Chapter 5

On the relationship between epistemology and methodology: A reanalysis of grammatical gender in additional-language Spanish

Aarnes Gudmestad

Virginia Polytechnic Institute and State University

In the current study I explore the relationship between epistemology and methodology through a reanalysis of production data on grammatical gender in additional-language Spanish that were analysed in Gudmestad et al. (2019). This reanalysis consists of a shift in the epistemology from the one adopted by Gudmestad et al., where gender marking, which occurs between nouns and both determiners and adjectives, is a unified linguistic phenomenon. In contrast, the assumption in the present investigation is that the acquisition of gender marking entails learning gender assignment and gender agreement, two different learning processes that are observable in language behaviour with determiners and adjectives, respectively. In order to reflect critically on the relationship between epistemology and methodology and specifically on its influence on the interpretation of learner data, I conduct a multi-step analysis that is guided by the differentiation between gender assignment, which can be observed on determiners, and gender agreement, which can be observed on adjectives. I also discuss how the interpretation of the findings can be impacted by the epistemology that guides the current study.

Keywords: Epistemology, methodology, grammatical gender, Spanish, SLA

1 Introduction

As attention has been increasingly paid to methodological reform in applied linguistics (Byrnes 2013; Phakiti et al. 2018), there have been calls for change on

many fronts, such as open science (Marsden & Plonsky 2018), the reporting of quantitative results (Larsen-Hall & Plonsky 2015), and the need for replication (Porte & McManus 2018). And consequently, the methodological norms in the field are changing (e.g., Marsden et al. 2018). Improvement in quantitative methods is one of the specific areas that has received the most consideration (e.g., Plonsky 2015) and its import is clear: The veracity of the findings that emerge from statistical tests is contingent on the appropriate use of those tests. Another, perhaps more global, issue that is equally important but seems to have garnered less explicit attention is the connection between methodology and epistemology (Ortega 2005). This relationship pertains to the ways in which methodological practices are linked to epistemology or "what counts as knowledge … and how this relates to truth, belief, and justification" (Young 2018: 40). In the current study, I aim to contribute to discussions about the connection between methodology and epistemology through a focus on grammatical gender. Specifically, I explore this relationship through a reanalysis of production data on grammatical gender in additional-language[1] Spanish, originally reported on in Gudmestad et al. (2019). This reanalysis follows from a change in the epistemology. Whereas Gudmestad et al. treated gender marking as a single phenomenon, in the current study, gender assignment and gender agreement are considered to be different learning processes that are observable in language behaviour with determiners and adjectives (see Section 2.2 on *Grammatical gender in additional languages* for details). I show how this change in epistemology can orient not only the data analysis but also the interpretation of the findings, thus fundamentally changing what counts as relevant knowledge in the field of second language acquisition (SLA).

2 Background

In this section, I first briefly describe the relationship between epistemology and methodology. I then discuss one specific assumption that exists in research on grammatical gender and that guides the current study. Lastly, I introduce grammatical gender in Spanish and I briefly describe Gudmestad et al. (2019), because I reanalyse the dataset from this previous study in the present investigation.

---

[1]"Additional-language" is an inclusive term that refers to any language learned after the first language (cf. The Douglas Fir Group 2016).

2.1 Methodology and epistemology

The methodological decisions that scholars make are linked to many facets of the research process. Ortega (2005), for example, highlights the relationship among methodology, epistemology, and ethics:

> Research communities make decisions about best ways to approach the task of producing evidence (methodology) on the basis of agreed-upon notions of the nature of what can, or cannot, be captured and explicated as evidence (epistemology) and by drawing on agreed-upon valuations of what is, or is not, worth understanding and transforming (axiology). (p. 317)

In brief, this connection among different components of scholarship means that when reflecting on methodological practices of interpreting data, it is also valuable to consider other aspects of the research process. While each of the three issues highlighted by Ortega is important, I focus the present investigation on the connection between epistemology and methodology.

Creswell & Creswell (2018: 5) note that epistemologies or ontologies are also called worldviews or paradigms by some scholars and that, regardless of the term, this dimension of research refers generally to the assumptions that researchers have about their discipline or the world that impact methodological decisions. An example of this link between epistemology and methodology is seen in recent calls for multivariate, quantitative analyses in learner corpus research (Gries 2005) and SLA (Plonsky & Oswald 2017). These researchers have argued for the need to move from univariate to multivariate analyses in quantitative scholarship because the latter better align with the complexities of the acquisitional process. In other words, since the epistemology is that there are numerous factors at play in the development of an additional language, then the methodological practices (in this case, the statistical analyses we conduct to examine language behaviour and acquisition) should align with this reality. To illustrate this relationship between epistemology and methodology, I now turn to grammatical gender in additional languages.

2.2 Grammatical gender in additional languages

One assumption that is made in some investigations on grammatical gender in SLA is that learners face two primary learnability issues, which are visible in the marking of gender on different sets of modifiers. The acquisitional challenges are learning the gender of the noun (gender assignment, a lexical property) and

matching the gender of a modifier with the gender of the noun (gender agreement or gender concord, a morphosyntactic property). What is more "learners ... need to acquire gender assignment for individual nouns in their internal grammars before being able to produce correct gender agreement in sentences" (Alarcón 2010: 268). Furthermore, some researchers (e.g., Ayoun 2007; Alarcón 2010; Kupisch et al. 2013) consider that gender marking on determiners reflects gender assignment (e.g., *la película* 'the$_F$ movie$_F$'), whereas gender marking on adjectives constitutes gender agreement (e.g., *duraznos amarillos* 'yellow$_M$ peaches$_M$'). With this distinction, data showing targetlike gender marking on determiners are interpreted to indicate that learners have acquired the appropriate gender of the noun, and data exhibiting targetlike gender marking between nouns and adjectives are understood to reflect learners' ability to match the gender of modifiers with a noun's gender. Investigations that subscribe to this epistemology have found lower rates of targetlike gender marking on adjectives compared to determiners, and this observation has been interpreted as an indication that the morphosyntactic marking of grammatical gender (i.e., gender agreement) is a more challenging learnability issue for learners than assigning a noun its appropriate gender (i.e., a lexical property). Thus, concerning the connection between epistemology and methodology, the assumption among Ayoun, Alarcón, and Kupisch et al. is that gender assignment and gender agreement are different learning processes that are observable in linguistic behaviour, as seen through gender marking on determiners and adjectives, respectively. In the present investigation, I adopt this epistemology, which I refer to as the assignment-agreement assumption. Under this epistemology, researchers can then make the methodological decision to conduct analyses that enable them to distinguish between gender marking on determiners and gender marking on adjectives. When differences are found between the two modifier types, they can be interpreted as evidence in support of this epistemology.

