A Longitudinal Study of Grammatical-Gender Marking in French as an Additional Language

Amanda Edmonds, Aarnes Gudmestad, Thomas Metzger

A longitudinal study of grammatical-gender marking in French as an additional language

Abstract

This investigation responds to the need for longitudinal data-driven research on additional-language acquisition by examining grammatical-gender marking among additional-language learners of French during a 21-month period, which included an academic year abroad (LANGSNAP corpus). The analysis of oral production consists of a generalized linear mixed model that examines a range of linguistic and extralinguistic factors shown to be important for gender marking in previous research, as well as a random effect for participant. Results show evidence of both change across time and consistency in the interlanguage. Drawing on variationism and usage-based approaches, we argue that longitudinal investigations that are focused on how learners use their additional language have much to offer our understanding of additional-language acquisition processes.

Keywords: Longitudinal, variationism, usage-based approaches, generalized linear mixed model, grammatical gender marking, French

Introduction

At its beginnings, the field of second language acquisition (SLA) was dominated by studies that offered rich descriptions of interlanguage, often by following a small number of learners over time (e.g., Hakuta 1974; Schumann 1978). In the following decades, and in response to diversification in theory and methods, the data at the heart of additional-language (AL) research also quickly diversified, with cross-sectional and experimental datasets becoming more prevalent. Although such data have obvious advantages, there has been a push in the last decade to reassert both the importance of longitudinal data in SLA and the worth of datasets that provide descriptions of learners in the process of making meaning as an important source of information on learning processes. This is visible in explicit calls for more longitudinal research (Ortega and Iberri-Shea 2005;

Hasko 2013) and in trends towards data-driven approaches to studying learner language, in which

descriptions of language use provide the basis for explanatory reflections (e.g., usage-based

approaches, Ortega, Tyler, Park and Uno 2016, variationist approaches, Geeslin with Long 2014,

dynamic systems theory, de Bot, Lowie and Verspoor 2007). We see both trends as positive

developments, as we believe that detailed accounts of language use over time have the potential to

provide unique contributions to knowledge about learning processes in real time. In the current

study, we participate in these two trends with an analysis of grammatical-gender marking in AL

French. This morphosyntactic structure was selected for study because of its high frequency in

language use and because it has been extensively researched, thus providing us with a large body of

previous findings from which to identify potential factors impacting use and, ultimately, on which to

build new knowledge. We draw on insights from usage-based and variationist approaches to analyze

data from the longitudinal LANGSNAP corpus[1] (see Mitchell, Tracy-Ventura and McManus 2017),

pulling together the rich – albeit piecemeal – insights from previous research in a single generalized

linear mixed model. We thus contribute both to the need for better description of learning over time

and to reflections on factors that explain the observed learning trajectory.

Background

We begin by briefly discussing the importance of longitudinal data in SLA. We then move to data-

driven approaches to SLA that provide both descriptions and explanations of language use, focusing

on usage-based approaches and variationism. We end this section with a presentation of

grammatical gender in French and previous research on gender in AL French.

Longitudinal AL research

In the last decades, AL acquisition research has been dominated by cross-sectional studies. While

such research has allowed the field to amass an important body of knowledge, calls for more

longitudinal data have increased. Ortega and Iberri-Shea (2005), Hasko (2013), and Tracy-Ventura

and Huensch (2018), among others, have highlighted the unique contribution that longitudinal

studies make to the understanding of AL acquisition. Focusing on the what the field loses by not engaging in more longitudinal research, Ortega and Byrnes (2008: 3) state that "the lack of an explicit or sustained focus on longitudinal questions has meant that after some 40 years of disciplinary history, we know little about the longitudinal pace and pattern of development in second language and literacy." One recent response to calls for longitudinal AL research has been the development of longitudinal learner corpora, which follow learners over time, providing the possibility of detailed, real-time investigations of AL acquisition (Myles 2008). However, such corpora "continue to be a rarity" (Meunier and Littré 2013: 63), despite their potential to provide a fertile testing ground for numerous questions relevant to AL research. With respect to grammatical-gender marking, the focus of the current study, previous research has been largely cross-sectional, leading to numerous observations about how learners mark gender in their AL. However, it is only with longitudinal datasets that we can determine whether these observations are valid across the learning trajectory. In the current study, we build on past cross-sectional analyses of gender-marking behavior in AL French to investigate the evolution of grammatical-gender marking in a longitudinal corpus.

Data-driven approaches in SLA

Selinker's seminal 1972 article advocated for the importance of data reflecting language use in AL research, stating that without clear descriptions of language use in an AL, experiments would be premature and their results, confusing (see p. 224). In her discussion of Selinker (1972), Ortega (2014: 176) agrees with the importance of description, although she recognizes that "interlanguage researchers have always felt the need to go beyond description" in an attempt to explain the processes and mechanisms at play in AL acquisition. In the now varied landscape of SLA, several analytical approaches are strongly data-driven and devoted to both detailed description and fine-grained explanations of language-learning processes. Two such approaches informed the design and the interpretation of results of the current study: usage-based approaches and variationism. We argue that these two approaches have much in common and that their unique strengths result in

complementary insights to the study of grammatical-gender marking (see Reif, Robinson and Pütz 2013 for studies drawing on both approaches).

Within SLA, usage-based approaches have gained ground in recent years (Ellis, Römer and O'Donnell 2016; Ortega et al. 2016). According to Tummers, Heylen and Geeraerts (2005), usage-based approaches cover several overlapping theories and approaches, including Construction Grammar, Emergent Grammar, and much research done within phraseology. In their view, the family of usage-based approaches shares several assumptions, namely that language use should be at the heart of linguistic inquiry and that the competence-performance and syntax-lexicon dichotomies are untenable. A corollary of these shared assumptions is that mental grammar emerges out of concrete experience with language (Bybee 2006), meaning that "language acquisition happens in and through use" (Ortega 2014: 193). Researchers adopting a usage-based approach within SLA (e.g., Cadierno and Eskildsen 2015; Ellis and Wulff 2015; Ortega et al. 2016; Jach 2018) view language learning as a situated, embodied process in which learners engage (via implicit learning mechanisms) with input, showing sensitivity to various patterns and regularities present in the target language. Characteristics of the input that tend to facilitate learning include higher frequency and greater salience of forms (i.e., sounds, words, constructions), as well as reliable mappings between forms and their functions (Ellis et al. 2016). Usage-based studies on AL acquisition have, among other things, documented how learners' AL productions provide evidence for the building up of language knowledge from the input (e.g., Eskildsen 2009) and modeled learners' sensitivity to a variety of linguistic factors in their use of structures (e.g., Wulff, Lester and Martínez-García 2014). Understanding how language use and input characteristics act as drivers of learning is at the heart of usage-based approaches to AL acquisition.

Research within the variationist paradigm emphasizes that language is characterized by systematic variation, which can be conditioned by linguistic (e.g., phonological environment, syntactic position) and/or extralinguistic (e.g., register, time) factors. The mention of extralinguistic factors is important: Variationism was developed within sociolinguistics (Labov 1966), and the socially situated and contextually determined nature of language use has been at the center of this

approach since its beginnings. Within SLA, most variationist research has focused on how learners

acquire structures that are variable in target-language input. These studies examine whether

interlanguage is constrained by similar linguistic and extralinguistic factors as native-speaker use.

Although less common, variationist approaches can also be utilized to describe and explain

interlanguage variation, meaning variation present in AL development but not in native-speaker use

(see Young 1991; Berdan 1996; Bayley and Langman 2004). The focus of the current study (variation

in grammatical-gender marking) constitutes an example of interlanguage variation.

