings, such as *problema* 'problem', *día* 'day', *mano* 'hand', changed significantly after an academic year in Spain or Mexico, and this change was maintained after their return to the United Kingdom. Despite evidence of stability during the 21-month period, which included a stay abroad, some development in their gender-marking behavior was observed.

Thus, this variationist analysis was able to offer new details about the additional-language development of gender marking by beginning to explain the variability present in the use and development of modifier gender. In particular, we drew on three features of variationist SLA in order to address this issue. One is that the object of study was the linguistic forms (feminine and masculine modifiers) that learners used. The second characteristic was that we aimed to explain variability in modifier gender by conducting a multivariate analysis that revealed how a range of linguistic and extra-linguistic factors influenced learners' systematic variable use. Third, with the help of the LANGSNAP corpus' longitudinal data, we investigated additional-language development by exploring whether the factors that impacted the use of modifier gender changed over time. Thus, despite the assumption of consensual norms that underpins much variationist research on instances of Type II variation, our analysis offers a proof of concept for the fruitful extension of the variationist framework to analyses that remain independent of this norm.

Challenges and Implications

Before concluding we offer a reflection on some of the challenges we faced in trying to do an analysis that is independent from a native-speaker or prescriptive norm. We comment on the decisions we made in response to these challenges, which we believe may have implications for future research in SLA.

The first challenge we confronted was precisely *how* to conduct an analysis of gender marking that moved away from native-speaker and prescriptive standards for additional-language learning. Whereas scholars advocating for this paradigm shift within the field of SLA have presented convincing theoretical and conceptual arguments (e.g., Bley-Vroman, 1983; Klein, 1998; Ortega, 2016), there is still work to be done in order to work out the concrete details as to how to go about conducting this type of analysis.¹¹ Following a call for reform that encouraged different ways of achieving this goal (Ortega, 2017), we chose to

¹¹See Murahata et al. (2016) for a methodological discussion within the multicompetence approach.

explore how an existing approach to SLA could be useful, and we opted for the variationist approach. Variationist SLA provided us with a framework in which our object of study shifted from an assessment of accuracy or targetlikeness, as was typical in previous research on gender marking in additional-language Spanish (cf. Alarcón, 2014), to an examination of the forms that participants used (i.e., feminine and masculine modifiers). Unlike traditional variationist SLA scholarship, however, we did not compare the additional-language participants in the current study to a group of native speakers of Spanish. These decisions allowed us to move both conceptually and methodologically away from native-speaker and prescriptive biases. Conceptually, we reconceived of additional-language development of gender marking as the use and evolution in the use of modifier gender, rather than improvement in accuracy. Methodologically, we introduced a new way of analyzing gender marking through a change in the dependent variable and we employed the multivariate analytical tools common in variationist SLA in order to explain the complex and systematic variability in the use of modifier gender. With variationist SLA, we were able to explain the intricacies in participants' variable use of gender marking and how this variability changed over time. In sum, we believe that the new knowledge that emerged from the current study demonstrates that the variationist approach can be beneficially adapted in order to move SLA away, both conceptually and methodologically, from native-speaker and prescriptive biases, and we believe that it is worth considering other existing frameworks to see how they might also be valuable in contributing to this paradigm shift.