It is important to recognize, however, that the assumption that links gender assignment with determiners and gender agreement with adjectives is not held among all researchers who have investigated grammatical gender (see also Bruhn de Garavito & White 2002; Montrul et al. 2008: 510). As Gudmundson (2013) observes:

> this difference is considered to be a theoretical one, difficult to apply in practice. The difference between assignment errors and agreement errors would be applicable to only a very small number of cases, produced several times by the same learner. This is very seldom the case, as agreement tokens frequently occur only once, and sometimes a correct form co-occurs with

> an incorrect form. It is thus impossible to judge whether an error is due to assignment or to agreement without running the risk of drawing incorrect conclusions. (p. 242)

In other words, linking the assignment-agreement assumption to analyses of targetlike behaviour according to modifier type is not without criticism, and Gudmundson highlights a methodological challenge to this epistemology: Researchers need to be able to observe a given learner's gender marking on multiple occurrences of the same noun (rather than just a single occurrence of a noun). For instance, if a participant uses the noun *libro* 'book' with a modifier only one time and the noun *mesa* 'table' with a modifier three times, then, according to Gudmundson, researchers can make observations about gender marking on *mesa* but not *libro.* The goal of the current study is not to take a position on whether or not the assignment-agreement assumption is valid but rather to reflect critically on the impact that it can have on methodological practices and the interpretation of data.

2.3 Grammatical gender in Spanish

In Spanish every noun has masculine or feminine gender and modifiers (i.e., determiners and adjectives) agree in gender with the noun they modify, as illustrated in (1). Gender is assigned according to biological sex for some nouns (*mujer* 'woman$_F$', *hombre* 'man$_M$'). For most nouns, however, the gender is assigned arbitrarily, such as those in (1). The canonical morpheme for nouns and modifiers is *o* for masculine and *a* for feminine, though there are exceptions (e.g., *mapa* 'map$_M$'). Furthermore, not all nouns and modifiers end in these vowels. Regarding nouns, there are other inflectional endings that are predictive of one gender (e.g., *tad* for feminine nouns as in *lealtad* 'loyalty' and *e* for masculine nouns, e.g., *bate* 'bat'), as well as endings that are not linked to a particular gender (e.g., *s*; *lunes* 'Monday$_M$' versus *oasis* 'oasis$_F$'; Teschner & Russell 1984). Concerning modifiers, not all determiners and adjectives are overtly marked for gender either (e.g., *tu* 'your' and *difícil* 'difficult').

(1) a. El   vestido roj-o
         the.M dress   red-M
         'The$_M$ red$_M$ dress$_M$.'

     b. La   bicicleta car-a
         the.F bike   expensive-F
         'The$_F$ expensive$_F$ bike$_F$.'

Research on grammatical gender in additional-language Spanish spans various theoretical and analytical approaches (e.g., Universal Grammar, variationist SLA), has examined language processing and production (cf. Alarcón 2014), and includes investigations that subscribe to the aforementioned assignment-agreement assumption (e.g., Alarcón 2010; Kupisch et al. 2013) and others that do not (e.g., Montrul et al. 2008; Grüter et al. 2012; Gudmestad et al. 2019). I focus here on Gudmestad et al., which serves as a starting point for the reanalysis in the current study. In Gudmestad et al., we examined gender marking in language production using the longitudinal corpus LANGSNAP (http://langsnap.soton.ac.uk, e.g., Mitchell et al. 2017). Our epistemology was that in language production researchers can make observations about one acquisitional challenge pertaining to grammatical gender – the marking of gender on modifiers. Thus, we made no distinction between gender assignment and gender agreement and analysed each instance of the use of a noun with a modifier (determiner or adjective) that was overtly marked for gender. We adopted a variationist approach (Geeslin & Long 2014), which means that we sought to account for the variability in learners' marking of grammatical gender over time by explaining the linguistic and extra-linguistic factors that conditioned the participants' use of targetlike gender marking (see the *Method* section below for more information on the data, participants, variables, etc.). In general, we found that numerous factors worked together to condition learners' use of targetlike gender marking and that the factor of noun ending helped to explain changes in use along the developmental trajectory. It is worth pointing out that modifier type (determiners versus adjectives) was one of the factors we investigated. And, while we found that learners were more likely to be targetlike in their gender marking with determiners compared to adjectives, we did not interpret these findings in relation to the assumption that determiners reflect a lexical property and adjectives a morphosyntactic one. We interpreted the findings, instead, as evidence of the complex nature of variability in language use and development, such that modifier type was just one of several linguistic features that impacts how learners develop the ability to mark gender on modifiers in a targetlike way. In the present study, I reanalyse the dataset from Gudmestad et al. through the lens of the assignment-agreement assumption.

3 The current study

In order to consider how the assignment-agreement assumption may influence methodological decisions and the interpretation of data pertaining to the additional-language development of grammatical gender marking, I reanalyse the data from Gudmestad et al. (2019). The current study consists of a three-step data analysis in which I examine determiners and adjectives separately. I then interpret the findings in light of the assignment-agreement assumption and reflect on how new knowledge can emerge from this epistemology. In general, this type of reanalysis, in which assumptions are modified, has the potential to shed light on the link between epistemology and methodology highlighted by Ortega (2005). More specifically, I aim to concretely demonstrate how an epistemological shift leads to a particular methodological decision that, in turn, leads the researcher down a new interpretive path.

3.1 Method

3.1.1 Data

I examined data from the LANGSNAP corpus. The corpus consists of production data collected over 21 months, which included an academic year abroad, from additional-language learners of Spanish. The data were collected six different times and at each point the participants completed three tasks: a written argumentative essay, an oral interview, and an oral narration.[2] For the essay, the participants were presented with a topic and asked to write a 200-word composition. The semi-guided interview consisted of opinion questions and questions about the participants' lives; it lasted about 20 minutes. The oral narration was a picture-based task. The participants looked over a set of images and then told the story in their own words. In the present investigation, I report on the data from all tasks that were collected at three of the data-collection periods (cf. Gudmestad et al. 2019). The first data-collection period, called pre-stay in the current study, was collected before the learners went abroad. The second data-collection point that I analysed was the third in-stay period in the LANGSNAP corpus (henceforth, in-stay); this data collection took place a year after the pre-stay and at the end of the academic year abroad. The final point was gathered 21 months after the pre-stay and was the second post-stay data collection in the LANGSNAP corpus (hereafter, post-stay).

---

[2] I analyse the data from the three tasks together.

3.1.2 Participants

I analysed data from 21 of the 27 learners of Spanish in the corpus.[3] They were all pursuing an undergraduate degree in Spanish at a British university and had been studying Spanish for an average of 5.4 years (SD = 3.4, range: 2–14 years). They ranged in age from 20 to 25 years (M = 20.8 years, SD = 1.6). Fifteen were women and six were men. Their first languages were Polish (n = 1), English (n = 19), and both English and Polish (n = 1). At the pre-stay, the participants completed a global proficiency measure – an elicited-imitation task. The group scored an average of 86.1 out of 120 points (SD = 12.7; range: 50–108). During the academic year abroad, they were teaching assistants (n = 10), exchange students (n = 9), and workplace interns (n = 2). Five participants were in Mexico and 16 were in Spain.