Usage-based and variationist approaches in SLA have much in common, including a focus on

language use and a rejection of the performance-competence dichotomy. In addition, both draw on

similar analytic techniques – namely multivariate regressions – to model language use (see Geeslin

with Long 2014 for variationism and Gries 2015 for usage-based approaches). However, each

approach brings different complementary strengths to the study of AL acquisition, which we draw on

in the current project. First, usage-based approaches underscore input characteristics as drivers of

learning. In the current project, usage-based reflections on the role of one such driver – frequency –

influenced our analysis and interpretation. As for variationism, this theory highlights the importance

of linguistic but also extralinguistic factors in explaining systematic variation, both of which will be

examined in the current study.

Gender marking in French

All French nouns have one of two genders: masculine or feminine. Whereas gender assignment

appears to be semantically motivated for nouns with biological gender (e.g., *frère*masc 'brother',

*sœur*fem 'sister') and for the members of a small number of semantic classes (e.g., languages, colors:

see Nelson 2005), researchers generally agree that the gender of the remaining nouns is not

determined by meaning. Instead, most accounts of grammatical gender in French have highlighted

the fact that morphophonological and orthographic noun endings constitute cues to the noun's

gender (see Tucker, Lambert and Rigault 1977; Lyster 2006; cf. Carroll 1989).[2] Tucker et al., for

example, found that the gender of approximately 80% of all nouns was predictable from their

morphophonological endings. The remaining one-fifth of nouns include homonyms (*la juge*fem 'the judge' vs. *le juge*masc 'the judge') or homophones (*le sel*masc 'the salt' vs. *la selle*fem 'the saddle') that differ with respect to gender, as well as nouns whose ending is predictive of the opposite gender. For example, nouns that end in <ette>/[ɛt] are generally feminine in French, whereas *squelette* 'skeleton' is masculine.

The expression of grammatical gender in French occurs on pronouns, past participles, and nominal modifiers (e.g., articles, demonstratives, adjectives). Only the third category will be examined in the current study, and it is important to point out that not all modifiers are informative of a noun's gender. For example, the plural definite article *les* and adjectives that end in an orthographic <e> (*jaune* 'yellow') have the same phonetic and orthographic form, regardless of whether the noun they modify is masculine or feminine. Other examples, such as *quel*masc/*quelle*fem 'which' and *fatigué*masc/*fatiguée*fem 'tired', show an orthographic gender distinction, although the pronunciation of the members of the two pairs is identical. In addition, it bears note that vowel-initial nouns have fewer gender-specifying contexts than those that begin with consonants, due to phenomena of elision (*l'année* 'the year'), liaison (*petit ami* 'boyfriend'), and the use of suppletive forms for certain determiners (*mon ami* 'my friend') and adjectives (*le nouvel an* 'the New Year'). In each of these cases, in oral French the modifier provides no information about noun gender, as the same phonological form is used before all vowel-initial nouns. This brief presentation of gender in French is not meant to be exhaustive (Ayoun 2007: 130-170 provides a thorough account), but rather serves to underscore the fact that the grammatical-gender cues available in the input are complex.

Acquisition of grammatical gender in AL French

The acquisition of grammatical gender involves several related learning problems (Carroll 1999: 45-51; Meisel 2018: 660-661). These include determining whether the language being learned has grammatical gender and, if so, how many gender categories are present. For gendered languages like French, each noun must be assigned to a gender category. In addition, the learner must determine the categories on which gender is marked (e.g., modifiers) and the domain over which this

agreement relationship holds. Previous research on gender in AL French has focused on one or more of these learning problems. First, numerous scholars have explored whether learners whose first language does not have grammatical gender (like English) can in fact acquire grammatical gender in AL French. In a review of previous literature having examined this issue, Prévost (2009: 318) concludes that the gender feature can, in fact, be acquired by such learners. With respect to the learning problem of gender assignment, past research has identified factors that influence learners' behavior. There is robust evidence that learners of French from different first-language backgrounds are more targetlike with masculine nouns (Harley 1979; Holmes and Dejean de la Bâtie 1999; Bartning 2000; Dewaele and Véronique 2001; Edmonds and Gudmestad 2018), suggesting that masculine gender may be assigned to nouns as a default. Other factors found to impact gender assignment include the first phoneme of the noun, with learners being less targetlike with vowel-initial nouns (Dewaele 2015), and noun frequency, where Surridge and Lessard (1984) reported greater success at assigning gender to frequent nouns. Finally, in studies involving native English-speaking learners of French, Hardison (1992) and Surridge and Lessard (1984) provided evidence that learners were sensitive to noun endings in assigning gender. With respect to gender agreement, it has been shown that learners are more targetlike in expressing agreement on determiners as opposed to on adjectives (Harley 1979; Bartning 2000; Dewaele and Véronique 2001; Granfeldt 2005; Ayoun 2007). Moreover, the distance between the noun and its modifier may negatively impact agreement (Prodeau 2005; Edmonds and Gudmestad 2018) as distance introduces "an additional memory component" (Holmes and Dejean de la Bâtie 1999: 500). Finally, Bartning (2000) and Ayoun (2007) both reported that more advanced learners show greater mastery in the expression of grammatical gender.

Taken together, research has examined the role of a wide array of factors on grammatical-gender marking in AL French:

(1)   Noun gender
        Noun-initial phoneme
        Noun frequency

Noun ending
Modifier type
Distance between noun and modifier
Proficiency level

Despite a sustained interest in gender marking in SLA generally and in AL French specifically, evidence for the factors in (1) remains largely piecemeal. In other words, robust descriptions and explanations based on analyses – be they quantitative or qualitative – of a single dataset that simultaneously consider a complex array of factors are conspicuously lacking with respect to the acquisition of this morphosyntactic phenomenon. This situation is reminiscent of what Ortega (2014: 177) characterized as a "puzzle […] whose pieces lie around while its meaning fails to emerge." We suggest that the piecemeal nature of previous findings is, in part, due to methodological choices. In particular, previous studies have focused on only one or a small number of these factors, leaving open the question of whether the full set significantly constrains gender-marking behavior and, if so, how. In addition, previous research has not described the learning trajectory using longitudinal data; those researchers who have addressed development have done so with cross-sectional designs (e.g., Bartning 2000; Ayoun 2007).[3] Thus, not only is it unclear what role the factors in (1) may play in concert with one another, but we also do not know how influential factors may change over time. In the current study, we focus on language use and draw on usage-based and variationist approaches to language to respond to the following research question: What factors predict variable gender marking in oral AL French, and how do they change over 21 months?

Method Section

Participants

The LANGSNAP project followed a total of 56 learners of Spanish and French over a period of 21 months. In the current article, we analyze the oral data from the first 20 learners of French followed by the LANGSNAP team (see Gudmestad, Edmonds and Metzger, in press, for an analysis of grammatical-gender marking by a subset of the learners of Spanish). The group includes two men

and 18 women, all of whom had obtained an end-of-secondary-education qualification in French and were enrolled in a French degree program in England throughout the project. At the outset, this subset of participants had studied French for an average of 10.45 years (*SD*: 2.3; range: 6-20)[4] and were on average 20 years old (*SD* = 0.38; range: 19-21). A measure of proficiency in AL French was obtained via an elicited imitation task administered at the beginning of the project, resulting in a score out of 120 points (Tracy-Ventura, McManus, Norris and Ortega 2014). The 20 participants for our project obtained between 36 and 97 points (*M* = 59.25, *SD* = 14.58). The participants all spent an academic year in France and were involved in three different activities during their stay: workplace internships (*n* = 3), university exchanges (*n* = 5), and teaching assistantships in elementary and secondary schools (*n* = 12). This difference will not be further explored in our study, as Mitchell, McManus and Tracy-Ventura (2015) demonstrated that placement type played no role in oral expression gains for learners in the LANGSNAP corpus.

Data collection

The LANGSNAP project collected data from each participant at six different points in time: before their stay abroad, three times during the academic year abroad, and twice after their return to England. At each data-collection point, the participant wrote an argumentative essay and completed two oral tasks (an interview and a narrative) in French. The semi-guided interviews lasted approximately 20 minutes and the narrative task required learners to tell a picture-book story in their own words.