A second challenge we encountered in trying to move away from a native or prescriptive norm was whether we should assess whether participants were becoming more proficient with grammatical gender (i.e., language development) and if so, how we should go about this kind of assessment (cf. Birdsong and Gertken, 2013). The majority of research within SLA makes reference to language development, with greater proficiency generally considered to correspond to language use or knowledge that is more in line with native-speaker use or knowledge or with prescriptive descriptions. In our analysis, we were able to make observations about development over time but these observations were not connected to notions of proficiency. Indeed, we saw that the use of feminine modifiers increased after a year spent in a target-language environment and that the influence of noun ending on the use of modifier gender evolved over the course of 21 months. While these insights contribute to a better understanding of the use of feminine (versus masculine) modifiers, they do not allow us – and certainly they were not designed or intended to allow us – to speak in terms of improvement per se. This is in stark contrast to most SLA research in general and to previous research on grammatical gender marking in Spanish in particular, where proficiency assessment has been done by examinations of accuracy or targetlikeness. This type of approach perpetuates the comparison of additional-language speakers to a native-speaker or prescriptive baseline. If we lose that baseline, we are necessarily confronted with the question of what is meant by proficiency and how to assess it. One possible solution, according to Cook (2016) and Ortega (2017), that would allow the field to continue to make observations in terms of proficiency would be to use other additional-language or multilingual speakers as

comparison (or baseline) groups, as long as the comparison does not perpetuate a deficit view of language acquisition.¹² In other words, comparing the multilingual speakers in the current study to another group of multilingual speakers can be a way of assessing proficiency with grammatical gender marking. However, we believe that this too begs the question of what it means to be proficient even for a multilingual comparison group. More specifically, in the absence of a benchmark, how might one determine that another group of multilingual speakers is indeed more proficient than the group under study? If SLA continues to be interested in questions of proficiency and if the field moves away from native-speaker and prescriptive norms, we agree with researchers such as Piller (2002) that this presumably requires new conceptualizations of proficiency (see also Monteiro et al., 2018).

Yet another challenge we faced concerned the role that previous research, which was shaped by a native-speaker norm, should have in the current study. This came up in two ways. First, it has been common practice for research to refer to previous studies in order to identify and motivate the variables that are examined in subsequent investigations. We have followed suit with the current study. Our dependent variable of modifier gender stems from previous (target-oriented) research that has observed that additional-language participants are more accurate in marking grammatical gender with masculine nouns than feminine nouns. Furthermore, all of our independent variables had been investigated previously in work on grammatical gender (cf. Gudmestad et al., 2019). The advantage is that they are justified by past research. However, one might question this decision on at least two grounds. First, why might we believe that the same variables thought to influence targetlike use of gender marking would also be involved in explaining the use of modifier forms? Second, is not the reliance on previous (targetlike-oriented) research for the identification of variables a way of introducing native-speaker bias into a project that precisely set out to avoid such bias? With respect to this first question, we decided to examine these factors in the current study because the linguistic phenomenon under investigation was the same (i.e., grammatical gender marking), even though the object of study had shifted (i.e., modifier gender rather than targetlikeness). In other words, in the absence of previous research on modifier use to guide us, we hypothesized that factors thought to influence targetlike gender marking might also impact the variable use of modifier gender. With respect to the second question, in the current study, we considered that the potential to introduce native-speaker or prescriptive bias differed as a function of the independent variable in question. Namely, we differentiated between, on the one hand, the factors of noun gender and initial proficiency and, on the other, other factors identified in previous research. This was done precisely in order to move away from a native-speaker or prescriptive target.

¹²On the basis of the present results, another possible avenue may be to explore development from the perspective of a move away from use characterized by default forms. In this case, greater proficiency would be defined as evolution from general reliance on default forms (such as masculine modifiers) towards greater use of non-default forms (such as feminine modifiers), as was the case between pre-stay and in-stay in the current study. Care would be needed, though, to make sure that the conceptualization and analysis of default forms is clearly distinct from notions of native-speaker and prescriptive norms.