3.1.3 Coding and analysis

The coding started by identifying each occurrence of a determiner or an adjective that modified a referent (K = 16,357); only those modifiers that met two criteria were then analysed (k = 11,832). The first criterion was that the modifiers needed to exhibit overt gender marking. For example, an adjective like *bonito/a* 'pretty' was included in the analysis because it has an inflectional gender morpheme, but adjectives like *interesante* 'interesting' were excluded because the form is the same, regardless of whether it modifies a feminine or masculine noun. Second, the current dataset consists only of nouns; pronouns that were modified by adjectives were not analysed (e.g., *ella está contenta* 'she is happy'). Following the assumption that gender marking on determiners and adjectives reflect different learning processes, I then separated the data by modifier type, determiners (k = 9,107) and adjectives (k = 2,725), in order to examine each modifier type separately. An example of the data is available in (2).

(2)   Tengo un-a   amig-a español
      have.1s INDEF-F friend-F spanish.M

      'I have a_F friend_F Spanish_M.' (Participant 165, post-stay, interview)

I analysed gender agreement (as seen on determiners) and gender assignment (as seen on adjectives) in three phases. The first two phases served to examine

---

[3]All of the data were coded by hand. Due to how labour-intensive this coding was, Gudmestad et al. (2019), and consequently the current study, analysed data from a subset of the participants and three of the six total data-collection points. The learners analysed in Gudmestad et al. and the current study were the first 21 participants in the corpus.

claims made in previous research about the differences between the two learning challenges mentioned in the literature review. The third step sought to further knowledge of the potential differences between these two processes by identifying factors that explain patterns in the data. The dependent variable for each step of the analysis was the targetlikeness of the gender marking: targetlike (the gender of the modifier matched the gender of the noun) or nontargetlike (the gender of the modifier differed from that of the noun).

For the first phase of the analysis, I explored the assertion that gender assignment is acquired before gender agreement and that, under the assignment-agreement assumption, this claim leads to the expectation that targetlike use with determiners is higher than that with adjectives (cf. Alarcón 2010). In order to address this issue, I identified the mean rate of targetlike use for adjectives and determiners at each data-collection point. With a two-way ANOVA, I also assessed whether the rates of targetlike use at each data-collection point were similar or different between determiners and adjectives.

Next, some previous research that subscribes to the assignment-agreement assumption appears to consider the acquisition of gender assignment to be binary: Either learners have acquired a noun's gender or they have not (e.g., Alarcón 2010). In order to address this claim in the second phase of the analysis, I sought to determine whether gender assignment (a lexical property) and gender agreement (a morphosyntactic property) resulted in categorical behaviour of gender marking. I examined targetlike assignment and agreement with individual nouns that participants used more than once at pre-stay; this assessment shows how many unique nouns that participants produced more than once exhibited categorical targetlike use.[4] It might be expected to find that, with determiners, learners exhibit either categorical targetlike or categorical nontargetlike use on individual nouns (i.e., rather than a mix of the two with a given noun, when a participant uses the noun more than once). However, the hypothesis for gender agreement may be different. Under the assumption that gender marking on adjectives reflects a morphosyntactic process, it may be reasonable to find that a noun, when used multiple times by a participant, shows targetlike agreement in some instances and nontargetlike agreement in others. This variability may be expected because the morphosyntactic features of an agreement relationship can differ each time a noun is used. For example, in one instance the adjective may be attributive, occurring in the noun phrase (*Tengo un gato blanco.* 'I have

---

[4]While this analysis may be valuable for each data-collection point, I focus on the pre-stay data in order to offer an example of what this type of analysis may contribute to the understanding of grammatical gender marking.

a white cat.') and in another case the adjective may be predicative, connected to the noun by means of a verbal phrase (*Mi gato es blanco.* 'My cat is white.').

In light of the findings from the second part of the analysis (see the *Results and Discussion* section), which provided preliminary evidence of variability with both gender assignment and agreement in language production, I sought to explain this variability in targetlike gender marking through multivariate analyses. Thus, for the third phase in the analysis, I adopted a variationist approach (Geeslin & Long 2014) in order to investigate variable gender assignment and agreement in language production. This approach, which was also employed in Gudmestad et al. (2019), models variable language behaviour by examining a range of factors (i.e., independent variables, fixed effects) simultaneously. Through two separate multivariate analyses, I identified which factors significantly impacted gender assignment and those that predicted gender agreement. If these two processes are indeed distinct, it may be expected to find that there are some conditioning factors that differ between the two learning challenges.

In order to conduct multivariate analyses, I examined nine fixed effects for both determiners and adjectives, all factors that were motivated by previous research (see Gudmestad et al. 2019 for justification of these factors): noun gender, noun ending, noun class, noun number, task, time, noun frequency (individual), noun log-frequency (language), and initial proficiency. Noun gender distinguishes between feminine and masculine nouns. For noun ending there were four categories: canonical, deceptive, predictive, and other endings. Canonical endings refer to masculine nouns that end in *o* and feminine nouns that end in *a*, and deceptive endings are the opposite: Masculine nouns ending in *a* and feminine nouns ending in *o*. Predictive endings are those that are strongly linked to one gender (e.g., *dad* is linked with feminine gender, Teschner & Russell 1984) and other endings are those that are not strongly connected with one gender (e.g., *s* Teschner & Russell 1984). Noun class differentiates between nouns with biological and arbitrary gender. Noun number explores possible differences between singular and plural nouns. Task pertains to the oral interview, the oral narration, and the written argumentative essay. Time distinguishes between the pre-stay, in-stay, and post-stay data-collection periods. The four remaining factors that were investigated for both adjectives and determiners were continuous factors. Noun frequency (individual) refers to the number of times that each learner produced a noun with a gender-marked modifier in a specific task and data-collection point. The noun log-frequency (language) refers to how often a noun occurs per one million words in the *Corpus del español* (Davies 2016--). Initial proficiency considers the score that each participant received on the elicited-