The current article analyzes the oral data, using the transcripts for interviews and narratives completed by the 20 participants at three points in the project: the beginning (pre-test: May/June 2011), the third visit abroad after learners had been in France for approximately 9-10 months (May/June 2012), and the end of the project, approximately 21 months after pre-test (January/February 2013). These three points will be referred to as pre-stay, in-stay, and post-stay, respectively. For the narrative, the same story was used for at pre-stay and in-stay, whereas a

different story was administered at post-stay. All data were recorded and then transcribed. Our analysis is based on the transcriptions provided on the LANGSNAP website.

Data coding

We analyzed every example of gender marking between a noun and either a determiner or an adjective, for a total of 11,218 occurrences of potential gender marking. A single noun could give rise to multiple tokens, as in (2):

(2)   *avec euh des*$_{masc/fem}$ *autres*$_{masc/fem}$ *étudiants*$_{masc}$ *anglaises*$_{fem}$ (Participant 118, interview, in-stay)[5]
        'with uh some other English students'

Here, the masculine noun *étudiants* is modified by one determiner (*des*) and two adjectives (*autres, anglaises*), resulting in three different tokens. However, not all were retained for the final analysis, given that gender marking on certain modifiers is not phonetically explicit in French. Thus, although it is clear that Participant 118 pronounced a masculine noun (*étudiants*) modified by a feminine adjective (*anglaises*), it is impossible to know whether *des* and *autres* are marked for feminine or masculine, as an invariant form for these two modifiers is used with all nouns. For this reason, such examples were removed from the final dataset. We faced a similar issue with a small set of nouns, including *ami*$_{masc}$/*amie*$_{fem}$ 'friend'. Although spelled differently, these nouns are pronounced identically in their masculine and feminine forms, thus making it impossible (on the basis of the pronunciation of the noun) to know which gender was intended by the speaker. Such examples were removed. LANGSNAP learners also used some interlanguage nouns and adjectives. Interlanguage nouns were excluded, as it was not possible to assign gender to such forms. Interlanguage adjectives, such as *les Français sont très agréants* 'the French are very *agréants*' (Participant 108, interview, pre-stay), were treated differently. Indeed, although the meaning of *agréant* is not entirely clear, the learner's pronunciation of this form clearly indicates that the adjective is marked for masculine gender. For this reason, such examples were included in our analysis. The final dataset, which contains only those observations in which gender marking was audible, amounts to 5,992 tokens, and includes 684 different nouns (305 feminine, 379 masculine).

The 5,992 observations were coded for our binary dependent variable, namely whether gender marking on the modifier was targetlike. When an occurrence was coded as targetlike, this meant that there was a match between the gender expressed on the modifier and the gender of the noun in French. To learn more about which factors collectively impacted variability in targetlike use, we then coded for 14 independent factors (Table 1). In order to identify factors potentially of interest in modeling the expression of grammatical gender, we followed the example of variationist approaches and looked to previous research on native and AL French.

Table 1. *Independent factors*

Factors	Coding categories
Initial proficiency	36-97
Time	pre-stay vs. in-stay vs. post-stay
Task	interview vs. narrative
Noun gender	feminine vs. masculine
Noun log-frequency (language)	0-1773.62
Noun frequency (individual)	1-27
Noun semantic class	yes vs. no
Noun class	grammatical gender vs. biological gender
Noun-initial phoneme	consonant vs. vowel
Noun-final phoneme	consonant vs. vowel
Noun-final orthographic rhyme	not predictive vs. predictive
Noun derivational morphology	absent vs. present
Modifier type	adjective vs. determiner
Syllable distance	0-16

For each token, we noted which participant uttered each token and his or her initial proficiency, as measured by the elicited imitation task (these scores were scaled by 100 to facilitate model fitting). Tokens were also coded for the data-collection time and task during which they were uttered.

Nine independent factors concern properties of the noun. For the factor noun gender, we noted the gender of the noun in French. As for noun frequency, we coded this factor in two different manners, in light of recent discussions of frequency effects within usage-based approaches (Ellis and Wulff 2015: 84). The first measure – noun log-frequency (language) – reflects the frequency of the noun in French and was determined by consulting the film portion of the *Lexique 3.8* database and by taking the lemmatized frequency counts per million words (New, Brysbaert, Veronis and Pallier 2007). We log-transformed these frequency counts to reduce skewness in the distribution. Whereas

this first operationalization of frequency provides an indication of possible input, the second

measure – noun frequency (individual) – looks at individual use. Indeed, as specified by usage-based

researchers such as Bybee (2006) and Tummers et al. (2005), mental grammar develops through

language use and, thus, an individual's usage patterns are expected to impact his or her mental

grammar. For this factor, we calculated how frequently each individual participant used a given noun

with a gender-marked modifier in each task and at each data-collection period. In addition, each

token was coded for whether the noun belonged to a semantic class predictive of gender (e.g.,

colors, languages), for whether the nominal referent showed grammatical or biological gender, and

for the first phoneme (consonant or vowel) of the noun. Additionally, three variables focused on the

ending of the noun: (a) noun-final phoneme identified the final phoneme as a consonant or vowel,

(b) noun-final orthographic rhyme used the predictability information provided by Lyster (2006) to

determine whether each noun's final orthographic rhyme was predictive of the noun's gender

(meaning that at least 89% of all nouns with that ending have the same gender), and (c) noun

derivational morphology coded for the presence of a derivational morpheme (following Surridge

1989).

Finally, two aspects of the modifier were coded. Each token was coded for whether the

modifier was an adjective or determiner (modifier type). Determiners included definite and indefinite

articles, indefinite determiners (e.g., *aucun* 'no'*, chaque* 'each'), demonstratives, possessives, and

interrogative determiners. Adjectives could occur both within the noun phrase (in either pre or

postnominal positions) and in predicative position. With the factor syllable distance, we recorded the

distance (in syllables) between the noun and the modifier, thus providing a measure of linear

distance between the noun and the modifier (cf. Holmes and Dejean de la Bâtie 1999: 500, who

suggest that greater distance results in higher memory costs). Modifiers located directly next to the

noun were coded "0"; the modifier located the farthest from its noun was found at a distance of 16

syllables. A sample of our coding is provided in Appendix A of the Supplemental Material.

With the exception of participant, all factors listed in Table 1 were examined as fixed effects in a generalized linear mixed model (Gelman and Hill 2006). After the significant fixed effects had been identified, we examined whether there was evidence of development over time by testing for the significance of the interaction between each significant fixed effect and the factor of time. As for the variable of participant, a random intercept was included for each participant. This factor enabled us to account for variability at the level of the learner while simultaneously examining variability in aggregate data. Moreover, treating participant as a random effect has the advantage of allowing for the possibility of generalizing beyond the sample of 20 learners in this corpus.

Data analysis

Following usage-based and variationist approaches, we conducted a multivariate analysis to explain how learners vary their gender-marking behavior according to a range of linguistic and extralinguistic factors and to determine how this variable use changes over time. We began by calculating the average targetlike rate of use for each individual and for the group at each data-collection point. Data were then analyzed using the statistical software R (for initial chi-square analyses and bootstrapping) and SAS (for the generalized linear mixed model). We checked for strong correlations between the independent variables using chi-square methods; when two effects were strongly correlated, we retained the effect that was best motivated by previous research. We then used bootstrapping on the remaining fixed effects to explore their potential importance. This step allowed us to narrow down the set of fixed effects further examined in this study. We then fit a single generalized linear mixed model to the full dataset, and non-significant factors were removed. Next, we explored whether any of the significant fixed effects interacted with time, using a top-down strategy to remove non-significant interactions. After having fit the model, we reexamined correlations between fixed-effect estimates for significant factors to check for the presence of strongly correlated covariates. Next, for the interactions and the one nominal factor with more than two levels (time), multiple comparisons between levels were carried out using the Tukey-Kramer method, which adjusts confidence interval widths to reduce Type-I error rate. Finally, we assessed

the quality of our model by calculating three different metrics: the percentage of correctly predicted observations, McFadden's R^2 (used instead of R^2 because the dependent variable in our model is a binary response), and the Bayesian Information Criterion (BIC).

