The interpretation of verbal moods in Spanish: A close replication of Kanwit and Geeslin (2014)

Aarnes Gudmestad[1] , Amanda Edmonds[2] , Carlos Henderson[3] and Christina Lindqvist[4]

[1]Department of Modern and Classical Languages and Literatures, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [2]Université Côte d'Azur, Nice, France; [3]University of Lund, Lund, Sweden; [4]University of Gothenburg, Gothenburg, Sweden
Corresponding author: Aarnes Gudmestad; Email: agudmest@vt.edu

Abstract
This study is a close replication of Kanwit and Geeslin (2014), a variationist investigation of the interpretation of verbal moods in adverbial clauses in Spanish. Whereas the first language (L1) of the second-language participants in the initial study was English, we explore whether Kanwit and Geeslin's results extend to other L1 populations—Swedish and French learners of Spanish. Participants in the replication study completed the same interpretation task and grammar test as those in the initial study. Results showed, for example, that multiple factors influenced their variable interpretation of verbal moods, there was evidence of change between course levels, and there were certain differences in interpretation between the French and Swedish groups. This study contributes to knowledge about the interpretation of a variable structure by enhancing the confirmatory power of some of the initial study's findings, while also suggesting that the learners' L1 leads to diverging findings.

Introduction

We conduct a close replication of Kanwit and Geeslin (Kanwit & Geeslin, K&G, 2014). These authors contributed to a fruitful line of research on sociolinguistic competence, using variationist methods to show that second-language (L2) learners acquire variable structures instantiated in the input (Geeslin with Long, 2014). Specifically, K&G (2014) conducted the first investigation of learners' *interpretation* of a variable structure (namely, mood distinction in Spanish adverbial clauses). Given the importance of this seminal article, we aim "to increase the confirmatory power of the original study" (Porte & McManus, 2019, p. 73) with a close replication[1] in which we expand the first language (L1) of the participants. Whereas the initial study involved L1 English speakers learning

---

[1]We adopt Porte and McManus's (2019) definition of close replication "whereby *only one* major variable is modified each time … and all others are kept as constant as possible to be better able to single out the kind of influences each has on the dependent variable" (p. 73).

Spanish, we recruited French- and Swedish-speaking participants. This replication study contributes to the growth of variationist second language acquisition (SLA) by conducting much-needed work on the interpretation of variable forms, thus providing a fuller picture of the development of sociolinguistic competence. This study also enriches both variationist SLA and SLA more broadly by increasing the L1-L2 combinations investigated.

Background

Variationist SLA

The current investigation is couched within variationist SLA. This line of inquiry aims to account quantitatively for the linguistic and extralinguistic factors that predict language variation and change (or development) and to advance the understanding of sociolinguistic competence among additional-language users. L2 studies on sociolinguistic variation demonstrate that learners develop similar sensitivity to the linguistic and extralinguistic factors that influence variable patterns in their input (Geeslin with Long, 2014). However, this research has largely focused on learner production in oral tasks or on the selection of variable forms through written contextualized instruments (Gudmestad, 2022), meaning that little is known about how learners develop the ability to interpret variable structures. Investigations on interpretation are important because they advance knowledge of the linguistic and social meaning that learners assign to the input they receive (Michalski, 2023). They are also valuable because learners may be able to interpret forms in linguistically and socially meaningful ways before this variation is manifested in language production, thus providing information about the developmental trajectory that cannot be uncovered in language production (Geeslin & Hanson, 2023). In the first study of its kind, K&G (2014) investigated the interpretation of a variable structure by English-speaking L2 learners, focusing on the interpretation of verbal moods (the subjunctive and indicative) in Spanish adverbial clauses.

Initial research study

K&G (2014) investigated L2 learners' and bilingual native speakers' interpretation of verbal moods in Spanish in contexts with a temporal adverb (i.e., *cuando* "when," *hasta que* "until," *después de que* "after"). Traditionally, in contexts following these adverbial conjunctions, the indicative is linked with a habitual interpretation, and the subjunctive is connected to the interpretation of an event or state that has not yet occurred (K&G, 2014, p. 490). Sociolinguistic research, however, has found evidence of variation in the use of verbal moods for each interpretation, including with the aforementioned adverbs (e.g., Murillo Medrano, 1999). More generally, research has demonstrated that verbal mood is undergoing a change in Spanish, with the use of the subjunctive gradually decreasing and the indicative increasing (Harris, 1974). Corpus linguistics has also shown that the subjunctive occurs infrequently in Spanish, compared to the indicative (Biber et al., 2006). These observations led K&G (2014) to question how categorically each mood is linked with a particular interpretation. The research questions that guided their study were (p. 498):

(1) Do NNSs [non-native speakers] interpret clauses in the indicative as habitual and clauses in the subjunctive as having not yet occurred? How do their interpretations

change across levels of proficiency? How do their interpretations compare with those of NSs [native speakers]?

(2) What linguistic and extralinguistic factors predict NNSs's mood interpretation? Do the factors change across proficiency levels? Do the same factors predict NSs mood interpretation?

To answer these questions, K&G (2014) designed an interpretation task, which enabled them to examine participants' interpretation of the subjunctive and indicative in adverbial clauses according to a series of predetermined linguistic factors (see **METHOD**). Participants also completed a grammar test and a background question-naire. The L2 participants were enrolled in a fifth-semester undergraduate Spanish-language course, a fourth-year undergraduate Spanish linguistics course, or a graduate program in Spanish. The bilingual NSs, who served as a benchmark for targetlikeness, had been living in the United States (U.S.) for at least one year. K&G (2014) analyzed the data quantitatively. Generally, the examination of cross-sectional data provided evidence of change along the developmental trajectory, showing, for instance, that verbal mood did not influence interpretation among the least proficient learners. This study was important because it initiated SLA research on the interpretation of variable structures and because it demonstrated that the traditional interpretations of verbal moods in adverbial clauses in Spanish are variable for learners and NSs.