imitation task before going abroad (see Section 3.1.2). Furthermore, I coded for one factor that was unique to determiners and one that was unique to adjectives – two factors that were not examined in Gudmestad et al. since all determiners and adjectives were analysed together. Determiner type was investigated for determiners only. This factor was motivated by Bruhn de Garavito & White (2002) who found higher targetlike gender marking with definite articles compared to indefinite articles. In the current study, I examined a wider array of categories: definite article (*la* 'the$_{\text{FEM}}$'), indefinite article (e.g., *un* 'a$_{\text{MASC}}$'), demonstrative-this (e.g., *estos* 'these$_{\text{MASC}}$'), demonstrative-that (e.g., *esa* 'that$_{\text{FEM}}$'), indeterminate (e.g., *alguna* 'some/any$_{\text{FEM}}$'), and possessive (*nuestra* 'our$_{\text{FEM}}$'). The factor investigated for adjectives only was adjective position. The three categories were pre (the adjective came before the noun in the same noun phrase), post (the adjective was after the noun in the same noun phrase), and other (the adjective was in a different phrase than the noun). Prior studies have offered conflicting evidence as to whether adjective position plays a role in additional-language development (e.g., Bartning 2000; Dewaele & Véronique 2001). Finally, participant was examined as a random effect, in order to account for variability among the learners. In terms of the analysis, I fit two mixed-effects regression models – one for determiners and one for adjectives – using the statistical software R (**RCoreTeam2017**). Factors not found to be significant were removed from the statistical models. After the significant fixed effects were identified, I explored interactions between time and each of the remaining fixed effects in order to make observations about change over time. I also tested for correlations between independent variables to ensure that there were no strong correlations among the factors included in each regression model. Finally, I reported the McFadden's R^2 (Smith & McKenna 2013) for each model, a metric that indicates whether each model fits the data well. With this third phase in the analysis, I compared the determiner model with the adjective model in order to make observations about similarities and differences between gender assignment and gender agreement.

3.2 Results and discussion

In this section, I present the findings of each of the three steps of the data analysis. I also discuss the findings in relation to the assignment-agreement assumption. As a reminder, my objective is not to take a stance on the validity of the assignment-agreement assumption. Instead I am to reflect on the role it plays in methodological decisions and data interpretation.

3.2.1 Rates of targetlike use

Table 1 provides the average rate of targetlike use for the learners according to modifier type (adjective or determiner) and time (pre-stay, in-stay, and post-stay). I conducted a two-way ANOVA to examine the effect of modifier type and time on targetlike use. The interaction between modifier type and time was not significant, $F(2, 120) = 2.152$, $p = 0.121$. However, the main effect for modifier type was significant ($F(1, 120) = 49.44$, $p < 0.001$), indicating that the participant group was more targetlike with determiners ($M = 96.69$, SD = 2.52) than adjectives ($M = 90.43$, SD = 7.52). This finding can be interpreted as evidence that is consistent with the epistemology that gender marking on determiners reflects gender assignment and gender marking on adjectives reflects gender agreement. Therefore, since learners need to acquire a noun's gender (a lexical property) before being able to use targetlike gender agreement (a morphosyntactic property), the finding that targetlike gender marking was higher with determiners than adjectives was expected based on previous research (e.g., Alarcón 2010). The main effect for time was also significant ($F(2, 120) = 15.705$, $p < 0.001$). The learners were more targetlike at in-stay ($M = 95.312$, SD = 4.045) and post-stay (M 95.340, SD = 4.260) compared to pre-stay ($M = 90.029$, SD = 8.456),[5] but there was no significant difference between in-stay and post-stay ($p = 1.000$). These findings suggest improvement in targetlikeness of gender marking as a whole during the academic year abroad that was maintained after returning to the United Kingdom.

Table 1: Rates of targetlike use (in percentages)

	Determiners			Adjectives		
	Pre-stay	In-stay	Post-stay	Pre-stay	In-stay	Post-stay
M	94.46	97.95	97.67	85.60	92.67	93.01
SD	2.69	1.40	1.60	9.91	4.12	4.82
Median	93.90	98.10	97.80	84.00	92.10	93.90
Range	89.0–98.7	94.8–100	93.4–100	64.3–100	85.0–100	79.3–98.1

---

[5]The *p* value*s* for the pre-stay and in-stay comparison and the pre-stay and post-stay comparisons are both *p* < 0.01.

3.2.2 Individual nouns at pre-stay

Next, I examined targetlikeness of gender marking for the individual nouns that each participant produced more than once with a modifier overtly marked for gender at pre-stay. This analysis focuses on nouns that learners used more than once at pre-stay, rather than those that were used just one time, given the argument by Gudmundson (2013: 242) that researchers need to examine multiple occurrences of a particular noun, in order to make observations about gender assignment and gender agreement.

The results for gender assignment (as seen on determiners at pre-stay) are presented in Table 2. For example, participant 150 used a total of 103 different nouns with a determiner overtly marked for gender at pre-stay, of which 50 were used more than once. With 48 of the nouns that she used more than once, the gender of the determiner was targetlike 100 percent of the time. In contrast, one noun (*rana* 'frog') exhibited variable targetlike use and one noun (*programa* 'program') was consistently used with a determiner that did not match the gender of the noun (i.e., categorical nontargetlike use). This participant also used 53 nouns just one time at pre-stay.

Continuing with Table 2, the results indicate that each participant assigned the targetlike gender to most nouns that they used at least twice. Only three participants used one unique noun multiple times and always used a determiner that differed in gender from the noun it modified (participants 150, 156, and 168). Moreover, each participant used between one and 10 nouns in which some instances exhibited targetlike use of gender on the determiner and others did not (i.e., use of both masculine and feminine determiners with the same noun). For example, participants 158 and 164 used both masculine and feminine determiners with four unique nouns. The nouns were *día* 'day', *objeto* 'object', *problema* 'problem', and *telenovela* 'soap opera' for participant 158 and *apartamento* 'apartment', *casa* 'house', *idea* 'idea', and *mujer* 'woman' for participant 164. Under the assumption that gender marking on determiners reflects gender assignment (i.e., the lexical property of noun gender), this observation may be surprising as it suggests evidence of variable knowledge or of varying degrees of the strength of the lexical representations between nouns and their gender (Halberstadt et al. 2018). Thus, it may be that a noun that exhibits variability in targetlike use on determiners has a weaker gender representation than a noun whose gender assignment is categorical. These findings contrast with previous research that has considered gender assignment to be a categorical property (cf. Alarcón 2010).

Turning to gender agreement at pre-stay, I assessed targetlike gender marking on adjectives for the unique nouns produced more than once by individual par-

Table 2: Unique nouns and targetlikeness with determiners at pre-stay.
Note: Percentages may not add up to 100 due to rounding.