Results

We begin by presenting rates of targetlike use across the 21 months represented in the corpus. We then provide the results from the chi-square and bootstrapping phases, which allowed us to specify the fixed effects examined in our model, before presenting the generalized linear mixed model.

Rates of targetlike gender marking

The boxplots with overlays in Figure 1 provide a visual representation of how targetlikeness with respect to gender marking evolved over time for the group and for each of the 20 participants. For each data-collection period (represented along the x-axis), the group average is presented as a diamond, the standard deviation as arrows, and the median is shown as the bold horizontal line. Before the 20 learners went abroad, a total of 2,242 examples of phonetically overt gender marking could be analyzed, of which 1,787 (79.7%) showed a match between the gender of the noun and the gender marked on the modifier. Overall targetlikeness increased at in-stay, with 85.6% (1,910/2,232) of all instances being targetlike. At post-stay, the average rate of targetlikeness for the group is 88.2% (1,339/1,518). In addition to providing information relative to the group, Figure 1 provides the targetlike rate of use for each individual; each dot represents the average targetlike rate of use for a single participant.
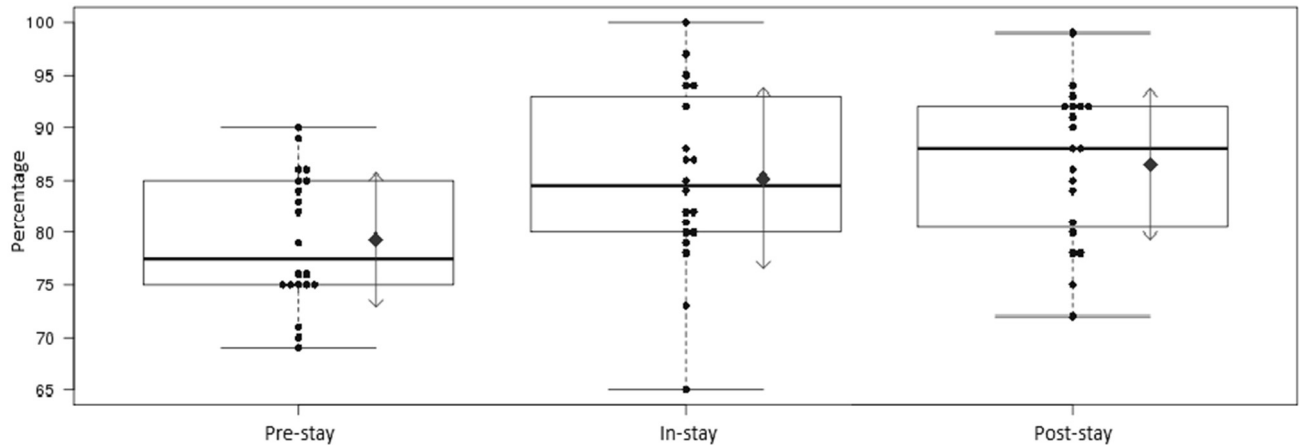
Figure 1. *Boxplots with individual overlaps for targetlike rates of use over time.*

A model of variable gender-marking behavior

Chi-square methods identified a high correlation between noun-final orthographic rhyme and modifier type. Given that noun ending was coded for in three different ways (noun-final orthographic rhyme, noun-final phoneme, noun derivational morphology), noun-final orthographic rhyme was removed and modifier type was retained. Bootstrapping for the remaining factors in Table 1 showed that two did not appear to account for significant variability in gender-marking behavior in this dataset. Thus, noun semantic class and noun derivational morphology were not explored in the generalized linear mixed model. When the model was run with the remaining fixed factors, the variables of task and noun log-frequency (language) were non-significant. These factors were removed and the model was re-run. The generalized linear mixed model identified nine independent fixed effects and two interactions with the factor of time that contributed significantly to the final model of variable gender-marking behavior.[6] Model details are provided in Table 2 (see Appendix B of the Supplemental Material for the participant random effect).

Table 2. *Generalized linear mixed model.*

Factors	Parameter estimate	*SE*	df	t value	*p* value	Confidence intervals	
						Lower	Upper
(Intercept)	1.2364	0.4152	18	2.98	0.0081	0.364	2.1087
Initial proficiency	0.01901	.006152	18	3.09	0.0063	0.006083	0.03193
Noun frequency (individual)	0.1132	0.02241	5959	5.05	< .0001	0.06925	0.1571
Syllable distance	-0.2993	0.03954	5959	-7.57	< .0001	-0.3768	-0.2218
Time [pre-stay]							
in-stay	0.729	0.1509	38	4.83	< .0001	0.4234	1.0345
post-stay	1.1195	0.191	38	5.86	< .0001	0.7329	1.5062
Noun gender [masculine]							
feminine	-0.781	0.116	19	-6.74	< .0001	-1.0237	-0.5383
Noun class [biological]							
grammatical	-0.3785	0.1539	19	-2.46	0.0237	-0.7007	-0.05626
Noun first phoneme [consonant]							
vowel	-0.253	0.1146	19	-2.21	0.0397	-0.4927	-0.01318
Noun final phoneme [vowel]							
consonant	-0.2264	0.07929	19	-2.86	0.0101	-0.3924	-0.06045
Modifier type [determiner]							
adjective	-0.8592	0.1256	19	-6.84	< .0001	-1.122	-0.5963
Time x Noun gender [pre-stay and masculine]							
in-stay x feminine	-0.5775	0.1774	38	-3.26	0.0024	-0.9366	-0.2183
post-stay x feminine	-0.3745	0.2143	38	-1.75	0.0886	-0.8083	0.05936
Time x Modifier type [pre-stay and determiner]							
in-stay x adjective	0.1923	0.1896	38	1.01	0.317	-0.1916	0.5761
post-stay x adjective	-0.4607	0.2156	38	-2.14	0.0391	-0.897	-0.02427

Note. For nominal variables, reference categories are given in square brackets. The reference point for the dependent variable is targetlike use.

The parameter estimates indicate the direction of the significant effect. Starting with the three continuous variables, we see that the estimates for initial proficiency and noun frequency (individual) are positive, indicating that the log-odds of using targetlike gender increase both as initial proficiency increases and as an individual uses a given noun more frequently with a gender-marked modifier. For syllable distance, the parameter estimate is negative, signifying that as the distance between the noun and the modifier increases, the log-odds of the gender of the modifier and noun matching decrease.

Moving now to the nominal variables, here the effects on log-odds are calculated by designating one category of the independent variable to be the reference category; the other categories are then compared against this reference category. Beginning with the factor of time, the results from in-stay and post-stay were compared to those from the pre-stay data-collection (reference category), and in both cases, the parameter estimates are significant and positive. This indicates that the learners were more likely to mark gender in a targetlike manner at in-stay and at post-stay, as compared to before leaving for France. As this variable has more than two categories, one might question whether there is a significant difference between the two non-reference categories. In other words, did learners show significant change between in-stay and post-stay? The confidence intervals, provided in Table 2, can be used to respond to this question. The intervals for in-stay and post-stay overlap, indicating that the two categories have statistically similar log-odds of targetlikeness and, thus, that the difference in gender-marking behavior at these two times is not considered different. The next nominal variable is noun gender, and here we see that the gender of the modifier and noun are less likely to match with feminine (vs. masculine) nouns. For noun class, learners were significantly less likely to be targetlike when marking gender with nouns with grammatical as opposed to biological gender. The phonological form of the noun was also found to significantly influence gender marking behavior: The log-odds of marking gender in a targetlike manner were lower when a noun began with a vowel and when a noun ended in a consonant. The final main effect concerned the variable of modifier type. The gender expressed on the modifier was

less likely to match the gender of the noun when the modifier was an adjective, as opposed to a determiner.