The replication study

Because research on the interpretation of sociolinguistic variation is in the early stages, we do not yet know whether K&G's (2014) findings generalize to other learner populations. Specifically, it is unclear whether learners whose L1 encodes the subjunctive-indicative contrast may show a different acquisitional pathway when compared to speakers reporting English as their only L1, perhaps showing the influence of verbal mood in interpretation earlier in the learning trajectory. This issue is worth investigating because research on crosslinguistic influence and aspect, for example, has shown that the presence of the grammaticalized aspect in the L1 can facilitate the learning of aspect in an L2 (cf. McManus, 2021). Thus, it seems reasonable to hypothesize that positive transfer may be at play in the acquisition of the subjunctive-indicative contrast for learners whose L1 contains a productive mood system. More generally, a focus on the role of the L1 is justified as it is an underexplored issue in variationist SLA (Bayley, 2005; Geeslin with Long, 2014, p. 162). Numerous authors (Geeslin with Long, 2014, p. 162; Judy & Perpiñán, 2015; Ortega, 2009, pp. 6–7), moreover, underscore the fact that SLA is dominated by work on English (as an L1 or L2), with Tracy-Ventura et al. (2021, p. 410) arguing that, "[I]f we are to better understand the process of L2 learning, it is essential that we expand our focus to include different languages, especially typologically distant languages." These reflections led us to conduct a close replication of K&G (2014), in which we modified one variable—the participants' L1—to study Spanish learners with two typologically distinct L1s (French and Swedish). Investigating these under-studied language combinations allows us to explore the generalizability of previous results.

While we believe there is an inherent value for SLA in investigating any under-explored language pairing, we chose Swedish and French because their mood systems differ. In Swedish, the present subjunctive is limited to fixed expressions (*Gud ske*SUBJC

lov! 'Thank heavens!'[2]), and the past subjunctive only occurs rarely with *vore* "to be" with a modal meaning (Holmes & Hinchliffe, 2013, pp. 311–312). The use of the subjunctive is more robust in French and shares various similarities with Spanish. Like Spanish, it tends to occur in subordinate clauses with different governors (e.g., *vouloir que* "to want that") and exhibits variation (Poplack et al., 2013). Regarding French synonyms of the temporal adverbials examined in the current study, grammarians report indicative use with *quand* "when," subjunctive use with *jusqu'à ce que* "until," and variation with *après que* "after" (Grévisse & Goosse, 2001, Section 866). Unlike Spanish that has distinct verb forms for each mood, many French forms in the present indicative and subjunctive are homonyms (e.g., *je chante* "I sing$_{\text{INDIC/SUBJC}}$"). The similarities between French and Spanish may mean that having French as an L1 facilitates the interpretation of Spanish mood, compared to Swedish and English (where the subjunctive is also limited).

Research questions

Our first two research questions are the same as those in K&G (2014), except for three modifications. One is that we do not compare learners to NSs. This is because our goal is to assess whether the same developmental patterns hold for other L1 groups, so it is not necessary to collect data from NSs. The second difference pertains to proficiency. Whereas K&G (2014) examined three groups, we analyzed two. The universities to which we had access for data collection do not have graduate-student populations that are large enough for quantitative analysis, thus we investigated two undergraduate course levels only and, therefore, opt for the term *course level*, rather than proficiency. The third difference is that we added a third question that focuses on the comparison of different L1 groups. Differences between K&G's research questions and ours are specified with strikethrough and bold text in our research questions.

(1) Do NNSs interpret clauses in the indicative as habitual and clauses in the subjunctive as having not yet occurred? How do their interpretations change across levels of proficiency (i.e., **course levels**)? ~~How do their interpretations compare with those of NSs?~~

(2) What linguistic and extralinguistic factors predict NNSs' mood interpretation? Do the factors change across levels of proficiency (i.e., **course levels**)? ~~Do the same factors predict NS mood interpretation?~~

(3) **How do the results for L1 speakers of French and Swedish compare to the L1 speakers of English in K&G (2014)?**

A summary of the differences between the initial and replication studies is in Table 1. L1 background gave rise to the hypothesis guiding this study and, as such, was the only major modification. The remaining changes were minor and were necessary either because of the major change or because they reflected the specificities of the research contexts.

Method

We followed the same procedure as K&G (2014), unless otherwise noted.

---

[2]This subjunctive form (*ske*) is identical to the infinitive form.

Table 1. Summary of changes between studies

Variable	Initial study	Replication study
L1	English	French, Swedish
Participant groups	L2 level 1, 2, 3, and NSs	L2 level 1, 2
Task	**Initial study**	**Replication study**
Language of responses	English	French or Swedish
Wording of 8 items	See **DATA COLLECTION** and Supplementary Materials	

Participants

We collected data from two contexts (France and Sweden) in intact classes with participants enrolled in undergraduate-level Spanish courses. In the initial study, the primary criterion used to gauge L2 proficiency was course level. A grammar test (see **DATA COLLECTION**) served as a secondary measure of Spanish proficiency. Because course levels may vary across different learning contexts, we used course level and the score on the grammar test to match the L2 proficiency of our participants to that of the Level 1 and 2 participants in K&G (2014) as closely as possible. In K&G (2014), the Level 1 participants were enrolled in a fifth-semester language course and the Level 2 participants were taking a fourth-year content course. Based on our shared knowledge of undergraduate programs in the United States, we roughly equated K&G's (2014) Level 1 and 2 with B1 and B2/C1 levels, respectively, in the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001). We then identified undergraduate courses in France and Sweden that aligned with these CEFR levels (see Supplementary Materials for details).

Following Marsden et al. (2018, p. 333), our target sample size for each L1 group was one that was at least as large as the sample size in K&G (2014). This goal was achieved (French: $n = 94$, Swedish: $n = 88$, English: $n = 64$[3]). After collecting data in intact classes, we removed participants who did not meet any one of the following four criteria. First, one participant each in France and Sweden did not complete all tasks, so they were excluded from the analysis. The remaining three criteria, not cited by K&G (2014), were necessary to ensure comparability between the initial and replication studies. In particular, two criteria were necessary to ensure that groups shared the central criterion of L1. To begin, individuals who reported Spanish as one of their L1s were removed from the analysis (France: $n = 5$; Sweden: $n = 13$). Next, we excluded participants who did not indicate the country's national language as an L1 (France: $n = 20$; Sweden: $n = 11$). The final criterion helped us ensure comparability in proficiency level between the initial and replication studies' participants. We compared the grammar-test scores of our participants to those in K&G (2014). The score range in the initial study for Level 1 learners was 8–20. To help ensure comparability, we excluded anyone who scored above the maximum score (20) for Level 1 learners in K&G (2014). This led to the removal of four Swedish participants; no French participants were excluded. We note, too, that no Level 1 participant in the replication study scored below the minimum score in K&G (2014). For Level 2, we removed participants who were not advanced enough according to the grammar test (i.e., anyone who scored below the lowest score of 12 in K&G's (2014) Level 2 group). We excluded six French and two Swedish participants. Descriptive statistics on the grammar test, which are provided in Tables 2 and 3, demonstrate that, for the most part, our participants scored similarly to their

---

[3]The sample size in K&G (2014) pertains to Levels 1 and 2 only because those are the levels we replicated.