Part.	Nouns used more than once						Nouns used once		Total
	All targetlike		Variable		All nontargetlike				
	#	%	#	%	#	%	#	%	
150	48	46.60	1	0.97	1	0.97	53	51.46	103
151	29	37.18	5	6.41	0	0	44	56.41	78
152	29	33.33	1	1.15	0	0	57	65.52	87
155	32	39.02	1	1.22	0	0	49	59.76	82
156	19	24.36	2	2.56	1	1.28	56	71.79	78
157	16	25.81	6	9.68	0	0	40	64.52	62
158	29	31.87	4	4.40	0	0	58	63.74	91
160	28	35.00	1	1.25	0	0	51	63.75	80
161	20	33.90	2	3.39	0	0	37	62.71	59
162	25	29.76	2	2.38	0	0	57	67.86	84
163	18	22.78	3	3.80	0	0	58	73.42	79
164	25	37.31	4	5.97	0	0	38	56.72	67
165	22	33.3	3	4.55	0	0	41	62.12	66
166	41	33.6	3	2.46	0	0	78	63.93	122
167	26	28.89	6	6.67	0	0	58	64.44	90
168	20	32.25	2	3.23	1	1.61	39	62.90	62
169	20	25.32	10	12.66	0	0	49	62.03	79
170	23	31.08	1	1.35	0	0	50	67.57	74
171	22	25.29	1	1.15	0	0	64	73.56	87
172	34	33.01	3	2.91	0	0	66	64.08	103
173	24	32.43	3	4.05	0	0	47	63.51	74

ticipants. These results are available in Table 3, which is organized like Table 2. Similar to gender assignment, learners exhibited targetlike gender agreement with most nouns. For example, among the 19 nouns that participant 150 used at least twice with an adjective, she exhibited targetlike gender agreement with 16 of them. Additionally, instances where learners used the same noun with an adjective multiple times but produced nontargetlike gender agreement categorically were uncommon. Participants 156, 157, 160, 166, and 167 each used one noun multiple times and were nontargetlike in their gender agreement every time they used that noun with an adjective. Moreover, there is variability in the marking

of gender on adjectives with some individual nouns. For instance, participants 150 and 163 exhibited variable gender marking on adjectives with three unique nouns. Participant 150 was variable with *chica* 'girl', *hombre* 'man', and *mujer* 'woman' and participant 163 was variable with *idea* 'idea', *identidad* 'identity', and *persona* 'person'.

Under the assignment-agreement assumption, one might expect to see a higher proportion of nouns that exhibit variable targetlike use with adjectives than with determiners, given the fact that morphosyntactic properties can differ based on the linguistic context. Focusing exclusively on nouns that the participants used more than once and that exhibited variability, the average number of nouns that were connected to variable behaviour was higher for adjectives than determiners: 17.81 percent (26/146 nouns) of the group's nouns that were used more than once by individual participants exhibited variable gender agreement, whereas 10.37 percent of their nouns (64/617 nouns) exhibited variable gender assignment.

3.2.3 Mixed-effects regression models

Findings from the second step of the analysis appeared to show variation in gender assignment and agreement. Specifically, the analysis in Section 3.2.2 pointed to the possibility that knowledge of the lexical property of gender assignment is not always categorical. It also indicated that targetlike gender agreement is not always categorical either. In light of these observations, it seems reasonable to look to research approaches in SLA that have implemented methodological tools for investigating variation in order to understand the factors that impact variability in gender assignment and agreement and to make comparisons between the two learning processes. In this vein, I adopt a variationist approach in order to investigate a range of factors that may condition variable gender marking on determiners and adjectives separately (see Section 3.1.3 for a general description of the type of multivariate analysis that is common in variationist SLA and for details on the factors I investigate).

I present the findings for two mixed-effects regression models (Tables 4–7). For the dependent variable and the nominal independent variables, both models compare a reference-point category of each variable to the other category (or categories) of the same variable. The reference point for the dependent variable is targetlike use and the reference points for all significant fixed effects are provided in the Tables 4 and 6 in brackets. The continuous fixed effects do not have reference points. The estimate listed with each category in the tables indicates whether there is a decrease (indicated by a negative estimate) or an increase (denoted by a positive estimate) in the log odds of targetlike use. The p value

Table 3: Unique nouns and targetlikeness with adjectives at pre-stay.
Note: Percentages may not add up to 100 due to rounding.

Part.	Nouns used more than once						Nouns used once		Total
	All targetlike		Variable		All nontargetlike				
	#	%	#	%	#	%	#	%	
150	16	40.00	3	7.50	0	0	21	52.50	40
151	6	28.57	1	4.76	0	0	14	66.67	21
152	5	21.74	0	0	0	0	18	78.26	23
155	4	16.67	1	4.17	0	0	19	79.17	24
156	3	13.04	1	4.35	1	4.35	18	78.26	23
157	2	12.5	1	6.25	1	6.25	12	75.00	16
158	8	26.67	0	0	0	0	22	73.33	30
160	1	4.76	2	9.52	1	4.76	17	80.95	21
161	4	19.05	2	9.52	0	0	15	71.43	21
162	8	25.00	0	0	0	0	24	75.00	32
163	8	25.00	3	9.38	0	0	21	65.63	32
164	0	0	1	4.17	0	0	23	95.83	24
165	9	24.32	1	2.70	0	0	27	72.97	37
166	12	34.29	0	0	1	2.86	22	62.86	35
167	3	10.00	2	6.67	1	3.33	24	80.00	30
168	3	18.75	2	12.5	0	0	11	68.75	16
169	7	26.92	2	7.69	0	0	17	65.38	26
170	1	5.00	1	5.00	0	0	18	90.00	20
171	2	7.41	1	3.70	0	0	24	88.89	27
172	5	16.13	1	3.23	0	0	25	80.65	31
173	8	36.36	1	4.55	0	0	13	59.09	22

(alpha level of $p < 0.05$) reveals whether the estimate is significant. When nominal independent factors have more than two categories (as is the case with noun ending, task, time, determiner type, and adjective position), it is also possible to assess whether there are significant differences between non-reference point categories (e.g., in-stay versus post-stay for time). This can be done by examining the confidence intervals of the non-reference point categories. Overlap between the confidence intervals of categories indicates that the log odds of targetlike use is similar. When the confidence intervals of two categories do not overlap, the log odds of targetlike behaviour can be considered to be different.

Beginning with the mixed-effects regression for determiners, it was found that targetlike gender assignment in this dataset was influenced by noun ending, task, noun gender, noun frequency (individual), initial proficiency, determiner type, and time (Table 4). Noun log-frequency (language), noun class, and noun number were not significant, and I found no significant interactions between time and the other significant fixed effects. Furthermore, none of the fixed effects were

Table 4: Results for the fixed effects in the regression model for determiners. *Note:* The reference point for the dependent variable is targetlike use. The reference points for the independent, nominal variables are in brackets in the table.

Effect	Estimate	SE	z	p	Lower	Upper
Intercept	3.54	0.73	4.84	1.30×10^{-6}	0.18	0.56
Noun ending [canonical]						
deceptive	−2.41	0.21	−11.48	$<2.00 \times 10^{-16}$	−2.82	−1.99
other	−0.69	0.15	−4.45	8.80×10^{-6}	−0.99	−0.38
predictive	−0.25	0.18	−1.35	0.18	−0.60	0.12
Task [written]						
narrative	−1.15	0.24	−4.76	1.95×10^{-6}	−1.63	−0.68
interview	−0.99	0.21	−4.69	2.77×10^{-6}	−1.42	−0.59
Noun gender [masculine]						
feminine	−1.33	0.14	−9.25	$<2.00 \times 10^{-16}$	−1.62	−1.05
Noun freq.[a]	0.13	0.03	3.84	0.0001	0.07	0.20
Initial prof.	1.69	0.80	2.11	0.03	0.04	3.38
Determiner type [definite]						
indefinite	−0.77	0.14	−5.62	1.94×10^{-8}	−1.03	−0.50
indeterminate	−1.60	0.31	−5.20	2.01×10^{-7}	−2.17	−0.96
possessive	−1.64	0.63	−2.60	0.01	−2.73	−0.17
demonstr. - this	−1.11	0.30	−3.69	0.0002	−1.67	−0.48
demonstr. - that	−1.57	0.46	−3.45	0.0006	−2.40	−0.57
Time [pre-stay]						
in-stay	1.01	0.14	7.10	1.22×10^{-12}	0.73	1.30
post-stay	0.85	0.17	4.95	7.39×10^{-7}	0.52	1.20

[a](individual)

strongly correlated. The results for the random effect for participant are available in Table 5. The McFadden's R^2 indicated a moderate fit for this model ($R^2_{McFadden}$ = 0.1339).