In addition to the main effects, two significant interactions with time were identified. Beginning with the time by noun gender interaction, we found that the log-odds of targetlikeness were lower for feminine nouns at in-stay (as compared to the reference categories of masculine and pre-stay). The interaction plot in Figure 2 illustrates this result. Here we see that although log-odds of targetlikeness increase for both masculine and feminine nouns between pre-stay and in-stay, the increase is steeper for masculine nouns (i.e., the gap between masculine and feminine nouns is larger at in-stay than at pre-stay). At post-stay, the difference between masculine and feminine nouns has reduced, as compared to in-stay.
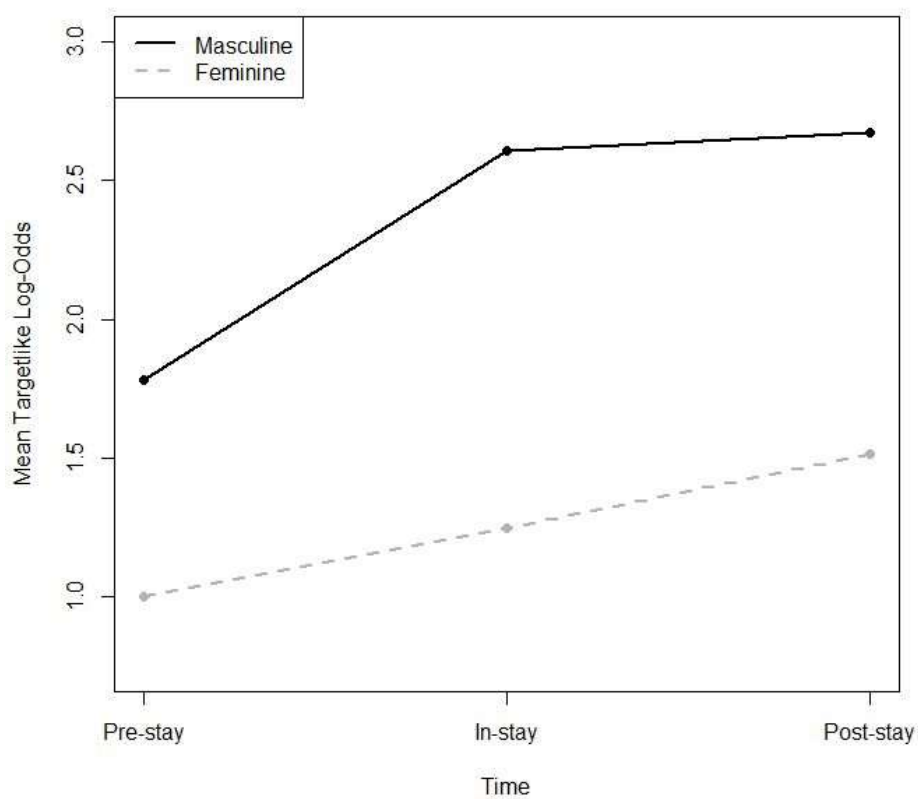

Figure 2. *Interaction plot for noun gender x time interaction.*

Time also significantly interacted with the factor of modifier type: Learners were significantly less likely to be targetlike with adjectives at post-stay as compared with determiners and pre-stay data.

Figure 3 shows that whereas log-odds of targetlike use with determiners increase from pre-stay to in-stay to post-stay, a different pattern is seen with adjectives. Log-odds of targetlike marking on adjectives increase between pre- and in-stay but then decrease between in- and post-stay. A large gap between behavior with determiners and adjectives is clearly visible at post-stay.
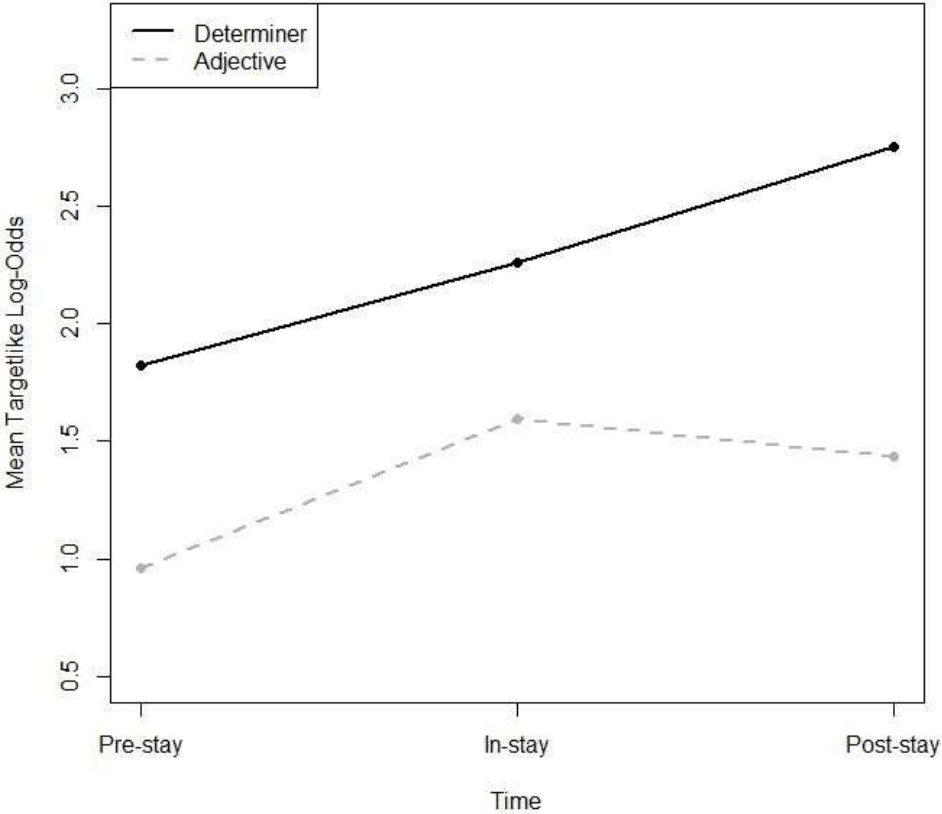

Figure 3. *Interaction plot for modifier type x time interaction.*

The quality of the model was assessed using three measures, all of which show that the generalized linear mixed model does a good job of fitting targetlike gender marking. First, this model predicted targetlike behavior with 85.3% accuracy. Second, we calculated McFadden's R^2 and obtained $R^2_{McFadden} = 0.1498$, indicating a moderate to good fit (see Smith and McKenna 2013 for the interpretation of this measure). Finally, the BIC compared our model with a null model (containing only the dependent variable) based on their log-likelihoods. This measure avoids overfitting by penalizing models with potentially extraneous parameters. A difference greater than 10 between the two models' BICs is considered to be strong evidence against the model with the higher BIC (Kass and

Raftery 1995). Despite the penalty for including nine fixed effects, two interactions, and a random effect for each participant, our model has a better (lower) BIC than the null model: 4575.56 (our model) versus 5268.67 (null model).

Discussion

We begin our discussion by responding to the research question that guided this study. We then discuss what a longitudinal, data-driven approach to the study of grammatical-gender marking has contributed to knowledge about AL acquisition. In particular, we argue for the importance of the use of longitudinal data, and we highlight the contributions of the current project with the help of insights from usage-based and variationist approaches to language.

Gender-marking behavior in a longitudinal corpus

What factors predict variable gender marking in oral AL French, and how do they change over 21 months? We analyzed 5,992 instances of overt gender marking using a generalized linear mixed model, which identified nine significant independent factors that conditioned variable targetlike use. Certain results lend additional support to trends identified in previous research, whereas others bring new findings to the fore. Beginning with the results that support previous research, we found that learners whose initial proficiency was higher at the outset of the project were more likely to mark grammatical gender in a targetlike manner (Bartning 2000; Ayoun 2007). Additionally, learners were less likely to express gender in a targetlike manner with feminine nouns (e.g., Bartning 2000) and when the noun began with a vowel (Dewaele 2015). Finally, log-odds of matching the gender expressed on the modifier with the gender of the noun were lower when the modifier was an adjective (e.g., Granfeldt 2005; Ayoun 2007) and as the number of syllables separating the modifier from the noun increased (Prodeau 2005).