Table 2. Descriptive statistics on grammar-test scores

Level	L1	N	M	Range	SD
1	English	29	13.40	8–20	2.96
	French	40	12.22	9–20	3.30
	Swedish	37	14.27	9–20	3.10
2	English	35	17.83	12–24	3.31
	French	22	17.00	13–23	3.56
	Swedish	20	20.20	15–23	2.84

Note. The L1 English participants are from K&G (2014, p. 501).

Table 3. Effect size for between-group comparisons

Course level	L1 comparison	d	CI
Level 1	French vs. English	−0.37	[−0.85, 0.11]
	Swedish vs. English	0.29	[−0.21, 0.77]
Level 2	French vs. English	−0.24	[−0.78, 0.29]
	Swedish vs. English	0.74	[0.17, 1.31]

corresponding course level in the initial study. Following Plonsky (2015), we calculated effect sizes between the L1 English groups in K&G (2014) and the L1 groups in the replication study (Table 3). These between-group comparisons indicate that there is no meaningful difference between the Level 1 comparisons or between the Level 2 French-English comparison. However, there is a medium effect size for the Level 2 Swedish-English comparison, suggesting that there may be a difference between groups, with the Swedish Level 2 participants showing a higher proficiency level in Spanish than the initial study's English-speaking Level 2 participants.

The French group ($n = 62$) ranged in age from 17 to 22 ($M = 18.97$, $SD = 1.25$). Forty-four were women, 17 were men, and one declined to answer. Twenty participants indicated that at least one other language was an L1, in addition to French (6 = Arabic, 2 = English, 2 = Russian, and one of each of the following: Algerian, Cape Verdean Creole/Portuguese, Chechen/Russian, Comorian, English/Lithuanian, Hebrew, Italian, Malagasy, Portuguese, and Romanian); 12 of these participants were in Level 1 and eight were in Level 2. All but one participant indicated that they knew languages other than their L1(s) (see the Supplementary Materials for these details for all participants). Forty-three participants had spent time in a Spanish-speaking country (in months: $M = 1.26$, Range = 0.1–10.5, $SD = 2.08$). Nineteen participants had either not spent time in a Spanish-speaking country or did not provide this information.

The average age of the Swedish group ($n = 57$) was 29.54 (Range = 18–78, $SD = 14.49$, one participant did not provide his age). Forty-three were women, and 14 were men. Five participants reported two L1s (four in Level 1 and one in Level 2): Swedish along with English ($n = 4$) or Italian ($n = 1$). The remaining 52 participants stated that Swedish was their only L1. Fifty-five participants indicated that they knew at least one other language. One participant reported not knowing any languages beyond her L1s (Swedish/English), and another participant elected not to provide this information. Forty-seven participants had spent time in a Spanish-speaking country (in months: $M = 5.94$, Range = 0.25–36, $SD = 8.48$). Ten participants had either not spent time in a Spanish-speaking country or did not provide this information.

It is worth acknowledging that the current participant pools reflect diverse language backgrounds that are characteristic of millions of speakers who do not have just one L1. We recognize that this diversity brings challenges to research that seeks to examine the role of the L1. We continue to use the term L1 in the current investigation but limit its use to refer to the one L1 that each group shares (namely, French and Swedish). We also note that, although some of our participants indicated that English was one of their L1s, there are still two relevant differences between these participants and those in K&G (2014). One is that the sociocultural context is different (i.e., the data come from three countries). The other difference is that the L1 English participants in the initial study did not report any L1s other than English (M. Kanwit, personal communication, April 17, 2023).

Data collection

As in the initial investigation, the participants in the current study completed three tasks. K&G (2014) shared the mood interpretation task and grammar test (what they call the language-proficiency exam) with us. Regarding the mood interpretation task, each of the 24 items contained a Spanish sentence that consisted of an independent clause and a subordinate clause introduced by an adverbial conjunction. All verb forms were in the present tense. Half of the items contained a subjunctive verb, and half contained an indicative verb in the subordinate clause. The participants were given three response options: an interpretation of habituality, an interpretation of an event that had not yet occurred, and the possibility to respond that both interpretations were possible. These responses were in English in the original task. They were translated into French and Swedish for the current study to align with the participants' L1. Eight items on the interpretation task were slightly modified, including the replacement of a U.S.-specific proper noun and small wording changes, following feedback from Spanish speakers in France and Sweden (see Supplementary Materials for the instrument and the modifications). These changes did not impact the coding (see **DATA CODING**).

The second task was a grammar test that was contextualized in a Spanish story. Throughout the story, there were 25 multiple-choice items that covered various Spanish grammatical structures. The only change made to the instrument was the language of the instructions. Originally in English, we translated them into French and Swedish. Finally, like in K&G (2014), the participants in the replication study completed a background questionnaire that elicited data on demographics and language experiences. We designed our own background questionnaire to gather sufficient details on participants' L1(s). Importantly, however, the questionnaire included all details that K&G (2014) provided on their participants (e.g., gender, age).

Data coding

Unless otherwise specified, we followed the same data-coding procedures found in K&G (2014). The dependent variable is identical in both studies—the response for each item on the interpretation task ("habitual", "not-yet-occurred", and "both" interpretations are possible). We coded for the same four independent linguistic variables in the interpretation task: verbal mood, verbal morphology regularity, clause order, and adverbial conjunction (see Supplementary Materials for an example of the coding).[4]

---

[4]These factors were motivated by previous research (see K&G, 2014).