Table 5: Results for the random effect in the determiner regression model

Participant	Intercept	Participant	Intercept	Participant	Intercept
150	0.29	160	0.18	167	−0.59
151	0.25	161	0.23	168	0.09
152	−0.14	162	0.50	169	−0.14
155	0.03	163	−0.09	170	0.18
156	−0.09	164	−0.01	171	−0.13
157	−0.13	165	−0.42	172	−0.33
158	0.32	166	−0.08	173	−0.17

For noun ending, the log odds of targetlike gender assignment were significantly lower with deceptive and other endings compared to canonical noun endings. Predictive and canonical endings were not statistically different. In the case of the non-reference point categories of noun ending, there was overlap between the other and predictive endings, which revealed that the log odds of targetlike gender assignment were similar between the two. However, the confidence intervals for deceptively marked nouns did not overlap with other and predictive endings and the values for the confidence intervals of deceptively marked nouns were lower than those for the other categories. This finding indicates that the log odds of targetlike use with deceptively marked nouns were lower than those of predictive and other endings. For gender assignment, these results suggest that noun ending played a role in whether learners assigned the targetlike gender to a noun. Specifically, deceptively marked nouns appeared to present learners with the greatest challenge.

For task, the log odds of targetlike gender assignment were lower with the oral narration and the oral interview compared to the written essay. The overlap in the confidence intervals for the two oral tasks also indicated that targetlike use was similar between the two. Thus, the findings demonstrated a difference between oral and written production. For gender assignment, these results were consistent with claims made by researchers who investigated explicit and implicit knowledge (e.g., Ellis 2006). Specifically, written tasks may enable learners to tap into their explicit knowledge more than they do in oral production so it

may be that this participant group has greater explicit knowledge, compared to implicit knowledge, of gender assignment.

The results for noun gender showed that the log odds of targetlike gender assignment were lower for feminine nouns compared to masculine nouns. Since previous research has demonstrated that the default gender for learners is masculine (e.g., López Prego & Gabriele 2012), this result may mean that the default facilitated gender assignment with masculine nouns.

The log odds of targetlike gender assignment were higher as the frequency with which learners used particular nouns increased and as their initial proficiency score increased. Additionally, the log odds of targetlike gender assignment were greater at in-stay and post-stay, compared to pre-stay, and the confidence intervals revealed that targetlike use was similar between in-stay and post-stay. The findings for these three factors showed that as learners become more proficient in the language, as they used individual nouns more often, and after they completed an academic year abroad, their knowledge of gender assignment, as seen through language production, improved.

Finally, for determiner type, the learner group exhibited higher log odds of targetlike gender assignment with definite articles compared to all other determiner types. The confidence intervals for all of the non-reference point categories overlapped, indicating that their log odds of targetlike use were similar. This finding was similar to Bruhn de Garavito & White (2002) who found that learners were more targetlike with definite articles than indefinite articles. It also appears to suggest that the assumption that gender marking on determiners signifies whether learners have assigned the targetlike gender to nouns needs to be nuanced, because all determiner types are not identical when it comes to assigning gender in language production.

Thus, returning to the assignment-agreement assumption, the current study's results do not align with previous research that considers gender assignment to be categorical (i.e., learners have either learned the gender of a noun or not, Alarcón 2010). Instead, they appear to support the observation that learners can show variable knowledge of a noun's gender in language use and that this variability is conditioned by a range of factors. More generally, they suggest that making assessments about the acquisition of gender assignment in language use involves an analysis that goes beyond a univariate examination of targetlike use of determiner gender.

Continuing with the mixed-effects regression for adjectives, eight fixed effects were significant: Noun ending, task, noun gender, noun log-frequency (language), initial proficiency, noun number, adjective position, and time signifi-

cantly impacted targetlike gender agreement. Noun class and noun frequency (individual) did not predict gender agreement. The interaction between time and initial proficiency was significant. However, because this interaction correlated with other main effects, I removed it from the model. The McFadden's R^2 indicated a moderate fit ($R^2_{McFadden}$ = 0.1563). The results for the fixed effects are available in Table 6 and the random effect results are in Table 7.

Table 6: Results for the fixed effects in the regression model for adjectives. *Note:* The reference point for the dependent variable is targetlike use.

Effect	Estimate	SE	z	p	Lower	Upper
Intercept	−0.94	0.84	−1.123	0.26	−2.69	0.741
Noun ending [canonical]						
deceptive	−2.26	0.33	−6.82	9.21×10^{-12}	−2.90	−1.589
other	−0.54	0.19	−2.88	0.004	−0.90	−0.16
predictive	−0.37	0.22	−1.71	0.09	−0.78	−0.06
Task [written]						
narrative	−0.65	0.27	−2.39	0.02	−1.19	−0.11
Interview	−0.35	0.19	−1.82	0.07	−0.74	0.02
Noun gender [masculine]						
feminine	−1.68	0.18	−9.25	$<2 \times 10^{-16}$	−2.05	−1.33
Noun log-freq.[a]	0.086	0.03	3.43	0.0006	0.04	0.14
Initial prof.	3.36	0.93	3.63	0.0003	1.52	5.30
Noun number [plural]						
Singular	0.43	0.15	2.82	0.005	0.13	0.74
Adj. position [other]						
pre	1.58	0.21	7.55	4.45×10^{-14}	1.17	1.99
post	0.91	0.20	4.46	8.38×10^{-6}	0.51	1.31
Time [pre-stay]						
in-stay	0.76	0.17	4.55	5.41×10^{-6}	0.43	1.09
post-stay	0.87	0.19	4.46	8.29×10^{-6}	0.49	1.25

[a](language)

Table 7: Results for the random effect in the determiner regression model

Participant	Intercept	Participant	Intercept	Participant	Intercept
150	0.41	160	−0.12	167	−0.72
151	−0.15	161	0.05	168	0.35
152	0.11	162	0.25	169	0.25
155	0.08	163	−0.41	170	−0.07
156	−0.32	164	0.11	171	−0.05
157	−0.17	165	0.23	172	−0.25
158	0.38	166	−0.02	173	−0.17

For noun ending, the log odds of targetlike gender agreement were lower with deceptively marked nouns and other noun endings compared to nouns with canonical endings and there was no significant difference between nouns with predictive endings and those with canonical endings. The confidence intervals indicated similarities (i.e., overlap) between other and predictive endings. The confidence intervals also demonstrated that the participants were less likely to use targetlike gender agreement with deceptively marked nouns compared to nouns with other and predictive endings. These findings suggest that targetlike gender agreement was most challenging for these learners when the noun has a deceptively marked ending. These results were similar those for determiners, which indicated that nouns with deceptive endings posed challenges for gender assignment.