In addition, four new findings emerged regarding learners' variable use of targetlike gender marking. First, these learners were less likely to be targetlike when a noun showed grammatical (vs. biological) gender. Whereas Carroll (1999) showed that Anglophone learners of French were able to

recognize and generalize biological gender using an experimental task, the current study demonstrates that learners use this distinction in producing AL French. Second, noun-final phoneme was significant, with the likelihood of targetlike gender being lower with consonant-final nouns. We know from previous research that learners are sensitive to noun endings in off-line gender-assignment tasks (Surridge and Lessard 1984; Hardison 1992); this study offers evidence that this sensitivity impacts language use. Third, these learners showed greater likelihood of marking grammatical gender in a targetlike manner with a given noun when they used that noun more frequently with modifiers overtly marked for gender. This operationalization of frequency is novel in the study of grammatical gender, and we return to its implications later in the discussion. Finally, although previous cross-sectional research has suggested that gender-marking behavior may change over time, the current study documents this change in a longitudinal dataset. Time was found to be a significant main effect in our model: Gender-marking behavior was significantly more targetlike at in-stay and at post-stay (vs. pre-stay), although targetlikeness was similar between the end of the stay abroad (in-stay) and the final data-collection period 8 months later (post-stay). In addition, two significant interactions with time were found. Thus, nine different linguistic and extralinguistic factors were found to be significant in this analysis, indicating that they work together to predict gender marking in oral AL French in the corpus investigated. Moreover, the impact of two of the linguistic factors changes over the 21-month period investigated.

What we gain from longitudinal, data-driven approaches to AL acquisition

Numerous researchers have argued for the necessity of longitudinal data in order to track learning in real time, thereby providing insights into the "longitudinal pace and pattern of development" (Ortega and Byrnes 2008: 3). The results from our longitudinal investigation indeed reveal trends that would be difficult to detect without longitudinal data. More specifically, many previous studies have reported that learners of AL French are more targetlike with masculine (vs. feminine) nouns and with determiners (vs. adjectives). Our investigation provides additional support for these findings, but importantly also nuances them. More specifically, each of these two factors interacted with time,

indicating that although this group of learners (like many groups before them) was overall less targetlike with feminine nouns and with adjectives, these global trends changed across time, patterns which would not be readily visible in cross-sectional data. In the current investigation, in which the learners did a stay abroad in a target-language environment, this evolution may result from the change in learning context. For the noun gender by time interaction, Figure 2 shows that whereas progress was seen with both masculine and feminine nouns after an academic year abroad, gains were steeper with masculine nouns. This suggests that gender marking with masculine nouns in oral production benefited more from the stay-abroad experience than did feminine nouns, which may reflect the fact that masculine nouns are more frequent than feminine ones in French (see Surridge and Lessard 1984: 46). As for the modifier type x time interaction, Figure 3 shows improvement with both determiners and adjectives after a year abroad. However, after learners' return to England, gender marking with determiners continued to progress, whereas behavior with adjectives regressed. This suggests that the stay abroad was effective in supporting development in gender marking on all modifiers, but that the home environment may not have provided sufficient, continued support for adjectives. This finding may reflect differences in the distribution of adjectives and determiners in French: Given that adjectives are less frequent than are determiners, and that there is more variety in the open class of adjectives (vs. the closed class of determiners), a context in which learners produce less French and are in contact with less French more generally (as is presumably the case in their home environment) may not be sufficient for the development of gender marking with adjectives. Although we have focused here on the two factors that showed change across time, it is just as relevant to highlight that the remaining six factors did not interact with time, suggesting that their influence remained stable over the course of the 21 months investigated, despite a change in learning context. This balance between change and stability in the developmental trajectory with gender marking has not been previously reported for AL French (but see Gudmestad et al., in press, for a similar discussion on gender-marking in Spanish), and future research would do well to focus on the interaction between time, learning context, and gender-

marking behavior suggested by these results. These observations demonstrate that findings from cross-sectional research can be both strengthened and, importantly, refined through longitudinal investigations of interlanguage.

In addition to arguing for the importance of longitudinal data, we have made the point (following Selinker 1972 and Ortega 2014) that the field of SLA stands to benefit from a renewed interest in detailed description of interlanguage and explanation of patterns found in those descriptions. Moreover, we suggested that data-driven approaches to AL acquisition are well-equipped to meet this goal, as language use and the identification of factors that characterize its variability are at the heart of such approaches. In the current study, we aimed to bring together the existing fragmentary observations on gender-marking behavior in a single analysis that drew upon usage-based perspectives and variationism. In this final section, we refer back to the complementary strengths of usage-based and variationist approaches to identify what this study contributes to SLA.

In usage-based approaches, distributional characteristics of the input (e.g., frequency, saliency, cue-interpretation reliability) are held to strongly influence language acquisition. In the case of the current study, special attention was given to one of these features, namely frequency. According to Tyler and Ortega (2018: 5), frequency is "a driver of learning mechanisms." However, as pointed out by Ellis and Wulff (2015), Ellis et al. (2016), Linford, Long, Solon and Geeslin (2016) and others, *frequency* is not a monolithic construct and can be operationalized and subsequently measured in multiple ways. In this study, we chose to analyze frequency in two manners: noun log-frequency (language) and noun frequency (individual). Consistent with most operationalizations of frequency in AL research, calculations for the first variable relied on a reference corpus in an attempt to reflect input frequency, as learners may show greater targetlikeness on nouns frequent in the input. For the second variable, we turned to researchers such as Tummers et al. (2005) and Bybee (2006), who speak of frequency effects that are a result of a speaker's own language use. In other words, a learner's interlanguage may be influenced by how often s/he makes use of a given lexical item or structure. In our analysis, the measure of external frequency did not significantly predict

gender-marking behavior. Thus, even though the stay abroad may have led to greater amounts of input, this finding suggests that the relative frequency of nouns does not constrain targetlike gender marking. We suggest that the input with which learners are confronted may explain this result. According to Ayoun (2007: 147), French does not provide "clear and unambiguous positive evidence" with respect to gender marking. Ayoun (2010) attempted to quantify this evidence by calculating the percentage of determiner phrases that included an overt cue to noun gender in a corpus of written native French. Results showed that only 50.14% of determiner phrases included a variable modifier, meaning that noun gender is often not unambiguously marked in the input. In the presence of input that does not consistently show gender marking, the role of input frequency may be reduced. This observation is consistent with Ellis (2015: 56), who stated that "[r]aw frequency of occurrence is less important than the contingency between cue and interpretation." On the other hand, the frequency variable that tapped into individual use was significant and showed that the more frequently an individual used a noun with a gender-marked modifier, the more likely s/he was to express gender in a targetlike manner with that noun. In other words, individual usage patterns impacted targetlikeness with gender marking. Taken together, the results from the two frequency variables show that whereas most AL research within a usage-based perspective has conceived of frequency with respect to the input learners receive, this is only part of the picture. More specifically, future research that looks to untangle the complex web of frequency effects in SLA will necessitate various operationalizations of frequency (see Ellis 2015), including more importance being given to frequency measures that reflect individual usage.