Verbal mood pertains to whether the verb in the dependent clause in the Spanish sentence of each task item is in the subjunctive or indicative. Verbal morphology regularity also focuses on the verbs in the subordinate clause of each Spanish sentence. Regular verbs are those that have the same stem in both moods (e.g., *yo tomo*$_{\text{INDIC}}$/ *tome*$_{\text{SUBJC}}$ "I take"), whereas irregular verbs have stems that differ between moods (e.g., *Marta tiene*$_{\text{INDIC}}$/*tenga*$_{\text{SUBJC}}$ "Marta has"). Clause order distinguishes between Spanish sentences in which the independent clause is preposed or postposed in relation to the subordinate clause. Finally, each Spanish sentence included one of three adverbial conjunctions (*cuando*, *después de que*, *hasta que*). Notably, "The instrument was designed in such a way that all of the possible combinations of each of the categories of these variables occurred one time on the instrument, making isolation of the effects of a single variable possible" (K&G, 2014, p. 501).

In addition to the independent linguistic factors, K&G (2014) coded for three independent extralinguistic variables. They called the first *proficiency*, which was the score on the grammar test. We call this factor grammar-test score because it clearly distinguishes the variable from the other proficiency metric, which is the second factor —course level (1 or 2). The third extralinguistic variable is participant. The only change made to the coding involved the addition of a fourth independent variable: L1 (French or Swedish).

Data analysis

K&G (2014) investigated all independent variables except for course levels as fixed effects in regression analyses; course levels were analyzed in separate analyses. We diverged from this approach in one way: Following current norms in the field, we included participant as a random intercept in the analysis (cf. Gries, 2021). Like the initial study, we analyzed each course level separately. Our analysis consisted of six steps. We first ran a crosstabulation that showed the frequency of response selection. Second, we fit one mixed-effects model (using lme4 in RStudio, RStudio Team, 2023) for each course level in which we analyzed the factors that predicted the response for all items on the task. Following the second step, we analyzed the indicative and subjunctive items separately. The third and fourth phases of the analysis consisted of crosstabulations and mixed-effects models for each course level and the indicative items. The final two steps were comprised of crosstabulations and mixed-effects models for each course level and the subjunctive items. In each regression, we considered interactions between L1 and other fixed effects, as this enabled us to identify possible differences in interpretation between L1s. Like K&G (2014), for the regression models, significant effects were determined by *p*-values, where α was set at 0.05 and p < α was significant. For model selection of the fixed effects, we followed a procedure similar to Gries (2021).

To facilitate comparison of the results between the initial and replication studies, we include results from K&G (2014) in our crosstabulation tables. We used three criteria for comparing the results, as shown in (a-c).

   a. For crosstabulations, we examined overall distributions. Our comparisons focused on similarities and differences in the response that was selected most often.[5]

---

[5]Unlike K&G (2014), we do not conduct chi-square tests on the crosstabulations in Tables 3, 5, and 6 because doing so would violate the statistical assumption of independence.

b. For regression models, we identified similarities and differences in the independent variables that were significant in each model.

c. For regression models, certain statistical details that are now commonly reported for regressions were not provided in K&G (2014). For this reason, our interpretation relied on the direction of effect estimates for significant factors in the replication study, which we compared to the crosstabulation distributions in the initial study (i.e., the response selected most or least often).

Results

Full dataset

It will be recalled that traditional interpretations link the subjunctive with events that have not-yet occurred and the indicative with events that occur habitually, and half of the task items included each mood. Thus, if participants' responses were fully aligned with traditional interpretations, the not-yet-occurred and habitual responses would each be selected 50 percent of the time, and the both response would never be chosen (K&G, 2014, p. 505). Table 4 shows the distribution of responses on the task for each participant group. Although the precise proportions differed, the tendency among each participant group was to select the not-yet-occurred interpretation most often (ranging from 38.0% for Level 1 French to 48.8% for Level 2 Swedish) and the both response least often (17.8% for Level 2 Swedish and 26.3% for Level 1 Swedish).

Following the crosstabulations, we fit two binary mixed-effect models for the full dataset, one for each course level. We collapsed the not-yet-occurred and both responses to identify "what factors would permit an interpretation that has been traditionally associated exclusively with the subjunctive" (K&G, 2014, p. 506). In Table 5, we share an overview of the significant effects in our models. In particular,

Table 4. Distribution of responses in full dataset

| | Level 1 | | | | | | Level 2 | | | | | |
| | English | | French | | Swedish | | English | | French | | Swedish | |
Response	#	%	#	%	#	%	#	%	#	%	#	%
Habitual	251	36.2	357	37.2	248	28.1	296	35.3	198	37.8	155	33.3
Not-yet-occurred	276	39.8	364	38.0	403	45.6	364	43.4	207	39.5	227	48.8
Both	167	24.1	238	24.8	232	26.3	179	21.3	119	22.7	83	17.8
Total	694	100.0	959	100.0	883	100.0	839	100.0	524	100.0	465	100.0

Note. L1 English participants are from K&G (2014, p. 505).

Table 5. Summary of the significant effects for the full-dataset regression models

Effect	Level 1	Level 2
L1	–	X
Mood	X	–
Adverb	X	X
Verbal morphology regularity	X	–
Clause order	X	X
Grammar-test score	–	–
L1*mood	–	X

Note. "X" denotes significant and "–" denotes a non-significant factor.

the mixed-effects models for both course levels demonstrate that mood played a significant role in the interpretation of this task. It was a significant main effect for Level 1, and in the Level 2 model, there was a significant interaction between L1 and mood. This differs somewhat from K&G (2014), who found mood to be a predictive factor for Level 2 but not Level 1. Because they found that mood was such an important predictor of interpretation, K&G (2014) focused on separate analyses for the indicative and subjunctive items. We follow suit and move now to the analyses by mood (for full statistical details corresponding to the models summarized in Table 5, see Supplementary Materials).

Separate analyses by mood

Next, we conducted separate analyses for each mood, beginning with the indicative. Table 6 provides the crosstabulations. With the indicative, which is traditionally connected with habituality, the habitual interpretation was generally selected most often. The one exception was the Level 1 Swedish group, who chose the not-yet-occurred response most frequently. Across course levels, we note similar rates of response selection between the Level 1 and 2 French groups. The Swedish group, however, increased their interpretation of the indicative as habitual by about 24 percent from Level 1 to 2.