Although noun gender has been extensively studied in the previous (target-oriented) research on gender marking, prescriptive grammar rules dictate that noun gender is the sole feature that conditions the use of modifier gender and, aside from a small group of nouns whose gender differs by geographical area, native speakers do not appear to exhibit sociolinguistic variation in their use of gender marking (cf. Gudmestad et al., 2019). This suggests that native-speaker use is influenced by one factor alone – noun gender –, just as grammar rules prescribe. Thus, in order to carry out an analysis of gender marking that was not reliant on native-speaker and prescriptive biases, we decided not to include noun gender as a potential explanatory factor in our analysis because the results from this factor can be interpreted to represent a native-speaker or prescriptive norm. A similar motivation led us to exclude initial proficiency from the analysis. Although an initial proficiency score was obtained for each participant at the outset of the project using an elicited imitation task, such a score reflects the participants' ability to imitate prescriptively accurate forms, including gendered forms. For this reason, the measure of (prescriptive) initial proficiency was excluded. Thus, the current analysis consisted only of factors that could be used to characterize language use without an implicit or explicit reference to a native-speaker or prescriptive standard. By relying on previous research to design the current study, one might suggest that our strategy for reform is one that Ortega (2017) would call “modest” and a more “ambitious” strategy would be a bottom-up one in which researchers conduct detailed, qualitative analyses of multilingual language use in order to identify emergent variables.

Additionally, we debated how or whether research that has attempted to move away from a native-speaker bias should make connections with previous work that was impacted by this bias. Again, it is common practice in SLA for researchers to make connections among investigations and to be explicit about how one study builds on existing knowledge. However, moving away from assessments of language that compare additional-language participants to native-speaker and prescriptive norms is a notable conceptual change. With such a change, how do researchers succeed in making connections between differently oriented analyses in order to build new knowledge or is it worthwhile to even make these comparisons? Or, does the field of SLA need to build an entirely parallel body of knowledge? Once again, we speculated that the extensive previous research provided a relevant starting point for the present investigation, in so far as prior studies informed our selection of the dependent variable and independent variables. Only subsequent research can show whether this is a justified position on our part. Importantly, we have not made explicit comparisons between the current study's results and prior investigations because they are not on the same footing, due to the difference in the object of study (i.e., accuracy versus modifier gender). Nevertheless, we feel that this is an important issue that SLA needs to consider if this paradigm shift becomes more integrated into the field.

CONCLUSION

In the present investigation, we heeded the call for SLA research to depart from analyses that compare additional-language participants to native-speaker or prescriptive benchmarks. In so doing, we drew

on strengths of an existing framework in the field – variationist SLA – in order to conduct an analysis of grammatical gender marking in Spanish that was independent from native-speaker and prescriptive targets. We shifted the object of study from accuracy to one that centered on additional-language participants' use of gender-marked modifiers. The results revealed new observations about the acquisition of gender marking in additional-language Spanish. Specifically, the variable use of modifier gender was conditioned by both linguistic (noun ending, syllable number, and modifier type) and extra-linguistic (time and task) factors, and noun ending helped to explain changes along the developmental trajectory. The current study has also demonstrated that existing approaches to SLA research can be adapted in order to help the field move beyond its native-speaker bias. Moreover, we discussed some of the challenges that we encountered when attempting to integrate this conceptual and methodological change into research on a well-studied linguistic phenomenon. In order to encourage researchers to reflect on how to respond to calls to avoid native-speaker and prescriptive biases, we feel that it is important to identify challenges inherent in this paradigm shift, as well as potential solutions to the issues encountered. By publicly reflecting on these challenges, we hope to encourage further dialogue on these issues.

DATA AVAILABILITY STATEMENT

The data presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://langsnap.soton.ac.uk/>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Southampton, United Kingdom. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AG and AE contributed to the conception and design of the study. AG is primarily responsible for the data coding. TM is largely responsible for the statistical analysis. AG wrote the paper. AG and AE contributed to revising the submitted version.