In the current study, variationist approaches inspired our investigation of both linguistic and extralinguistic factors in a single model, allowing us to move away from piecemeal, often univariate analyses of interlanguage to a recognition of systematic variation conditioned by multiple variables. Whereas linguistic factors have been identified in previous research, attention to the possible influence of extralinguistic variables is lacking. In the current project, we included two extralinguistic factors: time and task. Time was significant, and the importance of this factor has already been

addressed in connection with the longitudinal nature of the dataset. The difference between the interview and the narration, on the other hand, was not found to be significant. These two tasks differ in several respects, including the learners' familiarity with topics and the degree of control the learner has over the content they express. Despite these differences, gender-marking behavior did not significantly differ. Given the well-documented influence of task in AL behavior (Geeslin with Long 2014), future research may examine whether characteristics not represented in these two tasks impact gender marking. Moreover, it may be the case that differences between these two tasks are visible earlier in the learning trajectory. One additional promising avenue for future research with respect to task is possible with the LANGSNAP corpus itself. As was mentioned in the Method section, these same learners also completed a written argumentative essay (see Edmonds and Gudmestad 2018). Task effects could thus be investigated in future research by comparing gender-marking behavior across the three LANGSNAP tasks: interview, narration, essay.

Conclusion

The current study set out to respond to what Ortega and Byrnes (2008: 18) identified as "the dire need for longitudinal studies that can capture longer stretches of learners' developmental trajectories" by analyzing, in a bottom-up fashion, how a group of 20 learners expressed gender marking in oral AL French over 21 months. We focused on language use, which allowed us to offer important descriptive findings that support and refine results from previous research while simultaneously contributing to discussions concerning factors that explain gender-marking behavior. Two theoretical orientations that prioritize language use – namely, usage-based and variationist approaches – were drawn upon in order to design the study and interpret the results. This type of analysis is made possible by the availability of high-quality longitudinal learner corpora, particularly those that are shared freely and openly with the wider research community (Myles 2008; Tracy-Ventura and Huensch 2018). Indeed, the availability of learner corpora supports rich interlanguage description, allowing for complementary descriptions and analyses of the same dataset, thereby

enriching what we know about AL acquisition. The analysis presented here provides an example of just part of what longitudinal learner corpora have to offer to the field of SLA. We hope that other researchers may feel encouraged to take advantage of the many AL corpus resources currently available in order to undertake both rigorous descriptions and explanations of AL acquisition.

Ayoun, D. 2007. 'The second language acquisition of grammatical gender and agreement' in Ayoun, D. (ed.): *French Applied Linguistics*. Benjamins.

Ayoun, D. 2010. 'Corpus data: Shedding the light on French grammatical gender…or not,' *EUROSLA Yearbook* 10: 119-141.

Ayoun, D. 2018. 'Grammatical gender assignment in French: Dispelling the native speaker myth,' *Journal of French Language Studies* 28(1): 113-148.

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

Bayley, R. and **J. Langman.** 2004. 'Variation in the group and the individual: Evidence from second language acquisition,' *International Review of Applied Linguistics* 42: 303-318.

Berdan, R. 1996. 'Disentangling language acquisition from language variation' In Bayley, R. and D. R. Preston (eds.): *Second Language Acquisition and Linguistic Variation*. Benjamins.

Bybee, J. 2006. 'Usage to grammar: The mind's response to repetition,' *Language* 82(4): 711-733.

Carroll, S. 1989. 'Second-language acquisition and the computational paradigm,' *Language Learning* 39(4): 535-594.

Carroll, S. 1999. 'Input and SLA: Adults' sensitivity to different sorts of cues to French gender,' *Language Learning* 29(1): 37-92.

De Bot, K., W. Lowie and **M. Verspoor.** 2007. 'A Dynamic System Theory approach to second language acquisition,' *Bilingualism: Language and Cognition* 10(1): 7-21.

Dewaele, J.-M. 2015. 'Gender errors in French interlanguage: The effects of initial consonant versus initial vowel of the head noun,' *Arborescences: revue d'études françaises* 5: 7-27.

Dewaele, J.-M. and **D. Véronique.** 2001. 'Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study,' *Bilingualism: Language and Cognition* 4: 275-297.

Edmonds, A. and **A. Gudmestad.** 2018. 'Gender marking in written L2 French: Before, during, and after residence abroad,' *Study Abroad Research in Second Language Acquisition and International Education 3*(1): 59-82.

Ellis, N. C. 2015. 'Cognitive and social aspects of learning from usage' In Cadierno, T. and S. W. Eskildsen (eds.): *Usage-Based Perspectives on Second Language Learning*. de Gruyter Mouton.

Ellis, N. C., U. Römer and **M. B. O'Donnell.** 2016. *Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar.* Wiley.

Ellis, N. C. and **S. Wulff.** 2015. 'Usage-based approaches to SLA' In Cadierno, R. and Eskildsen, S. W. (eds.): *Usage-Based Perspectives on Second Language Learning.* de Gruyter Mouton.

Eskildsen, S. W. 2009. 'Constructing another language – usage-based linguistics in second language acquisition,' *Applied Linguistics* 30(3): 335-357.

Eskildsen, S. and **T. Cadierno.** 2015. *Usage-Based Perspectives on Second Language Learning*. de Gruyter Mouton.

Geeslin, K. L. with **A. Y Long.** 2014. *Sociolinguistics and Second Language Acquisition: Learning to Use Language in Context*. Routledge.

Gelman, A. and **J. Hill** 2006. *Applied Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Godfrey, L., C. Treacy and **E. Tarone.** 2014. 'Change in French second language writing in study abroad and domestic contexts,' *Foreign Language Annals* 47(1): 48-65.

Granfeldt, J. 2005. 'The development of gender attribution and gender concord in French: A comparison of bilingual first and second language learners' In Dewaele, J.-M. (ed.): *Focus on French as a Foreign Language: Multidisciplinary Approaches*. Multilingual Matters.

Gries, St. Th. 2015. 'Statistics for learner corpus research' In Granger, S., G. Gilquin and F. Meunier (eds.): *The Cambridge handbook of learner corpus research*. Cambridge University Press.

Gudmestad, A., A. Edmonds and **T. Metzger**. in press. 'Using variationism and learner corpus research to investigate grammatical gender marking in additional-language Spanish,' *Language Learning 69*(4)*.

Hakuta, K. 1974. 'Prefabricated patterns and the emergence of structure in second language acquisition,' *Language Learning* 24: 287-297.

Hardison, D. M. 1992. 'Acquisition of grammatical gender in French: L2 learner accuracy and strategies,' *The Canadian Modern Language Review* 48: 292-306.

Harley, B. 1979. 'French gender 'rules' in the speech of English-dominant, French-dominant, and monolingual French speaking children,' *Working Papers in Bilingualism* 19: 129-156.

Hasko, V. 2013. 'Capturing the dynamics of second language development via learner corpus research: A very long engagement,' *The Modern Language Journal* 97 Supplement: 1-10.

Holmes, V. and **B. Dejean de la Bâtie.** 1999. 'Assignment of grammatical gender by native speakers and foreign learners of French,' *Applied Psycholinguistics* 20: 479-506.

Jach, D. 2018. 'A usage-based approach to preposition placement in English as a second language,' *Language Learning* 68(1): 271-304.

Kass, R. E. and **A. E. Raftery.** 1995. 'Bayes factors,' *Journal of the American Statistical Foundation 90*(430): 773-795.

Labov, W. 1966. *The Social Stratification of English in New York City*. Center for Applied Linguistics.

Linford, B., A. Long, M. Solon and **K. L. Geeslin.** 2016. 'Measuring lexical frequency: Comparison groups and subject expression in L2 Spanish' In L. Ortega, A. E. Tyler, H. I. Park and M. Uno (eds.): *The Usage-Based Study of Language Learning and Multilingualism.* Georgetown University Press.

Lyster, R. 2006. 'Predictability in French gender attribution: A corpus analysis,' *Journal of French Language Studies* 16: 69-92.

Meisel, J. M. 2018. Early child second language acquisition: French gender in German children. *Bilingualism: Language and Cognition, 21*(4), 656-673.

Meunier, F. and **D. Littré.** 2013. 'Tracking learners' progress: Adopting a dual 'corpus cum experimental data' approach,' *The Modern Language Journal* 97 Supplement: 61-76.