We then fit two mixed-effects models for the indicative items, one for each course level. The results for the main effects are in Table 7 (see Supplementary Materials for statistical details for the random effect in each model). In these models, the both response is collapsed with the not-yet-occurred interpretation to be able "to examine the predictors of any context in which an indicative form allows an interpretation that is traditionally thought to require a subjunctive form" (K&G, 2014, p. 509). For the Level 1 participants, the adverb and clause order significantly predicted the interpretation of the indicative items on the task. L1, verbal morphology regularity, and grammar-test score were not significant, and there were no significant interactions between L1 and other fixed effects. Level 1 was less likely to select the habitual interpretation with indicative items when the adverb was *después de que* or *hasta que* compared to *cuando*. We were also interested in whether *después de que* and *hasta que* (the two non-reference point categories) were different from each other. For this, we used a pairwise Tukey test, with an adjusted *p*-value due to multiple comparisons (Table 8). This test indicated that these two adverbs were similar. Finally, the log-odds of choosing the habitual interpretation were lower when the main clause was first. These findings show similarities to the Level 1 group in K&G (2014). The two

Table 6. Distribution of responses on indicative items

	Level 1						Level 2					
	English		French		Swedish		English		French		Swedish	
Response	#	%	#	%	#	%	#	%	#	%	#	%
Habitual		40.3	191	39.9	133	30.2		50.5	101	38.7	127	54.5
Not-yet-occurred	N/A	36.5	158	33.0	185	42.0	N/A	23.3	94	36.0	71	30.5
Both		23.2	130	27.1	123	27.9		26.2	66	25.3	35	15.0
Total		100.0	479	100.0	441	100.0		100.0	261	100.0	233	100.0

Note. L1 English participants are from K&G (2014, p. 508).
N/A means not available.

Table 7. Results for fixed effects in the regression models (indicative items)

Effect	Estimate	SE	*p*-value	CI
		LEVEL 1		
(Intercept)	0.17	0.18	0.34	[–0.18, 0.52]
Adverb [*Cuando*]				
Después de que	–0.83	0.18	<0.0001	[–1.19, –0.47]
Hasta que	–1.17	0.19	<0.0001	[–1.55, –0.80]
Clause order [Post]				
Pre	–0.50	0.15	0.001	[–0.80, –0.20]
		LEVEL 2		
(Intercept)	0.86	0.29	0.003	[0.29, 1.43]
Adverb [*Cuando*]				
Después de que	–0.85	0.26	0.001	[–1.35, –0.34]
Hasta que	–1.23	0.26	<0.0001	[–1.75, –0.71]
Clause order [Post]				
Pre	–0.65	0.21	0.002	[–1.07, –0.24]

Note. The model fit the log-odds of the habitual response. Reference points for the independent variables are in brackets. Level 1: McFadden R^2 = 0.08; Bayesian Information Criterion [BIC] = 1131.2, versus 1200.58 for the null model. Level 2: McFadden R^2 = 0.13; BIC = 625.90 versus 688.11 for the null model.

Table 8. Pairwise differences of adverb (indicative items)

Comparison	Estimate	SE	*z*-ratio	*p*-value
		LEVEL 1		
Cuando vs. *después de que*	0.83	0.18	4.54	<0.0001
Cuando vs. *hasta que*	1.17	0.19	6.19	<0.0001
Después de que vs. *hasta que*	0.34	0.19	1.79	0.1742
		LEVEL 2		
Cuando vs. *después de que*	0.85	0.26	3.31	0.003
Cuando vs. *hasta que*	1.23	0.26	4.65	<0.0001
Después de que vs. *hasta que*	0.38	0.26	1.47	0.3034

significant factors in their regression model were adverb and clause order, and the direction of these effects aligned with our findings.

Like the Level 1 model, adverb and clause order significantly predicted interpretation of indicative items for Level 2. Grammar-test score, L1, and verbal morphology regularity were not significant; neither were there any significant interactions involving L1. The direction of the significant effects also paralleled those found for Level 1. The habitual interpretation was less likely with *después de que* and *hasta que* (vs. *cuando*), and interpretation between *después de que* and *hasta que* was similar (Table 8). The Level 2 participants were also less probable to select the habitual interpretation of indicative items when the main clause was first. Our Level 2 model for the indicative items diverges from the one in K&G (2014), who found that grammar-test score and participant (analyzed as a fixed effect) significantly predicted interpretation.

Finally, we turn to the subjunctive items. Table 9 demonstrates that the traditional interpretation—not-yet-occurred—was the most frequent interpretation for all participant groups, though the proportion ranged from 42.9 percent (Level 1 French) to 67.2 percent (Level 2 English and Swedish). The distribution of responses was similar between the two French groups. Once again, evidence of change in the same direction as the initial study is seen with the Swedish participants, who increased their interpretation of the subjunctive as not-yet-occurred from 49.5 percent at Level 1 to 67.2 percent at Level 2.

Table 9. Distribution of responses on subjunctive items

| | Level 1 | | | | | | Level 2 | | | | | |
| | English | | French | | Swedish | | English | | French | | Swedish | |
Response	#	%	#	%	#	%	#	%	#	%	#	%
Habitual	N/A	31.3	166	34.6	115	26.0	N/A	17.2	97	36.9	28	12.1
Not-yet-occurred		43.6	206	42.9	218	49.3		67.2	113	43.0	156	67.2
Both		25.1	108	22.5	109	24.7		15.6	53	20.2	48	20.7
Total		100.0	480	100.0	442	100.0		100.0	263	100.0	232	100.0

Note. L1 English participants are from K&G (2014, p. 508).