FUNDING

The publication of this paper was funded by Virginia Tech's Open Access Subvention Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.723496/full#supplementary-material>

REFERENCES

- Alarcón, I. V. (2010). Gender Assignment and Agreement in L2 Spanish: The Effects of Morphological Marking, Animacy, and Gender. *Stud. Hispanic Lusophone Linguist.* 3, 267–300. doi:10.1515/shll-2010-1076
- Alarcón, I. V. (2011). Spanish Gender Agreement under Complete and Incomplete Acquisition: Early and Late Bilinguals' Linguistic Behavior within the Noun Phrase. *Bilingualism* 14, 332–350. doi:10.1017/s1366728910000222
- Alarcón, I. (2014). "Grammatical Gender in Second Language Spanish," in *The Handbook of Spanish Second Language Acquisition*. Editor K. L. Geeslin (Malden, MA: Wiley Blackwell), 202–218. doi:10.1002/9781118584347.ch12
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Birdsong, D., and Gertken, L. M. (2013). In Faint Praise of Folly: A Critical Review of Native/Non-Native Speaker Comparisons, With Examples of Native and Bilingual Processing of French Complex Syntax. *Lia* 4, 107–133. doi:10.1075/lia.4.2.01bir
- Bley-Vroman, R. (1983). The Comparative Fallacy in Interlanguage Studies: the Case of Systematicity. *Lang. Learn.* 33, 1–17. doi:10.1111/j.1467-1770.1983.tb00983.x
- Bruhn de Garavito, J. B., and White, L. (2002). "The Second Language Acquisition of Spanish DPs: The Status of Grammatical Features," in *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection*. Editors A. T. Pérez-Leroux and J. M. Licerias (New York: Bilingual Press), 153–178. doi:10.1007/978-94-010-0291-2_6
- Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language* 82, 711–733. doi:10.1353/lan.2006.0186
- Cook, V., and Wei, L. (Editors) (2016). *The Cambridge Handbook of Linguistic Multi-Competence* (Cambridge, UK: Cambridge University Press).
- Cook, V. J. (1992). Evidence for Multicompetence. *Lang. Learn.* 42, 557–591. doi:10.1111/j.1467-1770.1992.tb01044.x
- Cook, V. (1997). Monolingual Bias in Second Language Acquisition Research. *Revista Canaria de Estudios Ingleses* 34, 35–49.
- Cook, V. (2016). "Premises of Multi-Competence," in *The Cambridge Handbook of Linguistic Multi-Competence*. Editors V. Cook and L. Wei (Cambridge, UK: Cambridge University Press), 1–25.
- Dabrowska, D. (2012). Different Speakers, Different Grammars: Individual Differences in Native Language Attainment. *Linguist. Approaches to Bilingual.* 3, 219–253. doi:10.1075/lab.2.3.01dab
- Davies, M. (2016-). *Corpus del español: Two billion words, 21 countries*. Available at: <http://www.corpusdelespanol.org/web-dial> (Accessed July 2017).
- de Prada Pérez, A. (2015). "First Person Singular Subject Expression in Spanish in Contact with Catalan," in *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective*. Editors A. M. Carvalho, R. Orozco, and N. Lapidus Shin (Washington, D. C.: Georgetown University Press), 121–142.
- Dewaele, J.-M., and Véronique, D. (2001). Gender Assignment and Gender Agreement in Advanced French Interlanguage: A Cross-Sectional Study. *Bilingualism* 4, 275–297. doi:10.1017/s136672890100044x
- The Douglas Fir Group (2016). A Transdisciplinary Framework for SLA in a Multilingual World. *Mod. Lang. J.* 100, 19–47. doi:10.1111/modl.12301
- Fernández-García, M. (1999). "Patterns of Gender Agreement in the Speech of Second Language Learners," in *Advances In Hispanic Linguistics: Papers From the 2nd Hispanic Linguistics Symposium*. Editors J. Gutiérrez-Rexach and F. Martínez-Gil (Somerville, MA: Cascadilla Press), 3–15.
- Finnemann, M. D. (1992). Learning Agreement in the Noun Phrase: The Strategies of Three First-Year Spanish Students. *Int. Rev. Appl. Linguistics Teach.* 30, 121–136. doi:10.1515/iral.1992.30.2.121
- Geeslin, K. L., and Long, A. Y. (2014). *Sociolinguistics and Second Language Acquisition: Learning to Use Language in Context*. New York, NY: Routledge. doi:10.4324/9780203117835
- Geeslin, K. L., García-Amaya, L. J., Hasler-Barker, M., Henriksen, N. C., and Killam, J. (2012). "The L2 Acquisition of Variable Perfective Past-Time Reference in Spanish in an Overseas Immersion Setting," in *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Editors K. Geeslin and M. Díaz-Campos (Somerville, MA: Cascadilla Proceedings Project), 197–213.