Mitchell, R., K. McManus, K. and **N. Tracy-Ventura.** 2015. 'Placement type and language learning during residence abroad' in Mitchell, R., N. Tracy-Ventra, and K. McManus (eds.): *Social Interaction, Identity and Language Learning During Residence Abroad.* Eurosla Monographs Series.

Mitchell, R., N. Tracy-Ventura, and **K. McManus.** 2017. *Anglophone Students Abroad: Identity, Social Relationships, and Language Learning*. Routledge.

Myles, F. 2008. 'Investigating learner language development with electronic longitudinal corpora: Theoretical and methodological issues' In Ortega, L. and H. Byrnes (eds.): *The Longitudinal Study of Advanced L2 Capacities*. Routledge.

Nelson, D. 2005. 'French gender assignment revisited,' *Word* 56: 19-38.

New, B., M. Brysbaert, J. Veronis, and C. Pallier. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28: 661-677.

Ortega, L. 2014. 'Trying out theories on interlanguage: Description and explanation over 40 years of L2 negation research' In Han, Z. and E. Tarone (eds.): *Interlanguage: Forty Years Later*. Benjamins.

Ortega, L. and **H. Byrnes.** 2008. 'The longitudinal study of advanced L2 capacities: An introduction' in Ortega, L. and H. Byrnes (eds.): *The Longitudinal Study of Advanced L2 Capacities*. Routledge.

Ortega, L. and **G. Iberri-Shea.** 2005. 'Longitudinal research in second language acquisition: Recent trends and future directions,' *Annual Review of Applied Linguistics* 25: 26-45.

Ortega, L., A. E. Tyler, H. I. Park and **M. Uno**. (eds.) 2016. *The Usage-Based Study of Language Learning and Multilingualism.* Georgetown University Press.

Prévost, P. 2001. *The Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism and L2 Acquisition.* Benjamins.

Prodeau, M. 2005. 'Gender and number in French L2: Can we find out more about the constraints on production in L2?' in Dewaele, J.-M. (ed.): *Focus on French as a Foreign Language*. Multilingual Matters.

Reif, M., J. Robinson and **M. Pütz.** (eds.) 2013. *Variation in Language and Language Use: Socio-Cultural and Cognitive Perspectives*. Peter Lang.

Schumann, J. H. 1978. *The Pidginization Process: A Model for Second Language Acquisition*. Newbury House.

Selinker, L. 1972. 'Interlanguage,' *International Review of Applied Linguistics 10*: 219-231.

Smith, T. J. and **C. M. McKenna.** 2013. 'A comparison of logistic regression pseudo R2 indices,' *Multiple Linear Regression Viewpoints* 39(2) : 17-26.

Surridge, M. E. 1989. 'Le genre grammatical en Français fondamental: données de base pour l'enseignement et l'apprentissage,' *La Revue canadienne des langues vivantes* 45: 664-674.

Surridge, M. and **G. Lessard.** 1984. 'Pour une prise de conscience du genre grammatical,' *La Revue canadienne des langues vivantes* 41: 44-52.

Tracy-Ventura, N. and **A. Huensch.** 2018. 'The potential of publicly shared longitudinal learner corpora in SLA research' In Gudmestad, A. and A. Edmonds (eds.): *Critical Reflections on Data in Second Language Acquisition*. Benjamins.

Tracy-Ventura, N., K. McManus, J. Norris and **L. Ortega.** 2014. 'Repeat as much as you can': Elicited imitation as a measure of oral proficiency in L2 French' in Leclercq, P., A. Edmonds, and H. Hilton (eds.): *Measuring L2 Proficiency: Perspectives from SLA*. Multilingual Matters.

Tucker, R., W. Lambert, and **A. Rigault.** 1977. *The French Speaker's Skill with Grammatical Gender: An Example of Rule-Governed Behavior*. Mouton

Tummers, J., K. Heylen and **D. Geeraerts.** 2005. 'Usage-based approaches in cognitive linguistics: A technical state of the art,' *Corpus Linguistics and Linguistic Theory* 1-2: 225-261.

Wulff, S., N. Lester and M. Martínez-García. 2014. 'That-variation in German and Spanish L2 English,' *Language&Cognition* 6: 271-299.

Young, R. 1991. *Variation in Interlanguage Morphology.* Peter Lang.

Appendix A. Examples of coding.

Factor	le **début** de de mai 'the **beginning** of of May'	**Jacques** était vraiment contente '**Jacques** was very happy'	je pense que ça sera une bonne **expérience** 'I think that it will be a good **experience**'	je pense que ça sera une bonne **expérience** 'I think that it will be a good **experience**'	c'est pas le même **chose** du tout 'it's not the same **thing** at all'
Dependent variable	Targetlike	Nontargetlike	Targetlike	Targetlike	Nontargetlike
Participant	111	107	100	100	106
Initial proficiency	49	36	97	97	61
Time	In-stay	Post-stay	Pre-stay	Pre-stay	In-stay
Task	Interview	Narrative	Interview	Interview	Interview
Noun gender	Masculine	Masculine	Feminine	Feminine	Feminine
Noun log-frequency (language)	114.31	3.12	72.04	72.04	1,773.62
Noun frequency (individual)	2	1	1	1	6
Noun semantic class	No	No	No	No	No
Noun class	Grammatical	Biological	Grammatical	Grammatical	Grammatical
Noun-initial phoneme	Consonant	Consonant	Vowel	Vowel	Consonant
Noun-final phoneme	Vowel	Consonant	Consonant	Consonant	Consonant
Noun-final orthographic rhyme	Predictive	Not predictive	Predictive	Predictive	Predictive
Noun derivational morphology	Absent	Absent	Present	Present	Absent
Modifier type	Determiner	Adjective	Determiner	Adjective	Determiner

Syllable distance	0	4	1	0	1

Note. The noun concerned by the coding is provided in bold. The modifier concerned by the coding is underlined.

Appendix B. Results for the random effect in the

regression model

Effect	Participant	Parameter estimate	SE
Intercept	100	0.3325	0.2767
Intercept	101	0.1449	0.2197
Intercept	102	-0.4236	0.1866
Intercept	104	0.3515	0.2169
Intercept	105	-0.03586	0.2017
Intercept	106	-0.2987	0.1681
Intercept	107	0.09924	0.2184
Intercept	108	0.3498	0.2118
Intercept	109	-0.3732	0.2117
Intercept	110	0.7240	0.2238
Intercept	111	-0.4494	0.1708
Intercept	112	-0.7196	0.1806
Intercept	113	0.1685	0.1803
Intercept	114	-0.4402	0.1891
Intercept	115	-0.4416	0.1780
Intercept	116	-0.1270	0.1966
Intercept	117	-0.00997	0.2149
Intercept	118	0.7041	0.2573
Intercept	119	0.005512	0.1996
Intercept	120	0.3173	0.2464

[1] http://langsnap.soton.ac.uk/.

[2] See Ayoun (2018) for an account of variability in native-speaker gender assignment.

[3] For an exception, see Godfrey, Treacy and Tarone (2014). In their longitudinal study of eight learners of French, they examine accuracy with gender marking as a proxy for general accuracy with respect to the development of competence in academic writing. No analysis of factors affecting grammatical-gender marking is undertaken.

[4] Participant 100, who was 20 years old at the beginning of the project, self-reported 20 years of French study. According to the LANGSNAP website, this participant reported some at-home contact with French.

[5] We cite all examples exactly as they appeared in the LANGSNAP transcripts. However, coding reflected phonetic realization.

[6] Although in this model there are a relatively large number of fixed effects considered relative to the number of participants, we believe that our conclusions are valid and the estimates are reliable for two reasons: (a) the model did not exhibit any problems in its convergence or inflation in the estimation of standard error; and (b) the correlation structure we have chosen safeguards against the problem of finding significance solely due to a large number of observations (see also Gelman and Hill 2006: 275).