Table 10. Results for fixed effects in the regression models (subjunctive items)

Effect	Estimate	SE	*p*-value	CI
		LEVEL 1		
(Intercept)	0.18	0.18	0.32	[−0.18, 0.54]
Adverb [*Cuando*]				
Después de que	−0.50	0.18	0.006	[−0.85, −0.14]
Hasta que	0.04	0.17	0.83	[−0.30, 0.38]
Verbal morphology regularity [Irregular]				
Regular	−0.81	0.15	<0.0001	[−1.11, −0.51]
Clause order [Post]				
Pre	0.51	0.15	0.0004	[0.23, 0.80]
		LEVEL 2		
(Intercept)	0.05	0.26	0.84	[−0.46, 0.56]
L1 [French]				
Swedish	1.20	0.29	<0.0001	[0.62, 1.78]
Adverb [*Cuando*]				
Después de que	−0.45	0.25	0.07	[−0.94, 0.03]
Hasta que	0.71	0.25	0.005	[0.22, 1.20]
Verbal morphology regularity [Irregular]				
Regular	−0.81	0.21	0.0001	[−1.22, −0.39]

Note. The model fit the log-odds of the not-yet-occurred response. Reference points for the independent variables are in brackets. Level 1: McFadden R^2 = 0.07; BIC = 1220.80 versus 1279.04 for the null model. Level 2: McFadden R^2 = 0.12; BIC = 638.20 versus 688.68 for the null model.

The final step consisted of separate mixed-effects models for each course level and the subjunctive items (see Table 10). In these models, the both and habitual responses are collapsed "to examine the predictors of responses for which the interpretation generally not found with subjunctive forms was allowed" (K&G, 2014, p. 509). For Level 1, adverb, verbal morphology regularity, and clause order significantly predicted interpretation. L1 and grammar-test score were not significant, and there were no interactions between L1 and other main effects. Regarding the adverb, the Level 1 participants were less likely to select the not-yet-occurred interpretation with *después de que* versus *cuando*. The participants' interpretation did not differ between *hasta que* and *cuando*. Table 11 shows that the participants were less probable to select the not-yet-occurred interpretation with *después de que* compared to *hasta que*. The results for verbal morphology regularity demonstrated that Level 1 was less likely to select the not-yet-occurred interpretation with regular, compared to irregular, verbs. Lastly, Level 1 was more probable to select the not-yet-occurred interpretation when the main clause was first. Our Level 1 model of subjunctive items is similar to the one in K&G (2014) with regard to the significant effects for verbal morphology regularity, clause order, and

Table 11. Pairwise differences of adverb (subjunctive items)

Comparison	Estimate	SE	z-ratio	p-value
LEVEL 1				
Cuando vs. *después de que*	0.50	0.18	2.73	0.017
Cuando vs. *hasta que*	−0.04	0.17	−0.22	0.975
Después de que vs. *hasta que*	−0.53	0.18	−2.94	0.009
LEVEL 2				
Cuando vs. *después de que*	0.46	0.25	1.83	0.160
Cuando vs. *hasta que*	−0.71	0.25	−2.83	0.013
Después de que vs. *hasta que*	−1.17	0.26	−4.51	<0.0001

the fact that the grammar-test score was not significant. However, findings differed with the adverb, which was not an influential factor in their study.

In the mixed-effects model for Level 2 and the subjunctive items, L1, verbal morphology regularity, and adverb were significant predictors. Clause order and grammar-test score were not significant, and there were no significant interactions between L1 and other fixed effects. The Swedish group was more likely to select the not-yet-occurred interpretation than the French group. For the adverb factor, *hasta que* was more likely to yield selection as not-yet-occurred compared to *cuando*. The pairwise Tukey test indicated that Level 2 participants were less probable to choose the not-yet-occurred interpretation with *después de que* versus *hasta que* (Table 11). There was no significant difference in interpretation between *después de que* and *cuando*. Lastly, the log-odds of selecting the not-yet-occurred interpretation were lower with regular, versus irregular, verbs. Our model was similar to K&G's (2014) in that the adverb and verbal morphology regularity were significant, and clause order and grammar-test score were not. Although both studies found the adverb to be an important predictor, its role differed insofar as K&G (2014) found that their Level 2 participants selected the not-yet-occurred interpretation most often with *cuando* and least often with *después de que*. The direction of the effect for verbal morphology regularity, however, was similar between the two investigations, with irregular verbs being linked to the traditional subjunctive interpretation.

Discussion

Main findings

The first research question addressed whether learners interpreted the indicative as habitual and the subjunctive as having not yet occurred, and whether their interpretations changed as the course level advanced. All groups, except for Level 1 Swedish, interpreted the indicative as habitual most often. However, the not-yet-occurred interpretation was also common, and each group chose the both response at least 15 percent of the time, demonstrating that learners' interpretation of the indicative was variable. Although change was not observed with the French group, the Swedish Level 2 group selected the habitual interpretation more often than Level 1. Turning to the subjunctive, each group interpreted this mood as not-yet-occurred most often, but just as with the indicative, their interpretation was variable. Like the indicative, the rates of subjunctive selection were similar for both French groups. It was the Swedish group who showed signs of change, increasing their interpretation of the subjunctive as not-yet-occurred by about 18 percent from Level 1 to Level 2. These findings are important because they attest to variation in the interpretation of verbal moods, regardless of the

course level and L1 under investigation, and because they suggest that there may be differences in development between L1 French and Swedish learners.

The second question focused on the linguistic and extralinguistic factors that predicted interpretation and whether these predictors changed cross-sectionally. When all task items were analyzed together, we found that mood significantly predicted interpretation for both course levels (as a main effect for Level 1 and in an interaction with L1 for Level 2). This result contrasts with K&G (2014), who found that although Level 2's interpretation was influenced by mood, Level 1 did not rely on subjunctive and indicative morphology in their interpretation of adverbial clauses. They subsequently concluded that this non-significant result was due to Level 1's limited proficiency in Spanish. The replication study's significant effect for mood at Level 1, however, suggests that the L1 may, at least partly, contribute to the role that mood plays in interpretation at early points along the developmental trajectory. The fact that the effect sizes between the Level 1 participants in the initial study and each L1 group in the replication study do not indicate meaningful differences in proficiency adds weight to the observation that this difference in the effect for mood may be attributed to the L1. Turning to the indicative items, adverb and clause order were significant for Levels 1 and 2, and the direction of these effects was similar between course levels. Thus, with the indicative items, we did not see evidence of development. With the subjunctive items, adverb, verbal morphology regularity, and clause order predicted Level 1's interpretation, and L1, adverb, and verbal morphology regularity influenced Level 2's interpretation. While the effect that verbal morphology regularity had on the subjunctive items was constant between course levels, the other predictors offered evidence of change. First, L1 impacted the Level 2 group only. Second, clause order impacted Level 1 only. Third, although adverb was significant for both groups, the effect it had on interpretation differed. Thus, evidence of development, as seen in differences between the predictors of the Level 1 and 2 models, appeared to be limited to the subjunctive mood. In K&G (2014), they found evidence of development among L1 English participants in the indicative and subjunctive items, which points to another difference between studies. The fact that the subjunctive occurs less frequently than the indicative in the input (Biber et al., 2006) may be why this mood is more susceptible to fluctuations in development.