For task, the log odds of targetlike gender assignment were lower with the oral narration compared to the written essay, and there was no significant difference between the oral interview and the essay. The overlap in the confidence intervals for the two oral tasks indicated that targetlike gender agreement is similar between the two. Although task constrained both gender assignment and agreement for these participants, it may be worth noting a difference between the two. Unlike the findings for gender assignment, which pointed to a difference between the oral and written modes, the interview task was statistically similar to both the essay and the oral narration in gender agreement.

The results for noun gender demonstrate that participants were less likely to be targetlike in their gender agreement with feminine nouns compared to masculine nouns. Just as with gender assignment, learners exhibited greater challenges with gender agreement when the nouns were feminine, perhaps pointing again

to the claim that the masculine gender is the default (López Prego & Gabriele 2012). Furthermore, participants were less likely to be targetlike in their gender agreement with plural nouns compared to singular nouns, which have also been considered to be a default for learners (López Prego & Gabriele 2012). The results for noun number constituted a difference between gender assignment and gender agreement, as this factor did not significantly predict targetlike use with determiners.

For the continuous factors, the log odds of targetlike gender agreement increased as noun log-frequency (language) increased; this factor was considered to be an indirect measure of input frequency (Gudmestad et al. 2019). The likelihood of targetlike gender agreement also increased as initial proficiency increases. In general, these findings demonstrated that experience with the language played a role in targetlike gender agreement. Moreover, while the results for initial proficiency were similar to those for gender assignment, the significant effects for frequency differed between determiners and adjectives. Noun frequency (individual) impacted gender assignment but noun log-frequency (language) constrained gender agreement.

Regarding adjective position, adjectives either before or after the noun in the same noun clause exhibited a higher log odds of targetlike gender agreement compared to adjectives that occurred outside of the noun clause, and there was overlap in the confidence intervals for the pre and post categories, indicating that targetlike use was similar between the two. In other words, the proximity between the noun and the adjective facilitated targetlike gender agreement.

Finally, the log odds of targetlike gender marking were higher at in-stay and post-stay compared to pre-stay and similar between in-stay and post-stay, indicating that learners' targetlike gender agreement improved during their academic year abroad and that this gain was maintained after returning home. This result was similar to the finding for gender assignment.

Thus, this multivariate analysis showed that noun ending, task, noun gender, noun log-frequency (language), initial proficiency, noun number, adjective position, and time were the factors that influenced targetlike gender agreement for this group of additional-language learners of Spanish. Considering the assumption that gender marking on adjectives is taken to reflect gender agreement, the findings can be interpreted to indicate that learners rely on a complex array of linguistic and extra-linguistic information in order to use this morphosyntactic property (i.e., agreement) in a targetlike way in language production.

4 Conclusion

Although it may seem obvious to say that an epistemology has bearing on research findings, it does not appear to be common in SLA for researchers to try out different perspectives in order to see where they lead in terms of the interpretation of data or to make this type of work publicly available. This is precisely what I set out to do in this chapter. In this vein, the current study has offered a reflection on the relationship between epistemology and methodology through a reanalysis of production data on grammatical gender in additional-language Spanish. This reanalysis was shaped by a shift in epistemology. In my previous collaborative project (Gudmestad et al. 2019), our assumption was that gender marking, with no distinction between agreement and assignment, was the linguistic issue under investigation. In the current study, however, I adopted a different perspective, one in which gender assignment and gender agreement were different learning processes that were manifested through gender marking on determiners and adjectives, respectively (cf. Alarcón 2010; Kupisch et al. 2013). Through the reanalysis of the data in Gudmestad et al., I explored, in the current chapter, possible methodological decisions that an investigation of gender assignment and agreement in language production might entail.

Under the assumption that gender assignment and gender agreement are different processes with different surface manifestations, the results from the present investigation's analysis can be interpreted as follows. First, the higher rates of targetlike use for determiners compared to adjectives support the understanding that gender assignment is acquired before gender agreement (Alarcón 2010). Second, regarding the examination of targetlike use with individual nouns that participants use more than once, the result that some nouns exhibited variability in targetlike use with determiners may indicate that, in language production, learners show evidence of variable knowledge of gender assignment, which is counter to what some researchers have suggested (e.g., Alarcón 2010). Moreover, the evidence of variability with individual nouns in the examinations of determiners and adjectives suggested that pursuing regression analyses in order to uncover the variable patterns of use was warranted. It is worth making explicit, however, that my observations about variable use in the current analysis and the methodological decision to pursue multivariate statistical analyses were influenced by the variationist orientation of my research program more generally (cf. Young 2018). The separate mixed-effects models for determiners and adjectives have resulted in three additional observations. One was that a range of factors help to account for when learners were more likely to show evidence of targetlike gender assign-

ment and agreement in language use. Another observation was that among the predictive factors, four impacted targetlike use on both determiners and adjectives: time, initial proficiency, noun gender, and noun ending. The epistemology that guided the present investigation may lead to expectations of finding some similarities between the two because gender assignment and agreement are related linguistic properties (i.e., they both deal with the gender of the noun). At the same time, though, the final observation that emerged from comparing the two mixed-effects models was that there were various differences in the factors impacting targetlike use between determiners and adjectives. In addition to finding that there were factors specific to each linguistic property that influenced use (determiner type and adjective position), the results also demonstrated that noun frequency (individual) only impacted gender assignment and that noun log-frequency (language) and noun number conditioned gender agreement only. Furthermore, although task was a significant constraint on both gender assignment and agreement, there were differences in the significant effects between the determiner and adjective models. These differences between the two mixed-effects models are expected, given the assumption that gender agreement and gender assignment are different learning properties that arguably have different acquisitional challenges. Thus, these multivariate analyses may be seen as bolstering to a degree the assignment-agreement assumption as they offered new details about how these learning properties differ in language use.