- Gudmestad, A., Edmonds, A., and Metzger, T. (2019). Using Variationism and Learner Corpus Research to Investigate Grammatical Gender Marking in Additional Language Spanish. *Lang. Learn.* 69 (4), 911–942. doi:10.1111/lang.12363
- Hall, J. K., Cheng, A., and Carlson, M. T. (2006). Reconceptualizing Multicompetence as a Theory of Language Knowledge. *Appl. Linguist.* 27, 220–240. doi:10.1093/applin/aml013
- Hashimoto, B. J., and Egbert, J. (2019). More Than Frequency? Exploring Predictors of Word Difficulty for Second Language Learners. *Lang. Learn.* 69, 839–872. doi:10.1111/lang.12353
- Kanwit, M. (2017). What We Gain by Combining Variationist and Concept-Oriented Approaches: The Case of Acquiring Spanish Future-Time Expression. *Lang. Learn.* 67, 461–498. doi:10.1111/lang.12234
- Kass, R. E., and Raftery, A. E. (1995). Bayes Factors. *J. Am. Stat. Assoc.* 90 (430), 773–795. doi:10.1080/01621459.1995.10476572
- Kinginger, C. (2009). *Language Learning and Study Abroad: A Critical Reading of Research*. London: Palgrave Macmillan. doi:10.1057/9780230240766
- Klein, W. (1998). The Contribution of Second Language Acquisition Research. *Lang. Learn.* 48, 527–549. doi:10.1111/0023-8333.00057
- Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Llanes, Á. (2011). The many Faces of Study Abroad: An Update on the Research on L2 Gains Emerged During a Study Abroad Experience. *Int. J. Multiling.* 8, 189–215. doi:10.1080/14790718.2010.550297
- Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-native Speakers*. Cambridge: Cambridge University Press.
- Mitchell, R., Tracy-Ventura, N., and McManus, K. (2017). *Anglophone Students Abroad: Identity, Social Relationships, and Language Learning*. New York, NY: Routledge. doi:10.4324/9781315194851-2
- Monteiro, K. R., Crossley, S. A., and Kyle, K. (2018). In Search of New Benchmarks: Using L2 Lexical Frequency and Contextual Diversity Indices to Assess Second Language Writing. *Appl. Linguist.* 41, 280–300. doi:10.1093/applin/amy056
- Montrul, S., Foote, R., and Perpiñán, S. (2008). Gender Agreement in Adult Second Language Learners and Spanish Heritage Speakers: The Effects of Age and Context of Acquisition. *Lang. Learn.* 58, 503–553. doi:10.1111/j.1467-9922.2008.00449.x
- Mougeon, R., Nadasdi, T., and Rehner, K. (2010). *The Sociolinguistic Competence of Immersion Students*. Clevedon, UK: Multilingual Matters. doi:10.21832/9781847692405
- Mulder, K., and Hulstijn, J. H. (2011). Linguistic Skills of Adult Native Speakers, as a Function of Age and Level of Education. *Appl. Linguist.* 32, 475–494. doi:10.1093/applin/amr016
- Murahata, G., Murahata, Y., and Cook, V. (2016). "Research Questions and Methodology of Multi-Competence," in *The Cambridge Handbook of Multi-Competence*. Editors V. Cook and L. Wei (Cambridge, UK: Cambridge University Press), 26–49.
- Nance, C., McLeod, W., O'Rourke, B., and Dunmore, S. (2016). Identity, Accent Aim, and Motivation in Second Language Users: New Scottish Gaelic Speakers' Use of Phonetic Variation. *J. Socioling.* 20 (2), 164–191. doi:10.1111/josl.12173
- O'Rourke, B., Pujolar, J., and Ramallo, F. (2015). New Speakers of Minority Languages: The Challenging Opportunity – Foreword. *Int. J. Sociol. Lang.* 231, 1–20. doi:10.1515/ijsl-2014-0029
- Ortega, L. (2013). SLA for the 21st Century: Disciplinary Progress, Transdisciplinary Relevance, and the Bi/multilingual Turn. *Lang. Learn.* 63, 1–24. doi:10.1111/j.1467-9922.2012.00735.x
- Ortega, L. (2014). "Ways Forward for a Bi/multilingual Turn in SLA," in *The Multilingual Turn: Implications for SLA, TESOL, and Bilingual Education*. Editor S. May (New York: Routledge), 32–53.
- Ortega, L. (2016). "Multi-competence in Second Language Acquisition: Inroads into the Mainstream?," in *The Cambridge Handbook of Multi-Competence*. Editors V. Cook and L. Wei (Cambridge, UK: Cambridge University Press), 50–76.
- Ortega, L. (2017). *The Bi/multilingual Turn in SLA: How Far Have We (Not) Come, and Why? 36th Second Language Research Forum*. Columbus, OH: The Ohio State University. October 12–15.
- Picoral, A., and Carvalho, A. M. (2020). The Acquisition of Preposition + Article Contractions in L3 Portuguese Among Different L1-Speaking Learners: A Variationist Approach. *Languages* 5, 45. doi:10.3390/languages5040045
- Piller, I. (2002). Passing for a Native Speaker: Identity and success in Second Language Learning. *J. Socioling.* 6, 179–208. doi:10.1111/1467-9481.00184

- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Regan, V., Howard, M., and Lemée, I. (2009). *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*. Clevedon, UK: Multilingual Matters. doi:10.21832/9781847691583
- Rehner, K. (2002). The Development of Aspects of Linguistic and Discourse Competence by Advanced Second Language Learners of French. *Dissertation Abstr. Int.* 63, 12, 2002. Doctoral dissertation.
- Sabourin, L., Stowe, L. A., and de Haan, G. J. (2006). Transfer Effects in Learning a Second Language Grammatical Gender System. *Second Lang. Res.* 22, 1–29. doi:10.1191/0267658306sr259oa
- Smith, T. J., and McKenna, C. M. (2013). A Comparison of Logistic Regression Pseudo R² Indices. *Mult. Linear Regression Viewpoints* 39, 17–26.
- Tarone, E. (2007). Sociolinguistic Approaches to Second Language Acquisition Research-1997-2007. *Mod. Lang. J.* 91, 837–848. doi:10.1111/j.1540-4781.2007.00672.x
- Teschner, R. V., and Russell, W. M. (1984). The Gender Patterns of Spanish Nouns: An Inverse Dictionary-Based Analysis. *Hispanic Linguist.* 1, 115–132.
- Tracy-Ventura, N., and Huensch, A. (2018). “Chapter 6. The Potential of Publicly Shared Longitudinal Learner Corpora in SLA Research,” in *Critical Reflections on Data in Second Language Acquisition*. Editors A. Gudmestad and A. Edmonds (Amsterdam: John Benjamins), 149–170. doi:10.1075/llt.51.07tra
- White, L., Valenzuela, E., Kozłowska-Macgregor, M., and Leung, Y.-K. I. (2004). Gender and Number Agreement in Nonnative Spanish. *Appl. Psycholing* 25, 105–133. doi:10.1017/s0142716404001067
- Young, R. (1991). *Variation in Interlanguage Morphology*. New York, NY: Peter Lang.
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2021 Gudmestad, Edmonds and Metzger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*