For the final research question, which compared the results in the initial and replication studies, our findings were similar to K&G's (2014) results in various ways, thus supporting their results. We focus this comparison on the separate analyses for each mood, given the importance that K&G (2014) gave to the need to examine each mood separately. These studies showed that L1 English, French, and Swedish learners exhibited variable interpretations of verbal moods following adverbial conjunctions, with most groups interpreting the indicative most often as habitual and the subjunctive most often as not-yet-occurred. Additionally, in terms of the factors predicting the interpretation of verbal moods, an overall comparison of our results to those in the initial study revealed that four factors significantly impacted interpretation, with more similarities than differences (see Table 12 for a summary). Beginning with Level 1, we note that the initial and replication studies' participants were similar (Table 12). The only difference concerned the adverb factor with the subjunctive items, which significantly predicted subjunctive interpretation for the L1 French and L1 Swedish participants but not the English-speaking learners in K&G (2014). In light of the small effect sizes of the grammar-test score between the initial and replication study participant groups, this difference in the effect for the adverb could be attributed to the role of the L1.

Table 12. Comparison summary of predictive factors

Variable	Level 1		Level 2	
	Indicative	Subjunctive	Indicative	Subjunctive
Adverb	S	D	D	D
Verbal morphology regularity	S	S	D	S
Clause order	S	S	D	S
Grammar-test score	S	S	D	S

Note. S = similarity between initial and replication studies, D = difference.

Despite numerous similarities among Level 1 groups, the comparison of the Level 2 participants between studies (Table 12) shows several differences and suggests the need to temper certain results from K&G (2014). First, we note that, like for Level 1, the only difference between the initial and replication studies for the subjunctive models concerned the factor adverb. Although the adverb was an influential factor in both studies, the direction of this effect differed. The not-yet-occurred interpretation appeared to be most likely with *hasta que* in the current study, whereas the participants in K&G (2014) selected the not-yet-occurred interpretation most often with *cuando*. Like K&G (2014), these data show different interpretation patterns as a function of the adverb, underscoring the important role of a lexical item in language variation. Differences are more striking when we turn to the Level 2 indicative models, which exhibited no similarities. We note, though, that these differences may have several sources, including methodological decisions, the role of proficiency, and the role of the L1. Beginning with methodological concerns, K&G (2014) analyzed participant as a fixed effect and found it to be significant in the Level 2 indicative model, thus asserting that there was notable individual variability in this group. The current study, however, included participant as a random effect in all models. This points to the inherent difficulty in comparing across studies, even in cases of close replication.

Different results for the indicative items among the Level 2 groups may also be partially due to the medium effect size on the grammar-test score between the L1 Swedish and L1 English participant groups. Although this possibility cannot be excluded, we believe that it is unlikely for two reasons. First, in the analysis of indicative items for Level 2, the L1 French and L1 Swedish participants did not perform in a significantly distinct manner. Second, the effect size comparison of proficiency level showed no meaningful difference between L1 French and L1 English participants. Considering these two points together, if the higher proficiency of the L1 Swedish learners were at play, we would not expect the L1 Swedish and L1 French learners (whose proficiency was more closely aligned with that of the L1 English participants) to behave in the same way.

Finally, our inclusion of two L1 groups revealed patterns that nuanced those from the initial study in two respects. First, the L1 French participants did not show a substantial shift in their frequency of interpretation of either verbal mood between course levels, unlike the L1 English participants, whereas the Swedish group was comparable to K&G's (2014) participants in that change in interpretation was observed. Second, the regression analysis of the Level 2 subjunctive items revealed that L1 significantly impacted interpretation, with the Swedish learners being significantly more likely to select the not-yet-occurred response than the French learners. On the basis of their data, K&G (2014, p. 524) presented developmental stages for mood interpretation. By uncovering certain differences among L1 groups in verbal mood

interpretation, this replication study suggests that these stages are not necessarily valid for every learner population. While the participants' L1 may account for these differences, we note that they cannot be explained on purely typological grounds. If typological differences concerning mood systems were consistently playing a key role in interpretation, we would expect Swedish and English learners to pattern similarly to each other and differently from French learners, which is not always the case. In short, this replication study strengthens the confirmatory power of the initial study, given the many common results between the two investigations, but it also suggests that learners' L1 is an important factor to consider because it seems to shape the interpretation of a variable structure, albeit in limited ways.

Future replication research

Given the fact that the current study is the first to replicate research on the interpretation of variable structures in SLA, there are numerous avenues for future scholarship. Replication studies that continue to examine variable mood distinction in Spanish should consider additional L1s. It is worth examining other L1 English participant pools as well to determine whether K&G's (2014) findings generalize to populations with the same L1. Additionally, about ten years have passed since the publication of the initial study, and methodological practices, undoubtedly, evolve over time. For example, it is now common to include task item as a random intercept in a regression to account for task-item-related variability (Winter, 2020, Section 14.4). Another change that has occurred in variationist SLA is the increased use of multinomial regression models (e.g., Gudmestad et al., 2020), which allow for the examination of three or more distinct levels of a dependent variable. The dependent variable in the current study has three levels (habitual, not-yet-occurred, both). We collapsed the both response with another response on the task and fit binary regression models to make our results as comparable to the initial study as possible. However, because analyzing the both response on its own may reveal new insights about the interpretation of variable moods (cf. Kanwit & Geeslin, 2018), a possible direction for future research is to conduct a conceptual replication that employs multinomial regression techniques. Finally, scholarship on the interpretation of variable structures is in the nascent stages, but as research on other L2s (e.g., French; Kanwit & Arnold, in press) and variable structures (e.g., copula contrast in Spanish, Kanwit & Geeslin, 2020) emerges, it will be important to replicate these findings, too.