More generally, because the analysis in this investigation uncovered numerous differences between assignment/determiners and agreement/adjectives, it also led to different conclusions from those drawn in Gudmestad et al. (2019), even though some of the findings are similar (e.g., the role of noun gender).[6] While it is not novel to say that analysing data differently may lead to different observations, explicit reflections on how epistemologies shape the research process are crucial. Ortega (2014: 194) explains, "by applying different theories, some findings appear to change only in the details and yet they seem to bring different 'interlanguage truths' to the fore for consideration". In a similar vein, Young (2018: 48) reflects on how applied linguists gain new knowledge and argues that

---

[6]One example of a difference is that Gudmestad et al. (2019) found that targetlike use changed over time with regard to noun ending. However, in the current study neither the mixed-effects model for determiners nor the one for adjectives contained an interaction between time and another significant fixed effect. Another difference is that noun number impacted targetlike use with adjectives in the present investigation, but in Gudmestad et al. it was not a conditioning factor. An example of a similarity is that the two mixed-effects models in the current study and the one in Gudmestad et al. pertains to the factor of noun gender. Each regression analysis showed that the likelihood of targetlike use was higher with masculine nouns.

"we know what we attend to and the habits of mind of researchers – their personal preferences as researchers and the early training they received – to a large extent determine the questions researchers ask, the design and implementation of research studies, and the way data are interpreted". In sum, the current study has sought to contribute to methodological reflections in SLA by considering the important role that epistemology plays both in the analysis and interpretation of learner data and, as a consequence, in the advancement of new knowledge. Further scholarship on the connection between epistemology and methodology is important for SLA, because it demonstrates concretely the direct relationship between researchers' (at times implicit) assumptions and the types of observations they make when interpreting learner data. There is value in making these assumptions more explicit in published research in order to illustrate concretely that knowledge is not absolute.

References

Alarcón, Irma V. 2010. Gender assignment and agreement in L2 Spanish: The effects of morphological marking, animacy, and gender. *Studies in Hispanic and Lusophone Linguistics* 3(2). 267–299.

Alarcón, Irma V. 2014. Grammatical gender in second language Spanish. In Kimberly L. Geeslin (ed.), *The handbook of Spanish second language acquisition*, 202–218. Malden, MA: Wiley Blackwell.

Ayoun, Dalila. 2007. The second language acquisition of grammatical gender and agreement. In Dalila Ayoun (ed.), *French applied linguistics*, 130–170. Amsterdam: Benjamins.

Bartning, Inge. 2000. Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica* 54. 225–237.

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 Muñoz Liceras (eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection*, 143–160. New York: Bilingual Press.

Byrnes, Heidi. 2013. Notes from the editor. *The Modern Language Journal* 97(4). 825–827.

Creswell, John. W. & J. David Creswell. 2018. *Research design: Qualitative, quantitative and mixed methods approaches*. 5th edn. Los Angeles, CA: London.

Davies, Mark. 2016-. *Corpus del Español: Two billion words, 21 countries*. http://www.corpusdelespanol.org/web-dial/.

Dewaele, Jean-Marc & Daniel Véronique. 2001. Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition* 4. 275–297.

The Douglas Fir Group. 2016. A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal* 100. 19–47.

Ellis, Rod. 2006. Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics* 27. 431–463.

Geeslin, Kimberly. L. & Avizia Yim Long. 2014. *Sociolinguistics and second language acquisition: Learning to use language in context*. New York, NY: Routledge.

Gries, Stefan Th. 2005. Statistics for learner corpus research. In Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research*, 159–181. Cambridge: Cambridge University Press.

Grüter, Theres, Casey Lew-Williams & Anne Fernald. 2012. Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research* 28(2). 191–215.

Gudmestad, Aarnes, Amanda Edmonds & Thomas Metzger. 2019. Using variationism and learner corpus research to investigate grammatical gender marking in additional-language Spanish. *Language Learning* 69. 911–949.

Gudmundson, Anna. 2013. Acquisition of grammatical gender and number agreement in Swedish learners of L2 Italian: Regularity and frequency effects. *Language, Interaction and Acquisition* 4(2). 232–255.

Halberstadt, Lauren, Jorge R. Valdés Kroff & Paola E. Dussias. 2018. Grammatical gender processing in L2 speakers of Spanish: The role of cognate status and gender transparency. *Journal of Second Language Studies* 1(1). 5–30.

Kupisch, Tanja, Deniz Akpınar & Antje Stöhr. 2013. Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism* 3(2). 150–179.

Larsen-Hall, Jenifer & Luke Plonsky. 2015. Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning* 65(S1). 127–159.

López Prego, Beatriz & Alison Gabriele. 2012. Examining the nature of variability in gender and number agreement in native and non-native Spanish. In Alia K. Biller, Esther Y. Chung & Amelia E. Kimball (eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development*, 315–327. Somerville, MA: Cascadilla Press.

Marsden, Emma, Kara Morgan-Short, Pavel Trofimovich & Nick C. Ellis. 2018. Introducing registered reports at language learning: Promoting transparency, replication, and a synthetic ethic in the language sciences [editorial]. *Language Learning* 68. 309–320.

Marsden, Emma & Luke Plonsky. 2018. Conclusion: Data, open science, and methodological reform in second language acquisition research. In Aarnes Gudmestad & Amanda Edmonds (eds.), *Critical reflections on data in second language acquisition*, 219–228. Amsterdam: Benjamins.

Mitchell, Rosamond, Nicole Tracy-Ventura & Kevin McManus. 2017. *Anglophone students abroad: Identity, social relationships, and language learning*. New York, NY: Routledge.

Montrul, Silvina, Rebecca Foote & Silvia Perpiñán. 2008. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58. 503–553.

Ortega, Lourdes. 2005. Methodology, epistemology, and ethics in instructed SLA research: An introduction. *The Modern Language Journal* 89. 317–327.

Ortega, Lourdes. 2014. Trying out theories on interlanguage: Description and explanation over 40 years of L2 negation research. In Zhao-Hong Han & Elaine Tarone (eds.), *Interlanguage: Forty years later*, 173–201. Amsterdam: Benjamins.

Phakiti, Aek, Peter DeCosta, Luke Plonsky & Sue Starfield. 2018. Applied linguistics research: Current issues, methods and trends. In Aek Phakiti, Peter De Costa, Luke Plonsky & Sue Starfield (eds.), *The palgrave handbook of applied linguistics research methodology*, 5–29. New York, NY: Palgrave.

Plonsky, Luke (ed.). 2015. *Advancing quantitative methods in second language acquisition*. New York, NY: Routledge.

Plonsky, Luke & Frederick L. Oswald. 2017. Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition* 39. 579–592.

Porte, Graeme & Kevin McManus. 2018. *Doing replication research in applied linguistics*. New York, NY: Routledge.

Smith, Thomas J. & Cornelius M. McKenna. 2013. A comparison of logistic regression pseudo R^2 indices. *Multiple Linear Regression Viewpoints* 39(2). 17–26.

Teschner, V., Richard & William. M. Russell. 1984. The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics* 1. 115–132.

Young, Richard F. 2018. Habits of mind: How do we know what we know? In Aek Phakiti, Peter De Costa, Luke Plonsky & Sue Starfield (eds.), *The Palgrave handbook of applied linguistics research methodology*, 31–53. New York: Palgrave Macmillan.