Conclusion

As the first close replication of K&G's (2014) pioneering work on the L2 interpretation of variation, the current study has extended the initial investigation to two new populations: L1 speakers of French and Swedish. In so doing, it has contributed to knowledge about the variable interpretation of verbal moods in Spanish by confirming many of the findings observed in the initial study and by identifying differences that may be attributed to learners' L1. Additional research is needed regarding the potential role of the L1 because the differences we observed between L1 French and L1 Swedish participants and between the initial and replication studies could also be attributed to variability in Spanish input in the three countries and to pedagogical differences in undergraduate Spanish courses. Nevertheless, the present investigation confirms that expanding the L1-L2 combinations investigated in SLA is important for knowledge

building in the field and that much work remains to determine the factors that lead to similarities and differences among learners of diverse backgrounds.

References

Bayley, R. (2005). Second language acquisition and sociolinguistics. *Intercultural Communication Studies*, *14*(2), 1–13. https://www-s3-live.kent.edu/s3fs-root/s3fs-public/file/01-Robert-Bayley.pdf

Biber, D., Davies, M., Jones, J., & Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora*, *1*, 1–37. https://www.mark-davies.org/articles/davies_30.pdf

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Strasbourg, France: Language Policy Unit. https://assets.cambridge.org/97805218/03137/frontmatter/9780521803137_frontmatter.pdf

Geeslin, K. L., & Hanson, S. (2023). Sociolinguistic approaches to communicative competence. In M. Kanwit & M. Solon (Eds.), *Communicative competence in a second language: Theory, method, and applications* (pp. 40–59). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003160779-4/sociolinguistic-approaches-communicative-competence-kimberly-geeslin-stacey-hanson

Geeslin, K. L., with Long, A. Y. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context*. Routledge.

Grévisse, M., & Goosse, A. (2001). *Le bon usage* (13th edition). DeBoeck Duclot.

Gries, S. Th. (2021). (Generalized linear) mixed-effects modeling: A learner corpus example. *Language Learning*, *71*, 757–798. https://doi.org/10.1111/lang.12448

Gudmestad, A. (2022). Eliciting variable structures across tasks. In K. L. Geeslin (Ed.), *The Routledge handbook of second language acquisition and sociolinguistics* (pp. 224–236). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003017325-21/eliciting-variable-structures-across-tasks-aarnes-gudmestad

Gudmestad, A., Edmonds, A., Donaldson, B., & Carmichael, K. (2020). Near-native sociolinguistic competence in French: Evidence from variable future-time expression. *Canadian Journal of Applied Linguistics*, *23*(1), 169–191. https://files.eric.ed.gov/fulltext/EJ1262425.pdf

Harris, M. (1974). The subjunctive mood as a changing category in Romance. In J. M. Anderson & C. Jones (Eds.), *Historical linguistics II* (pp. 169–188). North-Holland Publishing Company.

Holmes, P., & Hinchliffe, I. (2013). *Swedish: A comprehensive grammar* (3rd ed.). Routledge. http://soedi.se/filer/pdf/Swedish__A_Comprehensive_Grammar.pdf

Judy, T., & Perpiñán, S. (2015). The importance of crosslinguistic comparison in the study of the acquisition of Spanish. In T. Judy & S. Perpiñán (Eds.), *The acquisition of Spanish in understudied language pairings* (pp. 1–18). John Benjamins. https://benjamins.com/catalog/ihll.3.01jud

Kanwit, M., & Arnold, M. (in press). Lexical effects on mood interpretation in French adverbial clauses. *Canadian Journal of Linguistics*.

Kanwit, M., & Geeslin, K. L. (2014). The interpretation of Spanish subjunctive and indicative forms in adverbial clauses: A cross-sectional study. *Studies in Second Language Acquisition*, *36*, 487–533. https://www.jstor.org/stable/26330706

Kanwit, M., & Geeslin, K. L. (2018). Exploring lexical effects in second language interpretation: The case of mood in Spanish adverbial clauses. *Studies in Second Language Acquisition*, *40*, 579–603. https://doi.org/10.1017/S0272263117000262

Kanwit, M., & Geeslin, K. L. (2020). Sociolinguistic competence and interpreting variable structures in a second language: A study of the copula contrast in native and second-language Spanish. *Studies in Second Language Acquisition*, *42*, 775–799. https://www.cambridge.org/core/journals/studies-in-second-lan guage-acquisition/article/abs/sociolinguistic-competence-and-interpreting-variable-structures-in-a-sec ond-language/503DFEBAC25A261CD4CEDE33B4A44CC3

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*, 321–391. https://doi.org/10.1111/lang.12286

McManus, K. (2021). *Crosslinguistic influence and second language learning*. Routledge. https://doi. org/10.4324/9780429341663

Michalski, I. (2023). L2 sociolinguistic perception of stylistic variation: Attitudes toward variable linguistic features of Spanish. In S. Fernández Cuenca, T. Judy, & L. Miller (Eds.), *Innovative approaches to research in Hispanic linguistics: Regional, diachronic, and learner profile variation* (pp. 225–247). John Benjamins. https://doi.org/10.1075/ihll.38.10mic

Murillo Medrano, J. (1999). Subjuntivo e indicativo en las oraciones circunstanciales [Subjunctive and indicative in circumstantial sentences]. *Kanina*, *23*, 143–156. https://www.semanticscholar.org/paper/ Subjuntivo-e-indicativo-en-las-oraciones-Medrano/f747db50920b3ab9b996c7d065b063f4d7ee0c7c

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education. https://doi. org/10.4324/9780203777282

Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23–45). Routledge. https://experts.nau.edu/en/publications/ statistical-power-p-values-descriptive-statistics-and-effect-size

Poplack, S., Lealess, A., & Dion, N. (2013). The evolving grammar of the French subjunctive. *Probus*, *25*, 139–195. https://doi.org/10.1515/probus-2013-0005

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge. https:// www.routledge.com/Doing-Replication-Research-in-Applied-Linguistics/Porte-McManus/p/book/ 9781138657359

RStudio Team. (2023). *RStudio: Integrated development for R* [Computer software]. RStudio, PBC. http:// www.rstudio.com

Tracy-Ventura, N., Paquot, M., & Myles, F. (2021). The future of corpora in SLA. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 409–424). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781351137904-36/future-corpora-sla- nicole-tracy-ventura-magali-paquot-florence-myles

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge. https://doi.org/10.4324/ 9781315165547